Instance-dependent Optimality in Statistical Decision-making

Wenlong Mou

Acknowledgement

Research in this dissertation was partly funded by grants from the National Science Foundation and Office of Naval Research.

Instance-dependent Optimality in Statistical Decision-making

By

Wenlong Mou

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Martin J. Wainwright, Co-chair
Professor Peter L. Bartlett, Co-chair
Professor Michael I. Jordan
Associate Professor Adityanand Guntuboyina

Summer 2023

Instance-dependent Optimality in Statistical Decision-making

Abstract

Instance-dependent Optimality in Statistical Decision-making

by

Wenlong Mou

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Martin J. Wainwright, Co-chair

Professor Peter L. Bartlett, Co-chair

Data-driven and learning-based methodologies have been very popular in modern decision-making systems. In order to make optimal use of data and computational resources, these problems require theoretically sound procedures for choosing between estimators, tuning their parameters, and understanding bias/variance trade-offs. In many settings, asymptotic and/or worst-case theory fails to provide the relevant guidance.

In this dissertation, I present some recent advances that involve a more refined approach, one that leads to non-asymptotic and instance-optimal guarantees. Focusing on function approximation methods for policy evaluation in reinforcement learning, in Part I, I describe a novel class of optimal oracle inequalities for projected Bellman equations, as well as computationally efficient algorithms achieving them. In contrast to corresponding results for ordinary regression, the approximation pre-factor depends on the geometry of the problem, and can be much larger than unity. In Part II, I discuss optimal procedures for estimating linear functionals from observational data. Our theory reveals a rich spectrum of behavior beyond the asymptotic semi-parametric efficiency bound. It also highlights the fundamental roles of geometry, and provides concrete guidance on practical procedures and parameter choices.

To mom, dad, and Rui

Contents

Contents ii

List of Figures v

List of Tables vii

1 Introduction 1
1.1 High-level overview of the results 2
1.2 Related work not appearing in this thesis 6
1.3 Notation 7

I Reinforcement learning with function approximation 11

2 Optimal oracle inequalities for projected fixed-point equations 12
2.1 Introduction 12
2.2 Background 18
2.3 Main results for general projected linear equations 24
2.4 Consequences for specific models 36
2.5 Proofs 43
2.6 Discussion 52

3 Optimal and instance-dependent guarantees for Markovian stochastic approximations 54
3.1 Introduction 54
3.2 Problem set-up 58
3.3 Main results 65
3.4 Some consequences for specific problems 71
3.5 Proof of Proposition 3.1 81
3.6 Proof of Theorem 3.1 94
3.7 Discussion 100

II Off-policy estimation of linear functionals 102

4 A non-asymptotic theory for semi-parametric efficiency 103
4.1 Introduction . . . 103
4.2 Non-asymptotic and instance-dependent upper bounds . . . 108
4.3 Minimax lower bounds . . . 123
4.4 Proofs of upper bounds . . . 130
4.5 Proofs of minimax lower bounds . . . 139
4.6 Discussion . . . 149

5 Beyond semi-parametric efficiency: a kernel-based analysis 151
5.1 Introduction . . . 151
5.2 Problem set-up and preview . . . 158
5.3 Main results and their consequences . . . 162
5.4 Simulation studies . . . 175
5.5 Proofs . . . 180
5.6 Discussion . . . 199

Bibliography 201

A Proofs and discussion deferred from Chapter 2 217
A.1 Proofs deferred from Section 2.5.1 . . . 217
A.2 Proofs deferred from Section 2.5.2 . . . 220
A.3 Proof of Corollary 2.1 . . . 223
A.4 Proof of Theorem 2.3 . . . 225
A.5 Proof of Corollary 2.2 . . . 232
A.6 Proof of Theorem 2.4 . . . 233
A.7 Proof of the bounds on the approximation factor . . . 243
A.8 Proofs for the examples . . . 245
A.9 Additional simulation studies . . . 259

B Proofs and discussion deferred from Chapter 3 263
B.1 Additional related work . . . 263
B.2 Auxiliary truncation results . . . 265
B.3 Auxiliary results underlying Proposition 3.1 . . . 267
B.4 Auxiliary results underlying Theorem 3.1 . . . 271
B.5 Proof of Theorem 3.2 . . . 278
B.6 Proofs for the examples . . . 295

C Proofs and discussion deferred from Chapter 4 302
C.1 Proofs of auxiliary results in the upper bounds . . . 302
C.2 Proofs of the corollaries . . . 306
C.3 Some elementary inequalities and their proofs . . . 313
C.4 Empirical process results from existing literature . . . 314

D Proofs and discussion deferred from Chapter 5 **317**

D.1 Properties of effective dimension . . . 317

D.2 Relaxing the effective dimension condition . . . 319

D.3 Proof of technical lemmas . . . 321

D.4 Conditional variance estimation . . . 327

D.5 Proofs for the examples . . . 334

List of Figures

2.1 Plots of various approximation factor as a function of the discount factor γ in the policy evaluation problem. (See the text for a discussion.) (a) Results for an Erdös-Rényi random graph model with $N = 3000$, projected dimension $d = 1000$, and $a = 3$. The resulting number of vertices in the graph $\widetilde{G}$ is 2813. The value of $1 - \gamma$ is plotted in log-scale, and the value of approximation factor is plotted on the standard scale. (b) Results for a random geometric graph model with $N = 3000$, projected dimension $d = 2$, and $r = 0.1$. The resulting number of vertices in the graph $\widetilde{G}$ is 2338. Both the discount rate $1 - \gamma$ and the approximation factor are plotted on the log-scale. 28

4.1 Plots of the normalized MSE $n \cdot \mathbb{E}[|\widehat{\tau}_{n,\diamond} - \tau^*|]$ for $\diamond \in \{\omega, \mathbb{L}^2, \text{oracle}\}$ versus the sample size. Each marker corresponds to a Monte Carlo estimate based on the empirical average of 1000 independent runs. As indicated in the figure titles, panels (a–f) show the normalized MSE of estimators for combinations of parameters: exponent $\gamma \in \{0, 0.5, 1\}$ in the top, middle and bottom rows respectively, and propensity scores π_1 and π_2 in the left and right columns, respectively. For each run, we used 5-fold cross validation to choose the value of regularization parameter $\lambda_n \in [10^{-1}, 10^2]$. 116

5.1 Log-log plots of mean-squared error versus sample size n for four different estimators fo τ^*: procedure Opt-KRR is analyzed in this chapter, whereas CV-KRR is a related method with regularization parameter chosen by cross-validation. We also compare to the classical IPW estimate along with a truncated version of IPW. (a) Setting $\alpha = 0.5$ yields a propensity score with light tails, and our theory predicts classical n^{-1}-decay of the MSE for Opt-KRR. (b) Setting $\alpha = 2.0$ yields a heavy-tailed problem, and our theory guarantees consistency of Opt-KRR at the rate $n^{-3/4}$. 154

5.2 Plots of the mean-squared error $\mathbb{E}\big[\,|\widehat{\tau}_{n,\diamond} - \tau^*|^2\,\big]$ versus sample size n. Each curve corresponds to a different algorithm $\diamond \in \big\{\mathrm{ipw}, \mathrm{trunc}, \mathrm{opt}, \mathrm{cv}\big\}$. Each marker corresponds to a Monte Carlo estimate based on the empirical average of 2000 independent runs. For the cross-validated estimator $\widehat{\tau}_{n,\mathrm{cv}}$, the choice of regularization parameter is based on averaging the cross validation results of the first 50 runs. As indicated by the sub-figure titles, each panel corresponds to a problem setup $(\xi, \pi) \in \big\{\xi^{\mathrm{L}}, \xi^{\mathrm{N}}, \xi^{\mathrm{C}}\big\} \times \big\{\pi^{\mathrm{L}}, \pi^{\mathrm{N}}\big\}$. Both axes in the plots are given by logarithmic scales. Some of the curves may overlap with each other. 178

5.3 Plots of the mean-squared error $\mathbb{E}\big[\,|\widehat{\tau}_{n,\diamond} - \tau^*|^2\,\big]$ versus sample size n for $\diamond \in \big\{\mathrm{ipw}, \mathrm{trunc}, \mathrm{opt}, \mathrm{cv}\big\}$. The simulation parameters are exactly the same as Figure 5.2, except for the underlying problem instances. As indicated by the sub-figure titles, each panel corresponds to an exponent $\alpha \in \{0.5, 1, 2, 3\}$. We have already presented part of the results (the cases of $\alpha = 0.5$ and $\alpha = 2$) in Section 5.1. 181

A.1 A graphical illustration of the MRP instance constructed above. For this instance, we let $d = 1$, $|\mathcal{S}_1| = 4$ and $|\mathcal{S}_2| = 4$, so that the total number of states is $D = 10$. In the graph, solid rounds stand for states, and arrows stand for the possible transitions. The numbers associated to the arrows stand for the probability of the transitions, and the equations $r = \cdots$ standard for the reward at a state. The sets $\mathcal{S}_0$, $\mathcal{S}_1$ and $\mathcal{S}_2$ are separated by red dotted lines, and the sets Γ_1, $\bar{\Gamma}_1$, Γ_2, and $\bar{\Gamma}_2$ are marked by transparent rectangles. A blue round stands for a state with positive value function, and an orange round stands for a state with negative value function. 249

A.2 Plots of empirical MSE and theoretical bounds as a function of the discount factor γ in the policy evaluation problem. (See the text for a discussion.) (a) Results for an Erdös-Rényi random graph model with $N = 3000$, projected dimension $d = 100$, and $a = 30$. The resulting number of vertices in the graph $\widetilde{G}$ is 3000. The value of $1 - \gamma$ is plotted in log-scale, and the value of approximation factor is plotted on the standard scale. (b) Results for a random geometric graph model with $N = 3000$, projected dimension $d_0 = 3$, $k = 4$, and $r = 1$. The resulting number of vertices in the graph $\widetilde{G}$ is 2991 and the resulting feature dimension is $d = 31$. Both the discount rate $1 - \gamma$ and the approximation factor are plotted on the log-scale. The fluctuations in the plot are due to randomness of the estimators. 261

List of Tables

2.1 Bounds on the approximation factor $\frac{\mathbb{E}\|\widehat{v}_n - v^*\|^2}{\mathcal{A}(\mathbb{S}, v^*)}$ for proper estimators in different ranges of ambient dimension. Here, $c_q \in (0, 1)$ represents a constant depending only on the aspect ratio q. 36

Acknowledgments

Six years have passed, and I still clearly remember the sunny afternoon when I first arrived at Hearst street with two huge suitcases. So many things have changed in the world, at Berkeley, and inside myself, while there are also important invariants that remain. Looking back at the wonderful journey, the education at Berkeley has profoundly shaped my way of thinking and doing, with all the happiness and pains. There are so many great people I would like to thank, and I apologize in advance if I missed anyone.

Let me begin by thanking my two advisors, Martin Wainwright and Peter Bartlett, for the six years of inspiration and fun. Both of them are brilliant researchers and invaluable mentors. In my future career as a faculty, I want to be like them.

The interaction with Martin is a unique experience, which not only helped me developed my research tastes and academic career, but also brought happiness and laughter to life. The first meeting with Martin at the visit days made me quickly realize that Martin is a great fit – we share the same aesthetics for beautiful math, as well as the same grumps about bad research. The discussion with Martin has always been a pleasure. Martin is one of the sharpest researcher I have ever seen – he can always grasp the key ideas in complicated proofs, and provide thoughtful feedback. Martin's high standards about writing, presentation, and even coding skills also helped me a lot. Outside the world of concentration inequalities and convex geometry, we also share a lot in common, including the love for spicy food and messy offices.

I have been a big fan of Peter's courses when I first learned statistical learning theory back in Peking University as an undergraduate student, and it is truly a privilege working with him for my Ph.D. studies. Peter's cheerful character in life, caring and inclusive attitude towards others, and curiosity in research have always been a role model for me. I have been continually impressed by Peter's knowledge about literature – often times when I read about or come up with a new idea, Peter can find deep connections to a lesser-known paper back in the nineties. Apart from academia, I would also like to thank Peter for sharing with me lots of tourism attractions around the world. I am eager to visit these places one day.

I am also specially grateful to my committee members, Mike Jordan and Aditya Guntuboyina. During the past six years, I feel extremely fortunate to work with Mike on several collaborative projects. I would like to thank Mike for sharing great insights on the future directions of machine learning research, as well as helpful guidance on my career. Through the interaction with Mike's group with diverse research interests, I am able to explore different possibilities, refresh my minds, and create something new. In the past six years, I have always been trying to learn as much as possible from Mike. I would also like to thank Aditya for providing helpful feedbacks in my qualifying exam and thesis talk, and teaching a wonderful semester of theoretical statistics class. The solid training from Aditya lays the foundations of all the statistical research I have done during my Ph.D.

Next, let me thank my wonderful collaborators throughout the six years. Peng Ding is a great methodological researcher with an open mind towards modern tools from theoretical statistics and machine learning. I would like to thank Peng for introducing me to the fantastic area of causal inference. It is amazing how our collaboration goes well, even if we are from very different technical background. Thank you to Ashwin Pananjady, a good friend, a sharp researcher, and a helpful peer. The intellectual journey of discovery together with Ashwin has always been a pleasure, and Ashwin's advice on writing, presentation, and job market are indispensable. Special thanks to Nicolas Flammarion and Yi-An Ma, my first two collaborators at Berkeley. I would like to thank them for introducing me to the area of MCMC and countless whiteboard discussions that lead to the nice pieces of work. More importantly, I would like to thank the support and encouragement they offered me in my earlier years of Ph.D., which gave me the confidence to achieve what I have achieved now. We also had lots of fun together with shared interest outside research, ranging from hiking to food. Thanks to Yi-An for the beautiful natural views in San Diego. I would also like to thank Nhat Ho for the amazing sparks of idea we create together, when Bayesian theory and latent variable models meets modern probabilistic tools. Nhat's advice on job market and career development has also been a great help to me. I would like to thank Zheng Wen and Xi Chen for the wonderful summer I spent at Adobe research. I learned a lot from them about both research and career. Thanks to Koulik Khamaru for the inspiring discussion we had, as well as the wonderful smoothie. Thank you to Chris Junchi Li for the piece of art we created together. Thanks to Fangzhou Su for bringing the field of causal inference to my attention, and for tolerating me as I just get to learn how to mentor junior students. I would also like to thank Licong Lin and Yuhang Cai for their great creativity in our ongoing projects.

From a fresh graduate at college to a mature researcher, my academic development would be impossible without the great classes I have learned at Berkeley. Apart from Aditya, I would also like to thank many other teachers that offered great classes: David Aldous, Jim Pitman, Tom Courtade, Craig Evans, Daniel Tataru, Nike Sun, Will Fithian, Venkat Anantharam, Jiantao Jiao, Luca Trevisan, Avishay Tal, Song Mei, Lin Lin, Jacob Steinhardt, and Michael Christ. I would also like to thank Dawn Song, who guided me at the very beginning of the journey. Additionally, I would like to thank Arkadi Nemirovski at Georgia Tech. During the year of covid, I have audited Arkadi's modern convex optimization sequence remotely, which is a masterpiece.

Berkeley offers an unparalleled community of researchers in the area of statistical machine learning, broadly defined. Thriving in this environment, I benefit a lot from interaction with them. Exchange of ideas at group meetings, whiteboard discussion on random math questions, and informal conversations at lunches or hikes, could all become sparks of inspiration. Apart from the collaborators I have already mentioned above, I would like to thank my wonderful peers in this community: Tianyi Lin, Feng Ruan, Raaz Dwivedi, Yeshwanth Cherapanamjeri, Kush Bhatia, Yuting Wei, Chi Jin, Yuansi Chen, Aaditya Ramdas, Nilesh Tripuraneni, Niladri Chatterji, Andrea Zanette, Mike Celentano, Xiang Cheng, Lihua Lei, Avishek Ghosh, Efe Aras, Vipul Gupta, Soham

Phade, Kuan-Yun Lee, Banghua Zhu, Alex Tsigler, Juanky Perdomo, Aldo Pacchiano, Dong Yin, Lin Chen, Vidya Muthukumar, Cong Ma, Bin Shi, Ishaq Aden-Ali, Yannik Pitcan, Spencer Frei, Wei Hu, Chiao-Yu Yang, Devon Ding, Yaodong Yu, Yang Chu, Jianbo Chen, Zhuang Liu, Yu Sun, Zhe Cao, Zihao Chen, and Shizhan Zhu. Many thanks to my wonderful colleagues with whom I had lots of fun during my visit to MIT: Yaqi Duan, Reese Pathak, Eric Xia, Jian Qian, and Xiyu Zhai. Although we haven't formally collaborated, the intellectual exchange we had are invaluable to me.

The past six years have been filled with joy and excitement. In addition to the aforementioned, I would like to take this opportunity to thank a few other friends with whom I had lots of fun: Yixuan Li, Xiaohan Yan, Huibin Chen, Ruocheng Wang, Jin Zhu, Zhuqi Li and Zhenyi Li, Haoran Luo, Yu Zhang, Sijie Chen, Lijie Chen, Hao He, Yuchen Zhou and Lili Zheng, Zhi Wang, Shengtao Dai, Haoyang Wu, Tianru Wang, Zihao Zheng, Yu Tong, Yuan Yao, Jun Gao, Kai Zheng, Xusong Wang, Xinfeng Xie, Yi Zhao, Pan Xu, Qian Li, Shuyang Shi and Yichun Hu, Yuchen Wang, Liu Yang, and Nianze Jiao. Special thanks to Yixuan Li and Tianyi Lin for making the great dinner for my wedding guests recently, and for countless weekends of hikes, great food and vodka we had together.

I was introduced to the area of machine learning theory and data science by my undergraduate advisor, Prof. Liwei Wang, at Peking University. I am extremely grateful to Liwei for nearly ten years of inspiration. Liwei is a deep thinker and lifetime friend of mine. Our continual conversations keep me thinking out of the box, and caring about the role of theory in the future of artificial intelligence. I would also like to thank Prof. Nina Balcan at Carnegie Mellon, who provided me with generous guidance and support.

Finally, and most importantly, I would like to thank my family. Many thanks to my parents, for their unreserved love and consistent support. Backed by them, I was able to explore the world and pursue my dream as I want. Thank you to Rui Zhang, a talented researcher, a unique and interesting soul, an ally that I can always trust, and my dear wife. During my Ph.D. studies, you brought me so much joy and gave me so much support and inspiration. You made me a better self that I did not imagine before. I believe we have a bright future ahead of us.

Chapter 1

Introduction

In recent years, applications of machine learning and data science have profoundly changed people's everyday life. While these technologies have achieved tremendous success in lab environment, when applied to real-world decision-making problems, several puzzles still remain unresolved. In particular, it has been observed that many important choices are made in an ad hoc manner in practice, without principled guidance. These practical challenges put statistical theory at a critical position — in the world of decision-making, machine learning is not just about stacking more layers and tuning parameters. Instead, theoretically well-grounded methodologies are needed, concerning structures in the model, data, and the target. As the data in decision-making problems can be of low quality, inter-dependent, and expensive to collect, novel machine learning algorithms addressing these issues with optimal guarantees could lead to significant saving. Yet, existing statistical and computational theories are still lacking in many aspects.

During my Ph.D. studies, the focus of my research is drawn on the theoretical understanding into computational and statistical aspects of modern decision making applications. A popular class of methodologies in decision-making is the learning-based approaches, i.e., first learn a model or estimate some relevant functions, and then compute the target value or optimal policy from the learned model. For example, in Markov decision processes, this corresponds to reinforcement learning with function approximation; and in causal estimation problems with observational data, this corresponding to semi-parametric methods such as double machine learning. Though the learning-based approaches offer lots of flexibility in modeling, it is observed in practice that the classical principles in learning may fail for decision-making applications. This naturally motivates the recurring theme in my research:

> How to make learning optimal for decision-making?

In order to address this key question, I investigate the probabilistic and geometric structures underlying the problems, on which a rich class of statistical theories and computational methods are built. I aim at the fundamental questions about the possibility, optimality, and computational efficiency of statistical estimation. Notwithstanding the seemingly simplicity, these questions lead to fruitful outcomes, which, not only facilitates data-driven methods in decision-making, but also contributes to novel statistical theories.

Another important feature of the research results presented in this dissertation is their *instance-dependent optimality*. Minimax optimality has been the golden criteria to evaluate the performance of statistical estimation and learning algorithms. However, for many decision-making applications including reinforcement learning and causal inference, there are many methods that all achieve the global minimax optimality, but exhibit extremely different practical performances. This is because the worst-case problem instances within the problem class are usually too hard to solve with any practical sample size, so that the worst-case optimality theory fails to guide the practical choice. By way of contrast, in this dissertation, we adopt a more fine-grained criteria of optimality – the estimator has to be minimax optimal not only globally, but also locally within suitably defined neighborhood of any problem instance. In this way, we can obtain key quantities that governs the complexity of estimation associated to any problem instance.

The notion of instance-dependent optimality dates back to the local asymptotic minimax theory due to Le Cam and Hájek [117, 72]. However, in many modern decision-making applications, taking the asymptotic limit with sample size going to infinity can hide important finite-sample phenomena, especially when we are only able to model parts of the environment instead of the entire data-generating process. With a complicated model class and limited sample size, the asymptotic theory can be completely irrelevant. In this dissertation, we discover novel instance-dependent quantities that govern the non-asymptotic complexity, which are not covered by the classical local asymptotic minimax theory. In order to make the estimators adaptive to these complexities, novel statistical principles are revealed, which provide concrete guidance on practical choices.

The rest of the introduction is organized as follows: we first provide a high-level overview of the results in this dissertation; then, we summarize some additional related work of the author, which are not included in this dissertation; finally we introduce notations used throughout this dissertation.

1.1 High-level overview of the results

In this section, we provide a high-level overview of the main results presented in this dissertation.

1.1.1 Part I: reinforcement learning with function approximation

Approximate dynamic programming (ADP) and reinforcement learning (RL) provides a formalism of making optimal decisions in sequential settings. A central question in ADP and RL is the estimation of value function, cast as the problem of solving a (linear) Bellman fixed-point equation. In most practical applications, the state-action space is enormous or infinite (for example, the game of Go has a state space of cardinality

3^{361}, and most control problems work with continuous states), so that direct plug-in methods are impossible both computationally and statistically. A natural approach, therefore, is to use a function class $\mathcal{F}$ to approximate the solution v^*, and solve the *projected fixed-point equations*. Despite their extreme popularity in practice, statistical guarantees for projected fixed-point methods, as well as their optimality properties, are not clear. Drawing analogy to non-parametric estimation, for an estimator $\widehat{v}_n$, we seek to establish *oracle inequalities* to characterize its trade-off between approximation and statistical errors.

$$\mathbb{E}\big[\|\widehat{v}_n - v^*\|^2\big] \leq \text{Approximation factor} \times \inf_{v \in \mathcal{F}} \|v - v^*\|^2 + \text{Statistical error}_n(\mathcal{F}).$$

Ideally, we would want a unity approximation factor and an optimal statistical error depending on the localized complexities. Despite decades of efforts, however, such a guarantee is not achieved. The theoretical gap also bewilders practical model and parameter selection. This motivates our central question:

> Can we establish *optimal* oracle inequalities for policy evaluation just as in non-parametric regression?

The answer is “yes and no”: optimal oracle inequalities are established, but is qualitative different, exhibiting a richer spectrum of instance-dependent behavior. In Chapter 2, an instance-dependent approximation factor upper bound is established for projected fixed-point, which can be much larger than unity, but surprisingly, is information-theoretically optimal. In Chapter 3, optimal instance-dependent guarantees are established on the statistical error (under Markovian data), with the optimal sample complexity and efficient stochastic approximation schemes. The next two paragraphs describe these results.

Instance-dependent optimality of the approximation factor: Focusing on a low-dimensional linear subspace $\mathcal{F}$, the seminal work by Tsitsiklis and Van Roy establishes an approximation factor bound for the projected fixed point, laying the foundations of ADP with function approximation. Their bound depends on the effective horizon of the problem, which is usually large in practice. As a result, even when the value function is close to the class $\mathcal{F}$, such an approximation error is amplified by a large factor, leading to potentially poor solution. It is not clear whether this is unavoidable.

We start by showing an instance-dependent approximation factor upper bound for the projected fixed-point approach, which depends on a notion of mixing in the projected space. Such a bound recovers the horizon-based bound in the worst case, and improves existing instance-dependent results.

What is most surprising, though, is that the approximation factor we established is indeed information-theoretically instance-optimal, if we only have access to empirical data. In particular, we show that for a moderate sample size, the local minimax risk for estimation is *lower bounded* by the projection error $\inf_{v \in \mathcal{F}} \|v - v^*\|^2$ multiplied by exactly the same approximation factor as we established in the upper bound. This precisely characterizes the price of RL with function approximation: when the sample size cannot support estimating the entire underlying MDP model, an approximate factor must be paid.

Statistical complexity and stochastic approximation with Markovian data: Turning to the statistical error, we consider the stochastic approximation (SA) scheme in Euclidean space, where the data are from a Markov chain trajectory. This method is known as temporal difference (TD) methods, a default building block in most practical RL systems. Through a novel bootstrapping proof technique, my work establish risk bounds for SA with optimal dependency on the problem dimension and the mixing time of underlying Markov chain. Additionally, for the Polyak–Ruppert averaged iterates, we show an instance-dependent and optimal error upper bound that achieves the exact covariance of Markovian central limit theorem, again with an optimal sample complexity.

Interesting consequences are derived by combining the instance-optimal results Chapters 2 and 3. In RL literature, a classical approach for addressing the trade-off between approximation and statistical error is through a resolvent formalism of the Bellman operator – leading to the class of TD(λ) methods for $\lambda \in (0, 1)$. A long-standing puzzle is the choice of the tuning parameter λ, and our instance-dependent results make it possible to select λ optimally based on empirical estimates.

The main contents of this part is drawn, under minor modification, from the following two papers: Chapter 2 is from the paper [152] "Optimal oracle inequalities for solving projected fixed-point equations, with applications to policy evaluation", co-authored with Ashwin Pananjady and Martin J. Wainwright. Chapter 3 is from the paper [153] "Optimal and instance-dependent guarantees for Markovian linear stochastic approximation", co-authored with Ashwin Pananjady, Martin J. Wainwright, and Peter L. Bartlett.

1.1.2 Part II: off-policy estimation of linear functionals

The contextual bandit model provides a general framework for decision-making. In many applications including causal inference, the learning algorithms are not allowed to interactively explore the environment, and have access only to the observational data generated from a given behavior policy. A central problem is to estimate a linear functional, which covers various notions of treatment effects. This problem exhibits a semi-parametric nature, where the target is a scalar functional of a potentially complicated model.

The celebrated local asymptotic minimax theory by Le Cam, Hájek and Levit provides a canonical measure of instance-dependent optimality with sample size n tending to infinity. Applied to the off-policy estimation problems, the optimal efficiency bound consists of two terms: the variance of the target functional under random state, and the average of conditional variance in the observation re-weighted by the importance ratio. Focusing on achieving the optimal efficiency bounds under different models, various semi-parametric procedures are developed, which lays the foundations of modern theory for causal effect estimation and inference.

Hidden behind the elegant and general asymptotic theory, however, is a diverse spectrum of finite-sample behavior. In semi-parametric methodology, a function is estimated and substituted into an identifying equation so as to obtain an optimal estimator for the scalar. Due to mis-specification, complexity, and local geometry of the function class, the estimation of the nuisance component may dramatically impacts the non-asymptotic error, which are not captured by asymptotic theory of efficiency. Owing to the theoretical gaps, in practice, the design and choices of estimators for the nuisance components lack principled guidance. This motivates us to rethink semi-parametric efficiency from non-asymptotic perspectives, and ask the question:

> What is finite-sample instance-optimality for off-policy estimation? And how to achieve it?

Part of the results in this dissertation pave the way towards a complete recipe for this question. In Chapter 4, we show that the optimal risk is determined by how well we can estimate the nuisance component in a weighted $\mathbb{L}^2$-norm, in addition to the classical efficiency bound. When the efficiency bound itself is infinite, Chapter 5 shows that instance-optimal estimation is made possible by exploiting geometric structures of the function class.

Weighted $\mathbb{L}^2$-norm and optimal finite-sample efficiency: In practice, a class of two-stage semi-parametric procedures known as Augmented Inverse Propensity Weighted (AIPW) are widely used for off-policy estimation. In the first stage, the treatment effect function, defined as the conditional expectation of the outcome conditioned on the state-action pair, is estimated, so as to reduce the variance of the naïve important weight (IPW) method in the second stage. Two prominent questions naturally arise: which estimator should be used in the first stage for optimal estimation of the scalar? and since the estimation of a function can require a large sample size, is this necessary?

Focusing on the case with known behavior policy, we analyze such two-stage procedures, and establish a general finite-sample error upper bound. The risk is given by the sum of the asymptotic efficiency bound, and the estimation error for the nuisance function, measured under a re-weighted $\mathbb{L}^2$-norm. The optimality of such procedure is established in a strong sense, through a local non-asymptotic minimax lower bound containing exactly the same weighted $\mathbb{L}^2$-error term. We also show the necessity of the complexity-dependent sample size for achieving such bounds. Our results therefore exhibits the *equivalence* between optimal estimation for the scalar and the function, under the weighted $\mathbb{L}^2$-norm.

Optimal notion of efficiency beyond the $\sqrt{n}$ rate: The semi-parametric efficiency bound for off-policy evaluation involves certain moments of the importance ratio, which can be infinite even for many natural applications. In causal literature, this corresponds to the lack of overlap assumption, a notorious situation where many classical principles in the semi-parametric efficiency regime fails. However, infinite efficiency bound does not rule out the possibility for estimating the target functional, if structural assumptions are imposed on the treatment effect function.

Assuming that the treatment effect function belongs to a reproducing kernel Hilbert spaces (RKHS), we characterize the instance-optimal risk with non-asymptotic upper bounds and local minimax lower bounds, thereby exhibiting the correct notion of efficiency in such regime. The resulting rate varies from the (near-)parametric $\sqrt{n}$ to arbitrarily slow ones, depending on interplay between the geometry of the RKHS and the behavior policy. Furthermore, the estimator is adaptive in a strong sense: no knowledge about the policy is required, while the optimal risk is achieved under any policy.

The main contents of this part is drawn, under minor modification, from the following two papers: Chapter 4 is from the paper [155] "Off-policy estimation of linear functionals: non-asymptotic theory for semi-parametric efficiency", co-authored with Martin J. Wainwright and Peter L. Bartlett. Chapter 5 is from the paper [145] "Kernel-based off-policy estimation without overlap: instance-dependent optimality beyond semiparametric efficiency", co-authored with Peng Ding, Martin J. Wainwright, and Peter L. Bartlett.

1.2 Related work not appearing in this thesis

In this section, we briefly summarize some other papers of the author during his Ph.D., which are closely related to the topics above.

- **Efficient sampling algorithms and diffusion processes:** High-dimensional sampling is a fundamental computational task with a wide range of applications in data science. In particular, it allows us to compute Bayes optimal estimators efficiently, and quantify uncertainties in a Bayesian framework. Lying at the heart of continuous-space Markov chain Monte Carlo (MCMC) algorithms is the discretization of the Langevin stochastic differential equations (SDE). I have made several contribution to design and analysis of efficient sampling algorithms for large-scale learning and inference. In particular, in the paper [147], we provide a near-optimal analysis for the forward Euler discretization of Langevin diffusion, achieving near-optimal rates, polynomial dependence on the time horizon, and linear dimension dependence simultaneously under mild conditions that allow arbitrary non-convexity. In the papers [151] and [146], we develop sampling algorithms with state-of-the-art non-asymptotic performances by exploiting structures in learning problems. The analysis of Langevin diffusion processes also sheds light on the non-asymptotic contraction behavior and shape of the Bayesian posterior. In my work [148], we develop new tools for non-asymptotic contraction rates using the diffusion approach, which leads to near-optimal rates in some challenging setups.

- **Statistically optimal stochastic approximation algorithms:** The instance-dependent optimality guarantees for stochastic approximation methods presented in Chapter 3 have also been established for other setups. In the paper [150], we study Polyak–Ruppert averaged constant stepsize linear stochastic approximation with i.i.d. data. A fine-grained instance-dependent behavior is characterized, with an additional term in the covariance given by a Ricatti equation depending on the stepsize. In the paper [121], we propose a new stochastic optimization algorithm, ROOT-SGD, which achieves desirable instance-dependent guarantees at a near-optimal sample complexity, in the classical strongly convex and smooth setup. This algorithm is further extended to solve the fixed point of contractive operators in general Banach spaces [149]. The resulting algorithm, ROOT-SA, non-asymptotically achieves the optimal risk functional defined by the local asymptotic minimax limit, with a sharp sample complexity.

- **Reinforcement learning with general function approximation:** Part I of this dissertation focuses on establishing optimal oracle inequalities for policy evaluation using approximation with linear subspaces. While this form of function approximation is widely used in practice, other parametric and nonparametric classes provide more flexibility. In the paper [154], we extend the optimal oracle inequalities to general function classes under some geometric conditions in a tangent cone. We further exhibit the necessity of such geometric conditions. In an independent line of research, in the paper [156], we characterize the sample complexity of policy optimization using the eluder dimension of the policy class.

- **Causal estimation with high-dimensional data:** In observational studies, when the propensity score is unknown, a popular approach is to estimate the propensity score using logistic models. High-dimensional covariates bring about additional challenges in this semiparametric problem. Using a novel debiased procedure, in the paper [197], we establish asymptotic normality and non-asymptotic bounds with an improved dependence on problem dimension.

1.3 Notation

Here we summarize some notation used throughout this dissertation.

Basic notations: For a positive integer m, we define the set $[m] := \{1, 2, \cdots, m\}$. For any pair $(\mathbb{X}, \mathbb{Y})$ of real Hilbert spaces and a linear operator $A : \mathbb{X} \to \mathbb{Y}$, we denote by $A^* : \mathbb{Y} \to \mathbb{X}$ the adjoint operator of A, which by definition, satisfies $\langle Ax, \, y\rangle = \langle x, \, A^* y\rangle$ for all $(x, y) \in \mathbb{X} \times \mathbb{Y}$. For a bounded linear operator A from $\mathbb{X}$ to $\mathbb{Y}$, we define its operator norm as: $\|A\|_{\mathbb{X} \to \mathbb{Y}} := \sup_{x \in \mathbb{X} \setminus \{0\}} \frac{\|Ax\|_{\mathbb{Y}}}{\|x\|_{\mathbb{X}}}$. We use the shorthand notation $\|A\|_{\mathbb{X}}$ to denote its operator norm when A is a bounded linear operator mapping $\mathbb{X}$ to itself. When $\mathbb{X} = \mathbb{R}^{d_1}$ and $\mathbb{Y} = \mathbb{R}^{d_2}$ are finite-dimensional Euclidean spaces equipped with the standard inner product, we denote by $\|A\|_{\mathrm{op}}$ the operator norm in this case. We also use $\|\cdot\|_2$ to denote the standard Euclidean norm, in order to distinguish it from the Hilbert norm $\|\cdot\|$. Given a set A in a normed vector space with norm $\|\cdot\|_c$, we denote the diameter $\mathrm{diam}_c(A) := \sup_{x,y\in A}\|x-y\|_c$.

For a random object X, we use $\mathcal{L}(X)$ to denote its probability law. For any $x \in \mathcal{S}$, we use δ_x to denote the distribution that places all its mass on $\{x\}$, or equivalently, the Dirac δ-function at point x (defined as a tempered distribution). Given a vector $\mu \in \mathbb{R}^d$ and a positive semi-definite matrix $\Sigma \in \mathbb{R}^{d\times d}$, we use $\mathcal{N}(\mu, \Sigma)$ to denote the Gaussian distribution with mean μ and covariance Σ. We use $\mathcal{U}(\Omega)$ to denote the uniform distribution over a set Ω. Given a Polish space $\mathcal{S}$ and a positive measure μ associated to its Borel σ-algebra, for $p \in [1, +\infty)$, we define $\mathbb{L}^p(\mathcal{S}, \mu) := \big\{f : \mathcal{S} \to \mathbb{R},\ \|f\|_{\mathbb{L}^p} := \Big(\int_{\mathcal{S}} |f|^p d\mu\Big)^{1/p} < +\infty\big\}$. When $\mathcal{S}$ is a subset of $\mathbb{R}^d$ and μ is the Lebesgue measure, we use the shorthand notation $\mathbb{L}^p(\mathcal{S})$.

Distance between probability measures: We let $(\mathcal{S}, \rho)$ denote a metric space. For a pair (π, μ) of probability distributions on $\mathcal{S}$, let $\Gamma(\pi, \mu)$ denote the space of all possible couplings of μ and π. For any $p \geq 1$, the Wasserstein-p distance between π and μ is given by

$$\mathcal{W}_p(\pi, \mu) := \Big\{\inf_{\gamma\in\Gamma(\pi,\mu)} \int_{\mathcal{S}\times\mathcal{S}} \rho(x, y)^p d\gamma(x, y)\Big\}^{1/p}, \tag{1.1}$$

and the total variation distance between π and μ by

$$d_{\mathrm{TV}}(\pi, \mu) := \sup_{A\subseteq\mathcal{S}} |\pi(A) - \mu(A)|\,.$$

For any pair of probability distributions P and Q on the same space, we use $P \ll Q$ to denote the fact that P is absolute continuous with respect to Q, and use $\frac{dP}{dQ}$ to indicate the Radon-Nikodym derivative. Given $P \ll Q$, we define:

$$\begin{aligned} \text{KL Divergence:}\quad & D_{\mathrm{KL}}\left(P \parallel Q\right) := \mathbb{E}_P\big[\log \tfrac{dP}{dQ}(X)\big], \\ \chi^2 \text{ divergence:}\quad & \chi^2\left(P \parallel Q\right) := \mathbb{E}_P\big[\tfrac{dP}{dQ}(X) - 1\big], \\ \text{Max divergence:}\quad & D_\infty(P||Q) := \sup_{x\in\mathrm{supp}(Q)} \Big|\log \tfrac{dP}{dQ}(x)\Big|\,. \end{aligned}$$

Matrices in finite dimensions: We use $\{e_j\}_{j=1}^d$ to denote the standard basis vectors in the Euclidean space $\mathbb{R}^d$, i.e., e_i is a vector with a 1 in the i-th coordinate and zeros elsewhere. For two matrices $A \in \mathbb{R}^{d_1\times d_2}$ and $B \in \mathbb{R}^{d_3\times d_4}$, we denote by $A \otimes B$ their Kronecker product, a $d_1d_3 \times d_2d_4$ real matrix. For symmetric matrices $A, B \in \mathbb{R}^{d\times d}$, we use $A \preceq B$ to denote the fact $B - A$ is a positive semi-definite matrix, and denote by $A \prec B$ when $B - A$ is positive definite. For a positive integer d and indices $i, j \in [d]$, we denote by E_{ij} a $d \times d$ matrix with a 1 in the (i, j) position and zeros elsewhere. More generally, given a set $\mathcal{S}$ and $s_1, s_2, \in \mathcal{S}$, we define E_{s_1,s_2} to be the linear operator such that $E_{s_1,s_2} f(x) := f(s_2)\mathbf{1}_{x=s_1}$ for all $f : \mathcal{S} \to \mathbb{R}$.

Given any matrix $A = (a_{ij}) \in \mathbb{R}^{n\times m}$, its vectorization is obtained by concatenating its columns—viz. $\mathrm{vec}(A) := \begin{bmatrix} a_{11} & a_{2,1} & \cdots & a_{n1} & a_{12} & \cdots & a_{n2} & \cdots & a_{1m} & \cdots & a_{nm}\end{bmatrix}^\top \in \mathbb{R}^{nm}$. We use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote the largest and smallest eigenvalue of the matrix A, respectively. We use the following notation for matrix norms: for any matrix $A \in \mathbb{R}^{d_1 \times d_2}$, we use the notation $|\!|\!| A |\!|\!|_{\mathrm{op}}$, $|\!|\!| A |\!|\!|_F$ and $|\!|\!| A |\!|\!|_{\mathrm{nuc}}$ to denote its operator norm, Frobenius norm and nuclear norm, respectively.

Infinite-dimensional matrices: Above definitions can be extended to the infinite-dimensional setup. In particula, we define infinite-dimensional vectors and matrices formally as mappings from integers (pairs) to reals. Given an infinite-dimensional matrix A and a vector z, we define their product pointwise as

$$[Az]_i := \sum_{j=1}^{+\infty} A_{i,j} z_j, \quad \text{for any } i \in \mathbb{N}_+,$$

assuming that each summation is absolute convergent.

Let $\ell_0(\mathbb{N})$ be the set of infinite-dimensional vectors with finite support, i.e., finitely many non-zero entries. We say an infinite-dimensional symmetric matrix A is positive semi-definite, denoted by $A \succeq 0$, if

$$x^\top A x \geq 0, \quad \text{for any } x \in \ell_0(\mathbb{N}).$$

Note that for any vector space $\mathbb{V}$ in which $\ell_0(\mathbb{N})$ is dense, if the matrix A maps from $\mathbb{V}$ to $\mathbb{V}^*$, the definition can be easily extended to ensure that $x^\top A x \geq 0$ for any $x \in \mathbb{V}$. Given this notation, we can furthermore define the positive semi-definite ordering $A \succeq B$ if $A - B \succeq 0$.

Similarly, we can define the inverse of infinite-dimensional matrix. We call $B = A^{-1}$ if $B \cdot (Ax) = A(Bx) = x$ for any $x \in \ell_0(\mathbb{N})$. Once again, such definition can be easily extended to larger vector spaces by density arguments, assuming that both A and B are bounded linear operators acting on suitably defined spaces.

Empirical process tools: For any $\alpha > 0$, the *Orlicz norm* of a scalar random variable X is given by

$$\|X\|_{\psi_\alpha} := \sup \left\{ u > 0 \mid \mathbb{E}\big[e^{(|X|/u)^\alpha}\big] \leq 1 \right\}.$$

The choices $\alpha = 2$ and $\alpha = 1$ correspond, respectively, to the cases of sub-Gaussian and sub-exponential tails, respectively.

Given a metric space $(\mathbb{T}, \rho)$ and a set $\Omega \subseteq \mathbb{T}$, we use $N(\Omega, \rho; s)$ to denote the cardinality of a minimal s-covering of set Ω under the metric ρ. For any scalar $q \geq 1$

and closed interval $[\delta, D]$ we define the Dudley entropy integral

$$\mathcal{J}_q(\Omega, \rho; [\delta, D]) := \int_{\delta}^{D} \big[\log N(\Omega, \rho; s)\big]^{1/q} \, ds.$$

Given a domain $\mathcal{S}$, a bracket $[\ell, u]$ is a pair of real-valued functions on $\mathcal{S}$ such that $\ell(x) \leq u(x)$ for any $x \in \mathcal{S}$, and a function f is said to lie in the bracket $[\ell, u]$ if $f(x) \in [\ell(x), u(x)]$ for any $x \in \mathcal{S}$. Given a probability measure $\mathbb{Q}$ over $\mathcal{S}$, the size of the bracket $[\ell, u]$ is defined as $\|u - \ell\|_{\mathbb{L}^2(\mathbb{Q})}$. For a function class $\mathcal{F}$ over $\mathcal{S}$, the bracketing number $N_{\text{bra}}\big(\mathcal{F}, \mathbb{L}^2(\mathbb{Q}); s\big)$ denotes the cardinality of a minimal bracket covering of the set $\mathcal{F}$, with each bracket of size smaller than s. Given a closed interval $[\delta, D]$, the bracketed chaining integral is given by

$$\mathcal{J}_{\text{bra}}(\mathcal{F}, \mathbb{L}^2(\mathbb{Q}); [\delta, D]) := \int_{\delta}^{D} \sqrt{\log N_{\text{bra}}(\mathcal{F}, \mathbb{L}^2(\mathbb{Q}); s)} \, ds.$$

Part I

Reinforcement learning with function approximation

Chapter 2

Optimal oracle inequalities for projected fixed-point equations

Linear fixed point equations in Hilbert spaces arise in a variety of settings, including reinforcement learning, and computational methods for solving differential and integral equations. In this chapter, we study methods that use a collection of random observations to compute approximate solutions by searching over a known low-dimensional subspace of the Hilbert space. First, we prove an instance-dependent upper bound on the mean-squared error for a linear stochastic approximation scheme that exploits Polyak–Ruppert averaging. This bound consists of two terms: an approximation error term with an instance-dependent approximation factor, and a statistical error term that captures the instance-specific complexity of the noise when projected onto the low-dimensional subspace. Using information-theoretic methods, we also establish lower bounds showing that both of these terms cannot be improved, again in an instance-dependent sense. A concrete consequence of our characterization is that the optimal approximation factor in this problem can be much larger than a universal constant. We show how our results precisely characterize the error of a class of temporal difference learning methods for the policy evaluation problem with linear function approximation, establishing their optimality.

2.1 Introduction

Linear fixed point equations over a Hilbert space, with the Euclidean space being an important special case, arise in various contexts. Such fixed point equations take different names in different domains, including estimating equations, Bellman equations, Poisson equations and inverse systems [15, 110, 221]. More specifically, given a Hilbert space $\mathbb{X}$, we consider a fixed point equation of the form

$$v = Lv + b, \tag{2.1}$$

where b is some member of the Hilbert space, and L is a linear operator mapping $\mathbb{X}$ to itself.

When the Hilbert space is infinite-dimensional—or has a finite but very large dimension D—it is common to seek approximate solutions to equation (2.1). A standard approach is to choose a subspace $\mathbb{S}$ of the Hilbert space, of dimension $d \ll D$, and to search for solutions within this subspace. In particular, letting $\Pi_{\mathbb{S}}$ denote the orthogonal projection onto this subspace, various methods seek (approximate) solutions to the *projected fixed point equation*

$$v = \Pi_{\mathbb{S}}\big(Lv + b\big). \tag{2.2}$$

In order to set the stage, let us consider some generic examples that illustrate the projected fixed point equation (2.2). We eschew a fully rigorous exposition at this stage, deferring technical details and specific examples to Section 2.2.2.

Example 2.1 (Galerkin methods for differential equations)**.** *Let $\mathbb{X}$ be a Hilbert space of suitably differentiable functions, and let A be a linear differential operator of order k, say of the form $A(v) = \omega_0 v + \sum_{j=1}^{k} \omega_j v^{(j)}$, where $v^{(j)}$ denotes the j^{th}-order derivative of the function $v \in \mathbb{X}$. Given a function $b \in \mathbb{X}$, suppose that we are interested in solving the differential equation $A(v) = b$. This represents a particular case of our fixed point equation with $L = I - A$.*

Let $\mathbb{S}$ be a finite-dimensional subspace of $\mathbb{X}$, say spanned by a set of basis functions $\{\phi_j\}_{j=1}^{d}$. A Galerkin method constructs an approximate solution to the differential equation $A(v) = b$ by solving the projected fixed point equation (2.2) over a subspace of this type. Concretely, any function $v \in \mathbb{S}$ has a representation of the form $v = \sum_{j=1}^{d} \vartheta_j \phi_j$ for some weight vector $\vartheta \in \mathbb{R}^d$. Applying the operator A to any such function yields the residual $A(v) = \sum_{j=1}^{d} \vartheta_j A(\phi_j)$, and the Galerkin method chooses the weight vector $\vartheta \in \mathbb{R}^d$ such that v satisfies the equation $v = \Pi_{\mathbb{S}}((I - A)v + b)$. In Section 2.2.2.2, we describe in detail a specific version of the Galerkin method as applied to a second-order differential equation that underlies the so-called elliptic boundary value problem. ♣

Example 2.2 (Instrumental variable methods for nonparametric regression)**.** *Let $\mathbb{X}$ denote a suitably constrained space of square-integrable functions mapping $\mathbb{R}^p \to \mathbb{R}$, and suppose that we have a regression model of the form $Y = f^*(X) + \epsilon$. Here X is a random vector of covariates taking values in $\mathbb{R}^p$, the pair (Y, ϵ) denote scalar random variables, and $f^* \in \mathbb{X}$ denotes an unknown function of interest. For discussion of the existence and uniqueness of the various objects in this model, see Darolles et al. [46].*

In the classical setup of nonparametric regression, it is assumed that $\mathbb{E}[\epsilon \mid X] = 0$, an assumption that can be violated. Instead, suppose that we have a vector of instrumental variables *$Z \in \mathbb{R}^p$ such that $\mathbb{E}[\epsilon \mid Z] = 0$. Now let $T : \mathbb{X} \to \mathbb{X}$ denote a linear operator given by $T(f) = \mathbb{E}[f(X)|Z]$, and denote by $r = \mathbb{E}[Y|Z]$ a point in $\mathbb{X}$. Instrumental variable (IV) approaches to estimating f^* are based on the equality*

$$\mathbb{E}[Y - f^*(X) \mid Z] \;=\; r - T(f^*) = 0, \tag{2.3}$$

which is a linear fixed point relation of the form (2.1) with $L = I - T$ and $b = r$.

Now let $\{\phi_j\}_{j\geq 1}$ be an orthonormal basis of $\mathbb{X}$, and let $\mathbb{S}$ denote the subspace spanned by the first d such eigenfunctions. Then each function $f \in \mathbb{S}$ can be represented as $f = \sum_{j=1}^{d} \vartheta_j \phi_j$, and approximate solutions to the fixed point equation (2.3) may be obtained via solving a projected variant (2.2), i.e., the equation $f = \Pi_{\mathbb{S}}((I-T)f + r)$.

A specific example of an IV method is the class of temporal difference methods for policy evaluation, introduced and discussed in detail in Section 2.2.2.3. ♣

In particular instantiations of both of the examples above, it is typical for the ambient dimension D to be very large (if not infinite) and for us to only have sample access to the pair (L, b). This chapter treats the setting in which n observations $\{(L_i, b_i)\}_{i=1}^{n}$ are drawn i.i.d. from some distribution with mean (L, b). Letting v^* denote the solution to the fixed point equation (2.1), our goal is to use these observations in order to produce an estimate $\widehat{v}_n$ of v^* that satisfies an *oracle inequality* of the form

$$\mathbb{E}\|\widehat{v}_n - v^*\|^2 \leq \alpha \cdot \inf_{v \in \mathbb{S}} \|v - v^*\|^2 + \varepsilon_n. \tag{2.4}$$

Here we use $\|\cdot\|$ to denote the Hilbert norm associated with $\mathbb{X}$. The three terms appearing on the RHS of inequality (2.4) all have concrete interpretations. The term

$$\mathcal{A}(\mathbb{S}, v^*) := \inf_{v \in \mathbb{S}} \|v - v^*\|^2 \tag{2.5}$$

defines the *approximation error*; this is the error incurred by an oracle procedure that knows the fixed point v^* in advance and aims to output the best approximation to v^* within the subspace $\mathbb{S}$. The term α is the *approximation factor*, which indicates how poorly the estimator $\widehat{v}_n$ performs at carrying out the aforementioned approximation; note that $\alpha \geq 1$ by definition, and it is most desirable for α to be as small as possible. The final term ε_n is a proxy for the *statistical error* incurred due to our stochastic observation model; indeed, one expects that as the sample size n goes to infinity, this error should tend to zero for any reasonable estimator, indicating consistent estimation when $v^* \in \mathbb{S}$. More generally, we would like our estimator to also have as small a statistical error as possible in terms of the other parameters that define the problem instance.

In an ideal world, both desiderata hold simultaneously: the approximation factor should be as close to one as possible while the statistical error stays as small as possible. As we discuss shortly, such a “best-of-both-worlds” guarantee can indeed be obtained in many canonical problems, and “sharp” oracle inequalities—meaning ones in which the approximation factor is equal to one—are known [172, 44]. On the other hand, such oracle equalities with unit factors are not known for the fixed point equation (2.1). Tsitsiklis and Van Roy [204] show that if the operator L is $\gamma_{\max}$-contractive in the norm $\|\cdot\|$, then the (deterministic) solution $\bar{v}$ to the projected fixed point equation (2.2) satisfies the bound

$$\|\bar{v} - v^*\|^2 \leq \frac{1}{1-\gamma_{\max}^2} \inf_{v \in \mathbb{S}} \|v - v^*\|^2. \tag{2.6}$$

Since $\gamma_{\max}$ can be arbitrarily close to one, the pre-factor in the bound (2.6) can be much larger than one, in contrast to so-called "sharp" oracle inequalities for non-parametric regression. One motivating question for our work is whether or not this bound can be improved, and if so, to what extent.[1]

Our work is also driven by the complementary question of whether a sharp bound can be obtained on the statistical error of an estimator that, unlike $\bar{v}$, has access only to the samples $\Big\{(L_i, b_i)\Big\}_{i=1}^{n}$. In particular, we would like the statistical error ε_n to depend on some notion of complexity within the subspace $\mathbb{S}$, and *not* on the ambient space. Recent work by Bhandari et al. [16] provides worst-case bounds on the statistical error of a stochastic approximation scheme, showing that the *parametric rate* $\epsilon_n \lesssim d/n$ is attainable. In this chapter, we study how to derive a more fine-grained bound on the statistical error that reflects the practical performance of the algorithm and depends optimally on the geometry of our problem instance.

2.1.1 Contributions and organization

The main contribution of this chapter is to resolve both of the aforementioned questions, in particular by deriving upper bounds and information-theoretic lower bounds on both the approximation factor and statistical error that are *instance-dependent*. On one hand, these bounds demonstrate that in general, it is not possible to obtain an oracle inequality with a pre-factor equal to one, but that there are many settings in which the optimal approximation factor is much smaller than what is suggested by the worst-case bound (2.6). We also derive a significantly sharper bound on the statistical error of a stochastic approximation scheme that is instance-optimal in a precise sense. In more detail, the contributions of this chapter include the following:

- Theorem 2.1 establishes an instance-dependent upper bound of the form (2.4) for the Polyak–Ruppert averaged stochastic approximation estimator, whose approximation factor α depends in a precise way on the projection of the operator L onto the subspace $\mathbb{S}$, and the statistical error ϵ_n matches the Cramér–Rao lower bound for the instance within the subspace.

- In Theorem 2.2, we prove an information-theoretic lower bound on the approximation factor. It is a local analysis, in that the bound depends critically on the projection of the population-level operator. This lower bound certifies that the approximation factor attained by our estimator is optimal. To the best of our knowledge, this is also the first instance of an optimal oracle inequality with a non-constant and problem-dependent approximation factor.

[1] Note that one can achieve an approximation factor arbitrarily close to one provided that $n \gg D$. One way to do so is as follows: form the plug-in estimate that solves the original fixed point relation (2.1) on the sample averages $\frac{1}{n}\sum_{i=1}^{n} L_i$ and $\frac{1}{n}\sum_{i=1}^{n} b_i$, and then project this solution onto the subspace $\mathbb{S}$. In this chapter, our principal interest—driven by the practical examples of Galerkin approximation and temporal difference learning—is in the regime $d \ll n \ll D$.

- In Theorem 2.3, we establish via a Bayesian Cramér-Rao lower bound that the leading statistical error term for our estimator is also optimal in an instance-dependent sense.
- In Section 2.4, we derive specific consequences of our results for several examples, including problem of Galerkin approximation in second-order elliptic equations, as well as temporal difference methods for policy evaluation with linear function approximation. A particular consequence of our results shows that in a minimax sense, the approximation factor (2.6) is optimal for policy evaluation with linear function approximation (cf. Proposition 2.1).

The remainder of this chapter is organized as follows. Section 2.1.2 contains a detailed discussion of related work. We introduce formal background and specific examples in Section 2.2. Our main results under the general model of projected fixed point equations are introduced and discussed in Section 2.3. We then specialize these results to our examples in Section 2.4, deriving several concrete corollaries for Galerkin methods and temporal difference methods. Our proofs are postponed to Section 2.5, and technical results are deferred to the appendix.

2.1.2 Related work

Our study touches on various lines of related work, including stochastic approximation and its application to reinforcement learning, projected linear equation methods, as well as oracle inequalities for statistical estimation. Let us provide a brief discussion of these connections here.

Stochastic approximation: Stochastic approximation algorithms for both linear and nonlinear fixed-point equations play a central role in large-scale machine learning and statistics [174, 114, 161]. See the books [11, 20] for a comprehensive survey of the classical methods of analysis. In seminal work due to Polyak, Ruppert, and Juditsky [168, 169, 186], it was proposed to take the average of the stochastic approximation iterates, which stabilizes the algorithm and ensures a Gaussian limiting distribution. In fact, the averaged iterates are known to be asymptotically optimal in a local minimax sense [55]. Non-asymptotic guarantees matching this asymptotic behavior have also been established for other forms of stochastic approximation, as well as variance-reduced variants thereof [157, 98, 150, 121].

Stochastic approximation is also a fundamental building block for reinforcement learning algorithms, wherein the method is used to produce an iterative, on-line solution to the Bellman equation from data; see the books [200, 14] for a survey. Such approaches include temporal difference (TD) methods [198] for the policy evaluation problem and the Q-learning algorithm [216] for policy optimization. Variants of these algorithms also abound, including LSTD [24], SARSA [185], actor-critic algorithms [107], and gradient TD methods [199]. The analysis of these methods has received significant attention in the literature, ranging from asymptotic guarantees (e.g., [27, 204, 205]) to more fine-grained finite-sample bounds (e.g., [16, 194, 115, 164, 211, 212]). Our work contributes to this literature, since as a corollary of our general analysis, we are able to establish finite-sample upper bounds for temporal difference methods with Polyak–Ruppert averaging, as applied to the policy evaluation problem with linear function approximation.

Projected methods for linear equations: In 1915, Galerkin [65] first proposed the method of approximating the solution to a linear PDE by solving the projected equation in a finite-dimensional subspace. This method later became a cornerstone of finite-element methods in numerical methods for PDEs; see the books [60, 28] for a comprehensive survey. A fundamental tool used in the analysis of Galerkin methods is Céa's lemma [35]; in this chapter, we derive more general upper bounds on the approximation factor that capture this classical lemma as a special case. As mentioned before, in the specific context of reinforcement learning, projected linear equations were studied by Tsitsiklis and Van Roy [204], who first proved the upper bound (2.6) on the approximation factor under contractivity assumptions. These contraction-based bounds were further extended to the analysis of Q-learning in optimal stopping problems [205]. The connection between the Galerkin method and TD methods was observed by Yu and Bertsekas [226, 15], and the former paper provides an instance-dependent upper bound on the approximation factor. This analysis was later applied to Monte–Carlo methods for solving linear inverse problems [171, 170].

The Bellman equation can be written in infinitely many equivalent ways—by using powers of the transition kernel and via the formalism of resolvents—leading to a continuous family of projected equations indexed by a scalar parameter λ (see, e.g., Section 5.5 of Bertsekas [14]). Some of these forms can be specifically leveraged in other observation models; for instance, by observing the trajectory of the Markov chain instead of i.i.d. samples, it becomes possible to obtain unbiased observations for integer powers of the transition kernel. This makes it possible to efficiently estimate the solution to the projected linear equation for various values of λ, and underlies the family of TD(λ) methods [198, 24]. Indeed, Tsitsiklis and Van Roy [204] also showed that the worst-case approximation factor in equation (2.6) can be improved by using larger values of λ. Based on this observation, a line of work has studied the trade-off between approximation error and estimation measure in model selection for reinforcement learning problems [13, 187, 158, 210]. Understanding precise trade-offs between approximation and estimation error is crucial to model selection. However, unlike this body of work, our focus in this chapter is on studying the i.i.d. observation model; a detailed investigation into the Markov setting will be presented in Chapter 3.

Oracle inequalities: There is a large literature on mis-specified statistical models and oracle inequalities (e.g., see the monographs [136, 105] for overviews). Oracle inequalities in the context of penalized empirical risk minimization (ERM) are quite well-understood (e.g., [8, 104, 137]). Typically, the resulting approximation factor is exactly 1 or arbitrarily close to 1, and the statistical error term depends on the localized Rademacher complexity or metric entropy of this function class. Aggregation methods have been developed in order to obtain *sharp* oracle inequalities with approximation factor exactly 1 (e.g. [207, 31, 44, 172]). Sharp oracle inequalities are now available in a variety of settings including for sparse linear models [32], density estimation [45], graphon estimation [102], and shape-constrained estimation [10]. As previously noted, our setting differs qualitatively from the ERM setting, in that as shown in this chapter, sharp oracle inequalities are no longer possible. There is another related line of work on oracle inequalities of density estimation. Yatracos [224] showed an oracle inequality with the non-standard approximation factor 3, and with a statistical error term depending on the metric entropy. This non-unit approximation factor was later shown to be optimal for the class of one-dimensional piecewise constant densities [36, 23, 233]. The approximation factor lower bound in these papers and our work both make use of the birthday paradox to establish information-theoretic lower bounds.

2.2 Background

We begin by formulating the projected fixed point problem more precisely in Section 2.2.1. Section 2.2.2 provides illustrations of this general set-up with some concrete examples.

2.2.1 Problem formulation

Consider a separable Hilbert space $\mathbb{X}$ with (possibly infinite) dimension D, equipped with the inner product $\langle \cdot, \cdot \rangle$. Let $\mathfrak{L}$ denote the set of all bounded linear operators mapping $\mathbb{X}$ to itself. Given one such operator $L \in \mathfrak{L}$ and some $b \in \mathbb{X}$, we consider the fixed point relation $v = Lv + b$, as previously defined in equation (2.1). We assume that the operator $I - L$ has a bounded inverse, which guarantees the existence and uniqueness of the fixed point satisfying equation (2.1). We let v^* denote this unique solution.

As previously noted, in general, solving a fixed point equation in the Hilbert space can be computationally challenging. Consequently, a natural approach is to seek approximations to the fixed point v^* based on searching over a finite-dimensional subspace of the full Hilbert space. More precisely, given some d-dimensional subspace $\mathbb{S}$ of $\mathbb{X}$, we seek to solve the projected fixed point equation (2.2).

Existence and uniqueness of projected fixed point: For concreteness in analysis, we are interested in problems for which the projected fixed equation has a unique solution. Here we provide a sufficient condition for such existence and uniqueness. In doing so and for future reference, it is helpful to define some mappings between $\mathbb{X}$ and the subspace $\mathbb{S}$. Let us fix some orthogonal basis $\{\phi_j\}_{j \geq 1}$ of the full space $\mathbb{X}$ such that $\mathbb{S} = \text{span}\{\phi_1, \ldots, \phi_d\}$. In terms of this basis, we can define the projection operator

$\Phi_d : \mathbb{X} \to \mathbb{R}^d$ via $\Phi_d(x) := \big(\langle x,\, \phi_j \rangle\big)_{j=1}^{d}$. The adjoint operator of Φ_d is a mapping from $\mathbb{R}^d$ to $\mathbb{X}$, given by

$$\Phi_d(v) := \sum_{j=1}^{d} v_j \phi_j. \tag{2.7}$$

Using these operators, we can define the *projected operator* associated with L—namely

$$M := \Phi_d L \Phi_d^*. \tag{2.8}$$

Note that M is simply a d-dimensional matrix, one which describes the action of L on $\mathbb{S}$ according to the basis that we have chosen. As we will see in the main theorems, our results do not depend on the specific choice of the orthonormal basis, but it is convenient to use a given one, as we have done here.

Consider the quantity

$$\kappa(M) := \tfrac{1}{2}\lambda_{\max}\Big(M + M^\top\Big), \tag{2.9}$$

corresponding to the maximal eigenvalue of the symmetrized version of M. One sufficient condition for there be a unique solution to the fixed point equation (2.2) is the bound $\kappa(M) < 1$. When this bound holds, the matrix $(I_d - M)$ is invertible, and hence for any $b \in \mathbb{X}$, there is a unique solution $\bar{v}$ to the equation $v = \Pi_{\mathbb{S}}(Lv + b)$.

Stochastic observation model: As noted in Section 2.1, this chapter focuses on an observation model in which we observe i.i.d. random pairs (L_i, b_i) for $i = 1, \ldots, n$ that are unbiased estimates of the pair (L, b) so that

$$\mathbb{E}[L_i] = L, \quad \text{and} \quad \mathbb{E}[b_i] = b. \tag{2.10}$$

In addition to this unbiasedness, we also assume that our observations satisfy a certain second-moment bound. A weaker and a stronger version of this assumption are both considered.

Assumption 2.1(W). *(Second-moment bound in projected space) There exist scalars $\sigma_L, \sigma_b > 0$ such that for any unit-norm vector $u \in \mathbb{S}$ and any basis vector in $\{\phi_j\}_{j=1}^d$ we have the bounds*

$$\mathbb{E}\langle \phi_j,\, (L_i - L)u \rangle^2 \leq \sigma_L^2 \|u\|^2, \quad \textit{and} \tag{2.11a}$$

$$\mathbb{E}\langle \phi_j,\, b_i - b \rangle^2 \leq \sigma_b^2. \tag{2.11b}$$

Assumption 2.1(S). *(Second-moment bound in ambient space) There exist scalars $\sigma_L, \sigma_b > 0$ such that for any unit-norm vector $u \in \mathbb{X}$ and any basis vector in $\{\phi_j\}_{j=1}^D$ we have the bounds*

$$\mathbb{E}\langle \phi_j,\, (L_i - L)u \rangle^2 \leq \sigma_L^2 \|u\|^2, \quad \textit{and} \tag{2.12a}$$

$$\mathbb{E}\langle \phi_j,\, b_i - b \rangle^2 \leq \sigma_b^2. \tag{2.12b}$$

In words, Assumption 2.1(W) guarantees that the random variable obtained by projecting the "noise" onto any of the basis vectors $\phi_1, \ldots, \phi_d$ in the subspace $\mathbb{S}$ has bounded second moment. Assumption 2.1(S) further requires the projected noise onto any basis vector of the entire space $\mathbb{X}$ to have bounded second moment. In Section 2.4, we show that there are various settings—including Galerkin methods and temporal difference methods—for which at least one of these assumptions is satisfied.

2.2.2 Examples

We now present some concrete examples to illustrate our general formulation. In particular, we discuss the problems of linear regression, temporal difference learning methods from reinforcement learning[2], and Galerkin methods for solving partial differential equations.

2.2.2.1 Linear regression on a low-dimensional subspace

Our first example is the linear regression model when true parameter is known to lie approximately in a low-dimensional subspace. This example, while rather simple, provides a useful pedagogical starting point for the others to follow.

For this example, the underlying Hilbert space $\mathbb{X}$ from our general formulation is simply the Euclidean space $\mathbb{R}^D$, equipped with the standard inner product $\langle \cdot, \cdot \rangle$. We consider zero-mean covariates $X \in \mathbb{R}^D$ and a response $Y \in \mathbb{R}$, and our goal is to estimate the best-fitting linear model $x \mapsto \langle v, \, x \rangle$. In particular, the mean-square optimal fit is given by $v^* := \arg\min_{v \in \mathbb{R}^D} \mathbb{E}\big(Y - \langle v, \, X \rangle\big)^2$. From standard results on linear regression, this vector must satisfy the normal equations $\mathbb{E}[XX^\top] v^* = \mathbb{E}[YX]$. We assume that the second-moment matrix $\mathbb{E}[XX^\top]$ is non-singular, so that v^* is unique.

Let us rewrite the normal equations in a form consistent with our problem formulation. An equivalent definition of v^* is in terms of the fixed point relation

$$v^* = \Big(I - \frac{1}{\beta} \mathbb{E}[XX^\top] \Big) v^* + \frac{1}{\beta} \mathbb{E}[YX], \tag{2.13}$$

where $\beta := \lambda_{\max}(\mathbb{E}[XX^\top])$ is the maximum eigenvalue. This fixed point condition is a special case of our general equation (2.1) with the operator $L = I - \frac{1}{\beta}\mathbb{E}[XX^\top]$ and vector $b = \frac{1}{\beta}\mathbb{E}[YX]$. Note that we have

$$|\!|\!| L |\!|\!|_{\mathrm{op}} = |\!|\!| I - \frac{1}{\beta}\mathbb{E}[XX^\top] |\!|\!|_{\mathrm{op}} \leq 1 - \frac{\mu}{\beta} < 1,$$

where $\mu = \lambda_{\min}(\mathbb{E}[XX^\top]) > 0$ is the minimum eigenvalue of the covariance matrix.

[2] As noted by Bradtke and Barto [27], this method can be understood as an instrumental variable method [221], and our results also apply to this more general setting.

In the well-specified setting of linear regression, we observe i.i.d. pairs $(X_i, Y_i) \in \mathbb{R}^D \times \mathbb{R}$ that are linked by the standard linear model

$$Y_i = \langle v^*,\, X_i \rangle + \varepsilon_i, \quad \text{for } i = 1, 2, \cdots, n, \tag{2.14}$$

where ε_i denotes zero-mean noise with finite second moment. Each such observation can be used to form the matrix-vector pair

$$L_i = I - \beta^{-1} X_i X_i^\top, \quad \text{and} \quad b_i = \beta^{-1} X_i Y_i,$$

which is in the form of our assumed observation model.

Thus far, we have simply reformulated linear regression as a fixed point problem. In order to bring in the projected aspect of the problem, let us suppose that the ambient dimension D is much larger than the sample size n, but that we have the prior knowledge that v^* lies (approximately) within a known subspace $\mathbb{S}$ of $\mathbb{R}^D$, say of dimension $d \ll D$. Our goal is then to approximate the solution to the associated projected fixed-point equation.

Using $\{\phi_j\}_{j=1}^d$ to denote an orthonormal basis of $\mathbb{S}$, the population-level projected linear equation (2.2) in this case takes the form

$$\mathbb{E}\Big[(\Pi_{\mathbb{S}} X)(\Pi_{\mathbb{S}} X)^\top\Big]\bar{v} = \mathbb{E}\Big[Y \cdot \Pi_{\mathbb{S}} X\Big], \tag{2.15}$$

Thus, the population-level projected problem (2.15) corresponds to performing linear regression using the projected version of the covariates, thereby obtaining a vector of weights $\bar{v} \in \mathbb{S}$ in this low-dimensional space.

2.2.2.2 Galerkin methods for second-order elliptic equations

We now turn to the Galerkin method for solving differential equations, a technique briefly described in Section 2.1. The general problem is to compute an approximate solution to a partial differential equation based on a limited number of noisy observations for the coefficients. Stochastic inverse problems of this type arise in various scientific and engineering applications [162, 5].

For concreteness, we consider a second-order elliptic equation with Dirichlet boundary conditions.[3] Given a bounded, connected and open set $\Omega \subseteq \mathbb{R}^m$ with unit Lebesgue measure, let $\partial\Omega$ denote its boundary. Consider the Hilbert space of functions

$$\mathbb{X} := \dot{\mathbb{H}}^1(\Omega) = \Big\{ v : \Omega \to \mathbb{R},\ \int_\Omega \|\nabla v(x)\|_2^2 dx < \infty,\ v|_{\partial\Omega} = 0 \Big\}$$

equipped with the inner product $\langle u,\, v\rangle_{\dot{\mathbb{H}}^1} := \int_\Omega \nabla u(x)^\top \nabla v(x) dx$.

[3] It should be noted that Galerkin methods apply to a broader class of problems, including linear PDEs of parabolic and hyperbolic type [116], as well as kernel integral equations [171, 170].

Given a symmetric matrix-valued function a and a square-integrable function $f \in \mathbb{L}^2$, the *boundary-value problem* is to find a function $v : \Omega \to \mathbb{R}$ such that

$$\begin{cases} \nabla \cdot (a(x)\nabla v(x)) + f = 0 & \text{in } \Omega, \\ v(x) = 0 & \text{on } \partial\Omega. \end{cases} \tag{2.16}$$

We impose a form of uniform ellipticity by requiring that $\mu I_m \preceq a(x) \preceq \beta I_m$, for some positive scalars $\mu \leq \beta$, valid uniformly over x.

The problem can be equivalently stated in terms of the elliptic operator $A := -\nabla \cdot (a\nabla)$; as shown in Appendix A.8.3.1, the pair (A, f) *induces* a bounded, self-adjoint linear operator $\widetilde{A}$ on $\mathbb{X}$ and a function $g \in \mathbb{X}$ such that the solution to the boundary value problem can be written as

$$v^* = \Big(I - \frac{1}{\beta}\widetilde{A}\Big)v^* + \beta^{-1}g. \tag{2.17}$$

By construction, this is now an instance of our general fixed point equation (2.1) with $L := I - \frac{1}{\beta}\widetilde{A}$ and $b := \beta^{-1}g$. Furthermore, our assumptions imply that $\|L\|_{\mathbb{X}} \leq 1 - \frac{\mu}{\beta}$.

We consider a stochastic observation model that is standard in the literature (see, e.g., the paper [69]). Independently for each $i \in [n]$, let W_i denote an $m \times m$ symmetric random matrix with entries on the diagonal and upper-diagonal given by i.i.d. standard Gaussian random variables. Let $w_i' \sim \mathcal{N}(0, 1)$ denote a standard Gaussian random variable. Suppose now that we observe the pair $x_i, y_i \sim \mathcal{U}(\Omega)$; the observed values for the i-th sample are then given by

$$(a_i, f_i) := \big(a(x_i) + W_i, f(y_i) + w_i'\big) \quad \text{with} \quad x_i, y_i \sim \mathcal{U}(\Omega). \tag{2.18}$$

The unbiased observations (L_i, b_i) can then be constructed by replacing (a, f) with $\big(a_i\delta_{x_i}, f_i\delta_{y_i}\big)$ in the constructions above.

For such problems, the finite-dimensional projection not only serves as a fast and cheap way to compute solutions from simulation [130], but also makes the solution stable and robust to noise [91]. Given a finite-dimensional linear subspace $\mathbb{S} \subseteq \mathbb{X}$ spanned by orthogonal basis functions $(\phi_i)_{i=1}^d$, we consider the projected version of equation (2.17), with solution denoted by $\bar{v}$:

$$\bar{v} = \Pi_{\mathbb{S}}(L\bar{v} + b). \tag{2.19}$$

Straightforward calculation in conjunction with Lemma A.9 shows that equation (2.19) is equivalent to the conditions $\bar{v} \in \mathbb{S}$, and

$$\langle \widetilde{A}\bar{v},\, \phi_j\rangle_{\dot{\mathbb{H}}^1} = \langle g,\, \phi_j\rangle_{\dot{\mathbb{H}}^1} \quad \text{for all } j \in [d], \tag{2.20}$$

with the latter equality better known as the *Galerkin orthogonality condition* in the literature [28].

2.2.2.3 Temporal difference methods for policy evaluation

Our final example involves the policy evaluation problem in reinforcement learning. This is a special case of an instrumental variable method, as briefly introduced in Section 2.1. We require some additional terminology to describe the problem of policy evaluation. Consider a Markov chain on a state space $\mathcal{S}$ and a transition kernel $P : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$. It becomes a discounted Markov reward process when we introduce a reward function $r : \mathcal{S} \to \mathbb{R}$, and discount factor $\gamma \in (0, 1)$. The goal of the policy evaluation problem to estimate the value function, which is the expected, long-term, discounted reward accrued by running the process. The value function exists under mild assumptions such as boundedness of the reward, and is given by the solution to the Bellman equation $v^* = \gamma P v^* + r$, which is a fixed point equation of the form (2.1) with $L = \gamma P$ and $b = r$.

Throughout our discussion, we assume that the transition kernel P is ergodic and aperiodic, so that its stationary distribution ξ is unique. We define $\mathbb{X}$ to be the Hilbert space $\mathbb{L}^2(\mathcal{S}, \xi)$, and for any pair of vectors $v, v' \in \mathbb{X}$, we define the inner product as follows

$$\langle v,\, v' \rangle := \int_{\mathcal{S}} v(s) v'(s) d\xi(s).$$

In the special case of a finite state space, the Hilbert space $\mathbb{X}$ is a finite-dimensional Euclidean space with dimension $D = |\mathcal{S}|$ and equipped with a weighted ℓ_2-norm.

We consider the i.i.d. observation model in this chapter. For each $i = 1, 2, \cdots, n$, suppose that we observe an independent tuple $(s_i, s_i^+, R_i(s_i))$, such that

$$s_i \sim \xi, \;\; s_i^+ \sim P(s_i, \cdot), \;\text{ and } \mathbb{E}[R_i(s_i) | s_i] = r(s_i). \tag{2.21}$$

The i-th observation (L_i, b_i) is then obtained by plugging in these observations to compute unbiased estimates of P and r, respectively.

A common practice in reinforcement learning is to employ *function approximation*, which in its simplest form involves solving a projected linear equation on a subspace. In particular, consider a set $\{\psi_1, \psi_2, \cdots, \psi_d\}$ of basis functions in $\mathbb{X}$, and suppose that they are linearly independent on the support of ξ. We are interested in projections onto the subspace $\mathbb{S} = \text{span}(\psi_1, \ldots, \psi_d)$, and in solving the population-level projected fixed point equation (2.2), which takes the form

$$\bar{v} = \Pi_{\mathbb{S}}(\gamma P \bar{v} + r). \tag{2.22}$$

The basis functions ψ_i are not necessarily orthogonal, and it is common for the projection operation to be carried out in a somewhat non-standard fashion. In order to describe this, it is convenient to write equation (2.22) in the projected space. For each $s \in \mathcal{S}$, let $\psi(s) = [\psi_1(s) \;\; \psi_2(s) \;\; \ldots \;\; \psi_d(s)]$ denote a vector in $\mathbb{R}^d$, and note that we may write $\bar{v}(s) = \psi(s)^\top \bar{\vartheta}$ for a vector of coefficients $\bar{\vartheta} \in \mathbb{R}^d$. Now observe that equation (2.22) can be equivalently written in terms of the coefficient vector $\bar{\vartheta}$ as

$$\mathbb{E}_{s \sim \xi}[\psi(s)\psi(s)^\top]\bar{\vartheta} = \gamma \mathbb{E}_{s \sim \xi}\Big[\mathbb{E}_{s^+ \sim P(s,\cdot)}[\psi(s)\psi(s^+)^\top]\Big]\bar{\vartheta} + \mathbb{E}_{s \sim \xi}[r(s)\psi(s)]. \tag{2.23}$$

Equation (2.23) is the population relation underlying the canonical *least squares temporal difference* (LSTD) learning method [27, 24].

2.3 Main results for general projected linear equations

Having set-up the problem and illustrated it with some examples, we now turn to the statements of our main results. We begin in Section 2.3.1 by stating an upper bound on the mean-squared error of a stochastic approximation scheme that uses Polyak–Ruppert averaging. We then discuss the form of this upper bound for various classes of operator L, with a specific focus on producing transparent bounds on the approximation factor. Section 2.3.2 is devoted to information-theoretic lower bounds that establish the sharpness of our upper bound.

2.3.1 Upper bounds

In this section, we describe a standard stochastic approximation scheme for the problem based on combining ordinary stochastic updates with Polyak–Ruppert averaging [168, 169, 186]. In particular, given an oracle that provides observations (L_i, b_i), consider the stochastic recursion parameterized by a positive stepsize η:

$$v_{t+1} = (1-\eta) v_t + \eta \Pi_{\mathbb{S}} \big(L_{t+1} v_t + b_{t+1} \big), \quad \text{for } t = 1, 2, \ldots. \tag{2.24a}$$

This is a standard stochastic approximation scheme for attempting to solve the projected fixed point relation. In order to improve it, we use the standard device of applying Polyak–Ruppert averaging so as to obtain our final estimate. For a given sample size $n \geq 2$, our final estimate $\widehat{v}_n$ is given by taking the average of these iterates from time n_0 to n—that is

$$\widehat{v}_n := \frac{1}{n - n_0} \sum_{t=n_0+1}^{n} v_t. \tag{2.24b}$$

Here the "burn-in" time n_0 is an integer parameter to be specified.

The stochastic approximation procedure (2.24) is defined in the entire space $\mathbb{X}$; note that it can be equivalently written as iterates in the projected space $\mathbb{R}^d$, via the recursion

$$\vartheta_{t+1} = (1-\eta)\vartheta_t + \eta (\Phi_d L_{t+1} \Phi_d^* \vartheta_t + \Phi_d b_{t+1}). \tag{2.25}$$

The original iterates can be recovered by applying the adjoint operator—that is, $v_t = \Phi_d^* \vartheta_t$ for $t = 1, 2, \ldots$.

2.3.1.1 A finite-sample upper bound

Having introduced the algorithm itself, we are now ready to provide a guarantee on its error. Two matrices play a key role in the statement of our upper bound. The first is the d-dimensional matrix $M := \Phi_d L \Phi_d^*$ that we introduced in Section 2.2.1. We show that the mean-squared error is upper bounded by the approximation error $\inf_{v \in \mathbb{S}} \|v - v^*\|^2$ along with a pre-factor of the form

$$\alpha(M, s) = 1 + \lambda_{\max}\Big((I - M)^{-1}(s^2\, I_d - MM^T)(I - M)^{-T}\Big), \tag{2.26}$$

for $s = |\!|\!| L |\!|\!|_{\mathrm{op}}$. Our bounds also involve the quantity $\kappa(M) = \frac{1}{2}\lambda_{\max}\big(M + M^T\big)$, which we abbreviate by κ when the underlying matrix M is clear from the context.

The second matrix is a covariance matrix, capturing the noise structure of our observations, given by

$$\Sigma^* := \mathrm{cov}\Big(\Phi_d(b_1 - b) + \Phi_d(L_1 - L)\bar{v}\Big).$$

This matrix, along with the constants (σ_L, σ_b) from Assumption 2.1(W), arise in the definition of two additional error terms, namely

$$\mathcal{E}_n(M, \Sigma^*) := \frac{\mathrm{trace}\Big((I - M)^{-1}\Sigma^*(I - M)^{-\top}\Big)}{n}, \quad \text{and} \tag{2.27a}$$

$$\mathcal{H}_n(\sigma_L, \sigma_b, \bar{v}) := \frac{\sigma_L}{(1-\kappa)^3}\Big(\frac{d}{n}\Big)^{3/2}\Big(\|\bar{v}\|^2\sigma_L^2 + \sigma_b^2\Big). \tag{2.27b}$$

As suggested by our notation, the error $\mathcal{H}_n(\sigma_L, \sigma_b, \bar{v})$ is a higher-order term, decaying as $n^{-3/2}$ in the sample size, whereas the quantity $\mathcal{E}_n(M, \Sigma^*)$ is the dominant source of statistical error. With this notation, we have the following:

Theorem 2.1. *Suppose that we are given n i.i.d. observations $\{(L_i, b_i)\}_{i=1}^n$ that satisfy the noise conditions in Assumption 2.1(W). Then there are universal constants (c_0, c) such that for any sample size $n \geq \frac{c_0 \sigma_L^2 d}{(1-\kappa)^2}\log^2\Big(\frac{\|v_0 - \bar{v}\|^2 d}{1-\kappa}\Big)$, then running the algorithm (2.24) with*

$$\textit{stepsize } \eta = \tfrac{1}{c_0 \sigma_L \sqrt{dn}}, \qquad \textit{and burn-in period } n_0 = n/2$$

yields an estimate $\widehat{v}_n$ such that

$$\mathbb{E}\|\widehat{v}_n - v^*\|^2 \leq (1+\omega)\cdot\alpha(M, |\!|\!| L |\!|\!|_{\mathbb{X}}) \inf_{v \in \mathbb{S}} \|v - v^*\|^2 + c\Big(1 + \tfrac{1}{\omega}\Big)\cdot\big\{\mathcal{E}_n(M, \Sigma^*) + \mathcal{H}_n(\sigma_L, \sigma_b, \bar{v})\big\}, \tag{2.28}$$

valid for any $\omega > 0$.

We prove this theorem in Section 2.5.1.

A few comments are in order. First, the quantity $\alpha(M, \|\!|L|\!\|_{\mathbb{X}}) \inf_{v \in \mathbb{S}} \|v - v^*\|^2$ is an upper bound on the approximation error $\|\bar{v} - v^*\|^2$ incurred by the (deterministic) projected fixed point $\bar{v}$. The pre-factor $\alpha(M, \|\!|L|\!\|_{\mathbb{X}}) \geq 1$ measures the instance-specific deficiency of $\bar{v}$ relative to an optimal approximating vector from the subspace, and we provide a more in-depth discussion of this factor in Section 2.3.1.2 to follow. Note that Theorem 2.1 actually provides a family of bounds, indexed by the free parameter $\omega > 0$. By choosing ω arbitrarily close to zero, we can make the pre-factor in front of $\inf_{v \in \mathbb{S}} \|v - v^*\|^2$ arbitrarily close to $\alpha(M, \|\!|L|\!\|_{\mathbb{X}})$—albeit at the expense of inflating the remaining error terms. In Theorem 2.2 to follow, we prove that the quantity $\alpha(M, \|\!|L|\!\|_{\mathbb{X}})$ is, in fact, the smallest approximation factor that can be obtained in any such bound.

The latter two terms in the bound (2.28) correspond to estimation error that arises from estimating $\bar{v}$ based on a set of n stochastic observations. While there are two terms here in principle, we show in Corollary 2.1 to follow that the estimation error is dominated by the term $\mathcal{E}_n(M, \Sigma^*)$ under some natural assumptions. Note that the leading term $\mathcal{E}_n(M, \Sigma^*)$ scales with the local complexity for estimating $\bar{v}$, and we show in Theorem 2.3 that this term is also information-theoretically optimal. In Appendix A.9.2, we perform additional simulation studies on the statistical error terms, showing that the actual performance of Polyak–Ruppert averaging estimator is accurately characterized by the instance-dependent analysis.

In the next subsection, we undertake a more in-depth exploration of the approximation factor in this problem, discussing prior work in the context of the term $\alpha(M, \|\!|L|\!\|_{\mathbb{X}})$ appearing in Theorem 2.1.

2.3.1.2 Detailed discussion of the approximation error

As mentioned in the introduction, upper bounds on the approximation factor have received significant attention in the literature, and it is interesting to compare our bounds.

Past results: In the case where $\gamma_{\max} := \|\!|L|\!\|_{\mathbb{X}} < 1$, the approximation-factor bound (2.6) was established by Tsitsiklis and Van Roy [204], via the following argument. Letting $\widetilde{v} := \Pi_{\mathbb{S}}(Lv^* + b)$, we have

$$\begin{aligned}\|\bar{v} - v^*\|^2 \overset{(i)}{=} \|\bar{v} - \widetilde{v}\|^2 + \|\widetilde{v} - v^*\|^2 &= \|\Pi_{\mathbb{S}}(L\bar{v} + b) - \Pi_{\mathbb{S}}(Lv^* + b)\|^2 + \|\widetilde{v} - v^*\|^2 \\ &\overset{(ii)}{\leq} \|L\bar{v} - Lv^*\|^2 + \|\widetilde{v} - v^*\|^2 \\ &\overset{(iii)}{\leq} \gamma_{\max}^2 \|\bar{v} - v^*\|^2 + \|\widetilde{v} - v^*\|^2. \end{aligned} \tag{2.29}$$

Step (i) uses Pythagorean theorem; step (ii) follows from the non-expansiveness of the projection operator; and step (iii) makes use of the contraction property of the operator L. Note that by definition, we have $\alpha(M, \|\|L\|\|_{\mathbb{X}}) \leq (1 - \|\|L\|\|_{\mathbb{X}})^{-2}$, and so the approximation factor in Theorem 2.1 recovers the bound (2.6) in the worst case. In general, however, the factor $\alpha(M, \|\|L\|\|_{\mathbb{X}})$ can be significantly smaller. See Lemmas 2.1 and 2.2 to follow.

Yu and Bertsekas [226] derived two finer grained upper bounds on the approximation factor; in terms of our notation, their bounds take the form

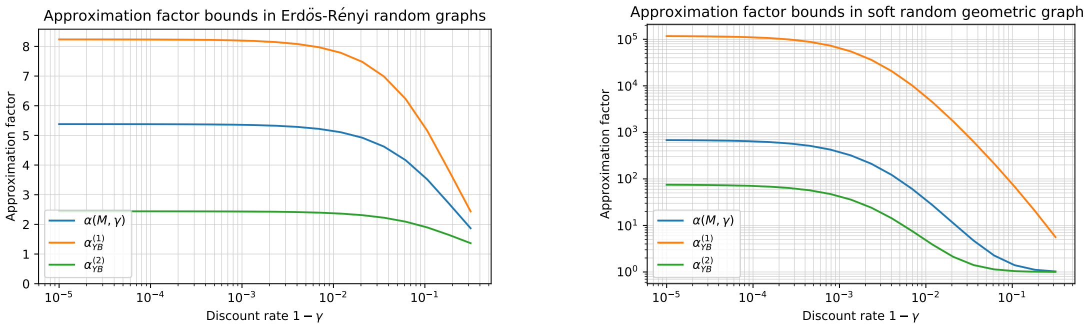
$$\begin{aligned} \alpha_{\mathsf{YB}}^{(1)} &:= 1 + \|\|L\|\|_{\mathbb{X}}^2 \cdot \lambda_{\max}\Big((I - M)^{-1}(I - M)^{-\top}\Big), \\ \alpha_{\mathsf{YB}}^{(2)} &:= 1 + \|\|(I - \Pi_{\mathbb{S}} L)^{-1} \Pi_{\mathbb{S}} L \Pi_{\mathbb{S}^\perp}\|\|_{\mathbb{X}}^2. \end{aligned}$$

It is clear from the definition that $\alpha(M, \|\|L\|\|_{\mathbb{X}}) \leq \alpha_{\mathsf{YB}}^{(1)}$, but $\alpha(M, \|\|L\|\|_{\mathbb{X}})$ can often provide an improved bound. This improvement is indeed significant, as will be shown shortly in Lemma 2.1. On the other hand, the term $\alpha_{\mathsf{YB}}^{(2)}$ is never larger than $\alpha(M, \|\|L\|\|_{\mathbb{X}})$, and is indeed the smallest possible bound that depends only on L and *not* b. However, as pointed out by Yu and Bertsekas, the value of $\alpha_{\mathsf{YB}}^{(2)}$ is not easily accessible in practice, since it depends on the precise behavior of the operator L over the orthogonal complement $\mathbb{S}^\perp$. Thus, estimating the quantity $\alpha_{\mathsf{YB}}^{(2)}$ requires $O(D)$ samples. In contrast, the term $\alpha(M, \|\|L\|\|_{\mathbb{X}})$ depends only on the projected operator M and the operator norm $\|\|L\|\|_{\mathbb{X}}$. The former can be easily estimated using d samples and at smaller computational cost, while the latter is usually known a priori. The discussion in Section 2.4 to follow fleshes out these distinctions.

A simulation study: In order to compare different upper bounds on the approximation factor, we conducted a simple simulation study on the problem of value function estimation, as previously introduced in Section 2.2.2.3. For this problem, the approximation factor $\alpha(M, \gamma)$ is computed more explicitly in Corollary 2.5. The Markov transition kernel is given by the simple random walk on a graph. We consider Gaussian random feature vectors and associate them with two different random graph models, Erdös-Rényi graphs and random geometric graphs, respectively. The details for these models are described and discussed in Appendix A.9.1.

In Figure 2.1, we show the simulation results for the values of the approximation factor. Given a sample from above graphs and feature vectors, we plot the value of $\alpha(M, \gamma)$, $\alpha_{\mathsf{YB}}^{(1)}$ and $\alpha_{\mathsf{YB}}^{(2)}$ against the discount rate $1 - \gamma$, which ranges from 10^{-5} to $10^{-0.5}$. Note that the two plots use different scales: Panel (a) is a linear-log plot, whereas panel (b) is a log-log plot. Figure 2.1 shows that the approximation factor $\alpha(M, \gamma)$ derived in Theorem 2.1 is always between $\alpha_{\mathsf{YB}}^{(1)}$ and $\alpha_{\mathsf{YB}}^{(2)}$. As mentioned before, the latter quantity depends on the particular behavior of the linear operator L in the subspace $\mathbb{S}^\perp$, which can be difficult to estimate. The improvement over $\alpha_{\mathsf{YB}}^{(1)}$, on the other hand, can be significant.

In the Erdös-Rényi model, all the three quantities are bounded by relatively small constant, regardless of the value of γ. The bound $\alpha(M, \gamma)$ is roughly at the midpoint between the bounds $\alpha_{\mathsf{YB}}^{(1)}$ and $\alpha_{\mathsf{YB}}^{(2)}$. On the other hand, the differences are much starker



(a) (b)

Figure 2.1: Plots of various approximation factor as a function of the discount factor γ in the policy evaluation problem. (See the text for a discussion.) (a) Results for an Erdös-Rényi random graph model with $N = 3000$, projected dimension $d = 1000$, and $a = 3$. The resulting number of vertices in the graph $\widetilde{G}$ is 2813. The value of $1 - \gamma$ is plotted in log-scale, and the value of approximation factor is plotted on the standard scale. (b) Results for a random geometric graph model with $N = 3000$, projected dimension $d = 2$, and $r = 0.1$. The resulting number of vertices in the graph $\widetilde{G}$ is 2338. Both the discount rate $1 - \gamma$ and the approximation factor are plotted on the log-scale.

in the random geometric graph case: The bound improves over $\alpha_{\mathsf{YB}}^{(1)}$ by several orders of magnitude, while being off from $\alpha_{\mathsf{YB}}^{(2)}$ by a factor of 10 for large γ. As we discuss shortly in Lemma 2.1, this is because the approximation factor $\alpha(M, \gamma)$ scales as $O\big(\frac{1}{1-\kappa(M)}\big)$ while $\alpha_{\mathsf{YB}}^{(1)}$ scales as $O\big(\frac{1}{(1-\kappa(M))^2}\big)$, making a big difference in the case where the constant $\kappa(M)$ is large.

Some useful bounds on $\alpha(M, \|L\|_{\mathbb{X}})$: We conclude our discussion of the approximation factor with some bounds that can be derived under different assumptions on the operator L and its projected version M. The following lemma is useful in understanding the behavior of the approximation factor as a function of the contractivity properties of the operator L; this is particularly useful in analyzing convergence rates in numerical PDEs.

Lemma 2.1. *Consider a projected matrix $M \in \mathbb{R}^{d \times d}$ such that $(I - M)$ is invertible and $\kappa(M) < 1$.*

(a) For any $s > 0$, we have the bound

$$\alpha(M, s) \leq 1 + |\!|\!|(I - M)^{-1}|\!|\!|_{op}^2 \cdot s^2 \leq 1 + \frac{s^2}{(1 - \kappa(M))^2}. \tag{2.30a}$$

(b) For $s \in [0, 1]$, we have

$$\alpha(M, s) \leq 1 + 2 \vert\!\vert\!\vert (I - M)^{-1} \vert\!\vert\!\vert_{op} \leq 1 + \frac{2}{1 - \kappa(M)}. \tag{2.30b}$$

See Appendix A.7.1 for the proof of this lemma.

A second special case, also useful, is when the matrix M is symmetric, a setting that appears in least-squares regression, value function estimation in reversible Markov chains, and self-adjoint elliptic operators. The optimal approximation factor $\alpha(M, \gamma_{\max})$ can be explicitly computed in such cases.

Lemma 2.2. *Suppose that M is symmetric with eigenvalues $\{\lambda_j(M)\}_{j=1}^d$ such that $\lambda_{max}(M) < 1$. Then for any $s > 0$, we have*

$$\alpha(M, s) = 1 + \max_{j=1,\ldots,d} \frac{s^2 - \lambda_j^2}{(1 - \lambda_j)^2}. \tag{2.31}$$

See Appendix A.7.2 for the proof of this lemma.

Lemma 2.1 reveals that there is a qualitative shift between the non-expansive case $\vert\!\vert\!\vert L \vert\!\vert\!\vert_{\mathbb{X}} \leq 1$ and the complementary expansive case. In the latter case, the optimal approximation factor always scales as $O\big(\frac{1}{(1-\kappa(M))^2}\big)$, but below the threshold $\vert\!\vert\!\vert L \vert\!\vert\!\vert_{\mathbb{X}} = 1$, the approximation factor drastically improves to become $O\big(\frac{1}{1-\kappa(M)}\big)$. It is worth noting that both bounds can be achieved up to universal constant factors. In the context of differential equations, the bound of the form (a) in Lemma 2.1 is known as Céa's lemma [35], which plays a central role in the convergence rate analysis of the Galerkin methods for numerical differential equations. However, the instance-dependent approximation factor $\alpha(M, \vert\!\vert\!\vert L \vert\!\vert\!\vert_{\mathbb{X}})$ can often be much smaller: the global coercive parameter needed in Céa's estimate is replaced by the bounds on the behavior of the operator L in the finite-dimensional subspace. The part (b) in Lemma 2.1 generalizes Céa's energy estimate from the symmetric positive-definite case to the general non-expansive setting. See Corollary 2.4 for a more detailed discussion on the consequences of our results to elliptic PDEs.

Lemmas 2.1 and 2.2 yield the following corollary of the general bound (2.28) under different conditions on the operator L.

Corollary 2.1. *Under the conditions of Theorem 2.1 and given a sample size $n \geq \frac{c_0 \sigma_L^2 d}{(1-\kappa)^2} \log^2 \left(\frac{\|v_0 - \bar{v}\|^2 d}{1-\kappa}\right)$:*

(a) There is a universal positive constant c such that

$$\mathbb{E}\|\widehat{v}_n - v^*\|^2 \leq c \left\{ \frac{\vert\!\vert\!\vert L \vert\!\vert\!\vert_{\mathbb{X}}^2}{\big(1 - \kappa(M)\big)^2} \cdot \inf_{v \in \mathbb{S}} \|v - v^*\|^2 + \frac{(\sigma_b^2 + \sigma_L \|\bar{v}\|^2)}{\big(1 - \kappa(M)\big)^2} \frac{d}{n} \right\} \tag{2.32a}$$

for any operator L, and its associated projected operator $M = \Phi_d L \Phi_d^*$.

(b) Moreover, when L is non-expansive ($\|L\|_{\mathbb{X}} \leq 1$), we have

$$\mathbb{E}\|\widehat{v}_n - v^*\|^2 \leq c\Big\{\frac{1}{1-\kappa(M)} \cdot \inf_{v \in \mathbb{S}} \|v - v^*\|^2 + \frac{(\sigma_b^2 + \sigma_L \|\bar{v}\|^2)}{\big(1-\kappa(M)\big)^2}\ \frac{d}{n}\Big\}. \tag{2.32b}$$

See Appendix A.3 for the proof of this claim.

As alluded to before, the simplified form of Corollary 2.1 no longer has an explicit higher order term, and the statistical error now scales at the parametric rate d/n. It is worth noting that the lower bound on n required in the assumption of the corollary is a mild requirement: in the absence of such a condition, the statistical error term $\frac{(\sigma_b^2+\sigma_L\|\bar{v}\|^2)}{(1-\kappa)^2}\ \frac{d}{n}$ in both bounds would blow up, rendering the guarantee vacuous.

2.3.2 Lower bounds

In this section, we establish information-theoretic lower bounds on the approximation factor, as well as the statistical error. Our eventual result (in Corollary 2.2) shows that the first two terms appearing in Theorem 2.1 are both optimal in a certain instance-dependent sense. However, a precise definition of the local neighborhood of instances over which the lower bound holds requires some definitions. In order to motivate these definitions more transparently and naturally arrive at both terms of the bound, the following section presents individual bounds on the approximation and estimation errors, and then combines them to obtain Corollary 2.2.

2.3.2.1 Lower bounds on the approximation error

As alluded to above, the first step involved in a lower bound is a precise definition of the collection of problem instances over which it holds; let us specify a natural such collection for lower bounds on the approximation error. Each problem instance is specified by the joint distribution of the observations (L_i, b_i), which implicitly specifies a pair of means $(L, b) = (\mathbb{E}[L_i], \mathbb{E}[b_i])$. For notational convenience, we define this class by first defining a collection comprising instances specified solely by the mean pair (L, b), and then providing restrictions on the distribution of (L_i, b_i). Let us define the first such component. For a given matrix $M_0 \in \mathbb{R}^{d \times d}$ and vector $h_0 \in \mathbb{R}^d$, write

$$\mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max}) := \Big\{(L, b) \;\Big|\; \begin{array}{c} \|L\|_{\mathbb{X}} \leq \gamma_{\max}, \quad \mathcal{A}(\mathbb{S}, v^*) \leq \delta^2, \quad \dim(\mathbb{X}) = D, \\ \Phi_d L \Phi_d^* = M_0, \quad \text{and} \quad \Phi_d b = h_0. \end{array}\Big\}.$$

In words, this is a collection of all instances of the pair $(L, b) \in \mathfrak{L} \times \mathbb{R}^D$ whose projections onto the subspace of interest are fixed to be the pair (M_0, h_0), and whose approximation error is less than δ^2. In addition, the operator L satisfies a certain bound on its operator norm.

Having specified a class of (L, b) pairs, we now turn to the joint distribution over the pair of observations (L_i, b_i), which we denote for convenience by $\mathbb{P}_{L,b}$. Now define the collection of instances

$$\mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b) := \Big\{ \mathbb{P}_{L,b} \;\Big|\; (L_i, b_i) \text{ satisfies Assumption 2.1(S) with constants } (\sigma_L, \sigma_b) \Big\}.$$

This is simply the class of all distributions such that our observations satisfy Assumption 2.1(S) with pre-specified constants. As a point of clarification, it is useful to recall that our upper bound in Theorem 2.1 only needed Assumption 2.1(W) to hold, and we could have chosen to match this by defining the $\mathbf{G}_{\mathsf{var}}$ under Assumption 2.1(W). We comment further on this issue following the theorem statement.

We are now ready to state Theorem 2.2, which is a lower bound on the worst-case approximation factor over all problem instances such that $(L, b) \in \mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max})$ and $\mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)$. Note that such a collection of problem instances is indeed *local* around the pair (M_0, h_0). Two settings are considered in the statement of the theorem: *proper* estimators when $\widehat{v}_n$ is restricted to take values in the subspace $\mathbb{S}$; and *improper* estimators, where $\widehat{v}_n$ can take values in the entire space $\mathbb{X}$. We use $\widehat{\mathcal{V}}_{\mathbb{S}}$ and $\widehat{\mathcal{V}}_{\mathbb{X}}$ to denote the class of proper and improper estimators, respectively. Finally, we use the shorthand $\mathbb{C}_{\mathsf{approx}} \equiv \mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max})$ for convenience.

Theorem 2.2. *Suppose $M_0 \in \mathbb{R}^{d \times d}$ is a matrix such that $I - M_0$ is invertible, and that the scalars (σ_L, σ_b) are such that $\sigma_L \geq \gamma_{\max}$ and $\sigma_b \geq \delta$. If the ambient dimension satisfies $D \geq d + \frac{12}{\omega} n^2$ for some scalar $\omega \in (0, 1)$, then we have the lower bounds*

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{S}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq (1 - \omega) \cdot \alpha(M_0, \gamma_{\max}) \cdot \delta^2 \quad \textit{and} \tag{2.33a}$$

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq (1 - \omega) \cdot \big(\alpha(M_0, \gamma_{\max}) - 1\big) \cdot \delta^2. \tag{2.33b}$$

See Section 2.5.2 for the proof of this claim.

A few remarks are in order. First, Theorem 2.2 shows that the approximation factor upper bound in Theorem 2.1 is information-theoretically optimal in the instance-dependent sense: in the case of proper estimators, the upper and lower bound can be made arbitrarily close by choosing the constant ω arbitrarily small in both theorems. Both bounds depend on the projected matrix M_0, characterizing the fundamental impact of the geometry in the projected space on the complexity of the estimation problem. The lower bound for improper estimators is slightly smaller, but for most practical applications we have $\alpha(M_0, \gamma_{\max}) \gg 1$ and so this result should be viewed as almost equivalent.

Second, note that we may also extract a worst-case lower bound on the approximation factor from Theorem 2.2. Indeed, for a scalar $\gamma_{\max} \in (0, 1)$, consider the family of instances in the aforementioned problem classes satisfying $|\!|\!| L |\!|\!|_{\mathbb{X}} \leq \gamma_{\max}$. Setting $M_0 = \gamma_{\max}^2 I_d$ and applying Theorem 2.2, we see that (in a worst-case sense over this class), the risk of any estimator is lower bounded by $\frac{1}{1 - \gamma_{\max}^2} \mathcal{A}(\mathbb{S}, v^*)$. This establishes the optimality of the classical worst-case upper bound (2.6).

Third, notice that the theorem requires the noise variances (σ_L, σ_b) to be large enough, and this is a natural requirement in spite of the fact that we seek lower bounds on the approximation error. Indeed, in the extreme case of noiseless observations, we have access to the population pair (L, b) with a single sample, and can compute both v^* and its projection onto the subspace $\mathbb{S}$ without error. From a more quantitative standpoint, it is worth noting that our requirements $\sigma_L \geq \gamma_{\max}$ and $\sigma_b \geq \delta$ are both mild, since the scalars $\gamma_{\max}$ and δ are typically order 1 quantities. Indeed, if both of these bounds held with equality, then Corollary 2.1 yields that the statistical error would be of the order $O(d/n)$, and so strictly smaller than the approximation error we hope to capture[4].

Observe that Theorem 2.2 requires the ambient dimension D to be larger than n^2. As mentioned in the introduction, we should not expect any non-trivial approximation factor when $n \geq D$, but this leaves open the regime $n \ll D \ll n^2$. Is a smaller approximation factor achievable when D is not extremely large? We revisit this question in Section 2.3.2.4, showing that while there are some quantitative differences in the lower bound, the qualitative nature of the message remains unchanged.

Regarding our noise assumptions, it should first be noted that the class of instances satisfying Assumption 2.1(W) is strictly larger than the corresponding class satisfying Assumption 2.1(S), and so our lower bound extends immediately to the former case. Second, it is important to note that imposing only Assumption 2.1(W) would in principle allow the noise in the orthogonal complement $\mathbb{S}^\perp$ to grow in an unbounded fashion, and one should expect that it is indeed optimal to return an estimate of the projected fixed point $\bar{v}$. As such, Theorem 2.2 constitutes a more meaningful lower bound, since we operate instead under the stronger Assumption 2.1(S) and enforce second moment bounds on the noise not only for basis vectors in $\mathbb{S}$, but also its orthogonal complement. Assumption 2.1(S) allows for other natural estimators: For instance, the plug-in estimator of v^* via the original fixed point equation (2.1) would now incur finite error. Our lower bound—which operates under the stronger assumption and is thus more challenging to establish—shows that the stochastic approximation estimator analyzed in Theorem 2.1 is optimal *even if* the noise in $\mathbb{S}^\perp$ behaves as well as that in $\mathbb{S}$.

2.3.2.2 Lower bounds on the estimation error

We now turn to establishing a minimax lower bound on the estimation error that matches the statistical error term in Theorem 2.1. This lower bound takes a slightly different form from Theorem 2.2: rather than studying the total error $\|\widehat{v}_n - v^*\|$ directly, we establish a lower bound on the error $\|\widehat{v}_n - \bar{v}\|$ instead.

[4] As a side remark, we note that our noise conditions can be further weakened, if desired, via a mini-batching trick. To be precise, given any problem instance $\mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)$ and any integer $m > 0$, one could treat the sample mean of m independent samples as a single sample, resulting in a problem instance in the class $\mathbf{G}_{\mathsf{var}}\big(\frac{\sigma_L}{\sqrt{m}}, \frac{\sigma_b}{\sqrt{m}}\big)$. The same lower bound still applies to the class $\mathbf{G}_{\mathsf{var}}\big(\frac{\sigma_L}{\sqrt{m}}, \frac{\sigma_b}{\sqrt{m}}\big)$, at a cost of stronger dimension requirement $D \geq d + \frac{12}{\omega} n^2 m^2$.

Indeed, the latter term is more meaningful to study in order to characterize the estimation error—which depends on the sample size n—since for large sample sizes, the total error $\|\widehat{v}_n - v^*\|$ will be dominated by a constant approximation error. As we demonstrate shortly, the term $\|\widehat{v}_n - \bar{v}\|$ depends on noise covariance and the geometry of the matrix M_0 in the *projected space*, while having the desired dependence on the sample size n. It is worth noting also that this automatically yields a lower bound on the error $\|\widehat{v}_n - v^*\|$ when we have $\bar{v} = v^*$.

We are now ready to prove a local minimax lower bound for estimating $\bar{v} \in \mathbb{S}$, which is given by the solution to the projected linear equation $\bar{v} = \Pi_{\mathbb{S}}(L\bar{v} + b)$. While our objective is to prove a local lower bound around each pair $(L_0, b_0) \in \mathfrak{L} \times \mathbb{X}$, the fact that we are estimating $\bar{v}$ implies that it suffices to define our set of local instances in the d-dimensional space of projections. In particular, our mean parameters (L, b) are specified by those pairs for which $\Phi_d L \Phi_d^*$ is close to $M_0 := \Phi_d L_0 \Phi_d^*$, and $\Phi_d b$ is close to $h_0 := \Phi_d b_0$. Specifically, let $\bar{v}_0$ denote the solution to the projected linear equation $\bar{v}_0 = \Pi_{\mathbb{S}}(L_0 \bar{v}_0 + b_0)$, and define the neighborhood

$$\mathfrak{N}(M_0, h_0) := \Big\{ (M', h') : \|M' - M_0\|_F \leq \sigma_L \sqrt{\frac{d}{n}}, \text{ and } \|h' - h_0\|_2 \leq \sigma_b \sqrt{\frac{d}{n}} \Big\}, \tag{2.34}$$

which, in turn, defines a local class of problem instances (L, b) given by

$$\mathbb{C}_{\mathsf{est}} := \Big\{ (L, b) \mid \big(\Phi_d L \Phi_d^*, \Phi_d b \big) \in \mathfrak{N}(M_0, h_0) \Big\}.$$

We have thus specified our local neighborhood in terms of the mean pair (L, b), and as before, it remains to define a local class of distributions on these instances. Toward this end, define the class

$$\begin{aligned} &\mathbf{G}_{\mathsf{cov}}(\Sigma_L, \Sigma_b, \sigma_L, \sigma_b) \\ &:= \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b) \cap \Big\{ \mathbb{P}_{L,b} \,\Big|\, \mathrm{cov}\Big(\Phi_d(b_1 - b) \Big) \preceq \Sigma_b \quad \text{and} \quad \mathrm{cov}\Big(\Phi_d(L_1 - L)\bar{v}_0 \Big) \preceq \Sigma_L \Big\}, \end{aligned} \tag{2.35}$$

corresponding to distributions on the observation pair (L_i, b_i) that satisfy Assumption 2.1(S) and whose "effective noise" covariances are dominated by the PSD matrices Σ_L and Σ_b.

Note that Assumption 2.1(S) implies the diagonal elements of above two covariance matrices are bounded by σ_b^2 and $\sigma_L^2 \|\bar{v}_0\|^2$, respectively. In order to avoid conflicts between assumptions, we assume throughout that for all indices $j \in [d]$, the diagonal entries of the covariance matrices satisfy the conditions

$$(\Sigma_b)_{j,j} \leq \sigma_b^2 \qquad \text{and} \qquad (\Sigma_L)_{j,j} \leq \sigma_L^2 \|\bar{v}_0\|^2. \tag{2.36}$$

We then have the following theorem for the estimation error $\|\widehat{v}_n - \bar{v}\|$, where we use the shorthand $\mathbf{G}_{\mathsf{cov}} \equiv \mathbf{G}_{\mathsf{cov}}(\Sigma_L, \Sigma_b, \sigma_L, \sigma_b)$ for brevity.

Theorem 2.3. *Under the setup above, suppose the matrix $I - M_0$ is invertible, and suppose that $n \geq 16\sigma_L^2 \|\!|(I - M_0)^{-1}\|\!|_{op}^2 d$. Then there is a universal constant $c > 0$ such that*

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{est}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \geq c \cdot \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b).$$

See Appendix A.4 for the proof of this claim.

The estimation error lower bound in Theorem 2.3 is the worst-case instantiation of the statistical error term $\mathcal{E}_n(M, \Sigma^*)$ in Theorem 2.1 within the local problem class, up to a universal constant. Indeed, in the asymptotic limit $n \to \infty$, the regularity of the problem can be leveraged in conjunction with classical Le Cam theory (see, e.g., [209]) to show that the asymptotic optimal limiting distribution is a Gaussian law with covariance $(I - M)^{-1}\Sigma^*(I - M)^{-\top}$. (See the paper [98] for a detailed analysis of this type in the special case of policy evaluation in tabular MDPs.) This optimality result holds in a "local" sense: it is minimax optimal in a small neighborhood of radius $O(1/\sqrt{n})$ around a given problem instance (M_0, h_0). Theorem 2.3, on the other hand, is non-asymptotic, showing that a similar result holds provided n is lower bounded by an explicit, problem-dependent quantity of the order $\sigma_L^2 d \|\!|(I - M_0)^{-1}\|\!|_{\mathrm{op}}^2$. This accommodates a broader range of sample sizes than the upper bound in Theorem 2.1.

2.3.2.3 Combining the bounds

Having presented separate lower bounds on the approximation and estimation errors in conjunction with definitions of local classes of instances over which they hold, we are now ready to present a corollary which combines the two lower bounds in Theorems 2.2 and 2.3.

We begin by defining the local classes of instances over which our combined bound holds. Given a matrix-vector pair (M_0, h_0), covariance matrices (Σ_L, Σ_b), ambient dimension $D > 0$, and scalars $\delta, \gamma_{\max}, \sigma_L, \sigma_b > 0$, we begin by specifying a collection of mean pairs (L, b) via

$$\mathbb{C}_{\mathsf{final}}(M_0, h_0, D, \delta, \gamma_{\max}) := \bigcup_{(M', h') \in \mathfrak{N}_n(M_0, h_0)} \mathbb{C}_{\mathsf{approx}}(M', h', D, \delta, \gamma_{\max}). \tag{2.37}$$

Clearly, this represents a natural combination of the classes $\mathbb{C}_{\mathsf{approx}}$ and $\mathbb{C}_{\mathsf{est}}$ introduced above. We use the shorthand $\mathbb{C}_{\mathsf{final}}$ for this class for brevity. Our collection of distributions $\mathbb{P}_{L,b}$ is still given by the class $\mathbf{G}_{\mathsf{cov}}$ from equation (2.35).

With these definitions in hand, we are now ready to state our combined lower bound.

Corollary 2.2. *Under the setup above, suppose that the pair (σ_L, σ_b) satisfies the conditions in Theorem 2.2 and equation (2.36), and that the matrix M_0 satisfies $\|\!|M_0\|\!|_{op} \leq \gamma_{\max} - \sigma_L\sqrt{d/n}$. Moreover, suppose that the sample size and ambient dimension satisfy*

$n \geq 16\sigma_L^2 \|(I - M_0)^{-1}\|_{op}^2 d$ and $D \geq d + 36n^2$, respectively. Then the following minimax lower bound holds for a universal positive constant c:

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{final}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq c \cdot \Big\{ \big(\alpha(M_0, \gamma_{\max}) - 1\big) \cdot \delta^2 + \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b) \Big\}.$$

We prove this corollary in Appendix A.5. It is relatively straightforward consequence of combining Theorems 2.2 and 2.3.

The combined lower bound matches the expression $\alpha(M_0, \gamma_{\max})\mathcal{A}(\mathbb{S}, v^*) + \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b)$, given by the first two terms of Theorem 2.1, up to universal constant factors. Recall from our discussion of Theorem 2.1 that the high-order term $\mathcal{H}_n(\sigma_L, \sigma_b, \bar{v})$ represents the "optimization error" of the stochastic approximation algorithm, which depends on the coercive condition $\kappa(M_0)$ instead of the natural geometry $I - M_0$ of the problem. While we do not expect this term to appear in an information-theoretic lower bound, the leading estimation error term $\mathcal{E}_n(M_0, \Sigma_L + \Sigma_b)$ will dominate the high-order term when the sample size n is large enough. For such a range of n, the bound in Theorem 2.1 is information-theoretically optimal in the local class specified above. More broadly, consider the class of all instances satisfying Assumption 2.1(S), with $\kappa(M) \leq \kappa$ and $\|L\|_{\mathbb{X}} \leq 1$. Then the bound in Theorem 2.1 is optimal, in a worst-case sense, over this class as long as the sample size exceeds the threshold $\frac{c\sigma_L^2}{(1-\kappa)^2} d$.

2.3.2.4 The intermediate regime

It remains to tie up some loose ends. Note that the lower bound in Theorem 2.2 requires a condition $D \gg n^2$. On the other hand, it is easy to see that the approximation factor can be made arbitrarily close to 1 when $n \gg D$. (For example, one could run the estimator based on stochastic approximation and averaging—which was analyzed in Theorem 2.1—with the entire Euclidean space $\mathbb{X}$, and project the resulting estimate onto the subspace $\mathbb{S}$.) In the middle regime $n \ll D \ll n^2$, however, it is not clear which estimator is optimal.

In the following theorem, we present a lower bound for the approximation factor in this intermediate regime, which establishes the optimality of Theorem 2.1 up to a constant factor.

Theorem 2.4. *Suppose $M_0 \in \mathbb{R}^{d \times d}$ is a matrix such that $I - M_0$ is invertible, and that the scalars (σ_L, σ_b) satisfy $\sigma_L \geq 1 + \gamma_{\max}$ and $\sigma_b \geq \delta$. If the ambient dimension satisfies $D \geq d + 3qn^{1+1/q}$ for some integer $q \in \left[2, \log n \wedge \frac{1}{\sqrt{2(1-\gamma_{\max} \wedge 1)}}\right]$, then we have the lower bound*

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq \frac{\alpha(M, \gamma_{\max}) - 1}{4q^2} \cdot \delta^2.$$

See Appendix A.6 for the proof of this theorem.

Theorem 2.4 resolves the gap in the intermediate regime, up to a constant factor that depends on q. In particular, the stochastic approximation estimator (2.24) for projected equations still yields a near-optimal approximation factor. Compared to Theorem 2.2, Theorem 2.4 weakens the requirement on the ambient dimension D and covers the entire regime $D \gg n$. Furthermore, using the same arguments as in Corollary 2.2, this theorem can also be combined with Theorem 2.3 to obtain the following lower bound in the regime $D \geq d + 3qn^{1+1/q}$, for any integer $q > 0$:

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{final}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq c \cdot \left\{ \frac{\alpha(M_0, \gamma_{\max}) - 1}{q^2} \cdot \delta^2 + \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b) \right\}.$$

$q = \lim_{n\to\infty} \frac{\log D_n}{\log n}$	$[2, \infty)$	$(1, 2)$	$(0, 1)$
Lower bound	$\alpha(M_0, \gamma_{\max})$	$c_q \cdot \alpha(M_0, \gamma_{\max})$	1
Upper bound	$\alpha(M_0, \gamma_{\max})$	$\alpha(M_0, \gamma_{\max})$	1

Table 2.1: Bounds on the approximation factor $\frac{\mathbb{E}\|\widehat{v}_n - v^*\|^2}{\mathcal{A}(\mathbb{S}, v^*)}$ for proper estimators in different ranges of ambient dimension. Here, $c_q \in (0, 1)$ represents a constant depending only on the aspect ratio q.

Let us summarize our approximation factor lower bounds in the various regimes. Consider a sequence of problem instances $\left(\mathbb{P}_{L,b}^{(n)}\right)_{n=1}^{\infty}$ with increasing ambient dimension D_n. Let the projected dimension d, noise variances (σ_L, σ_b), oracle error δ, projected matrix $\Phi_d L^{(n)} \Phi_d^* = M$, and the operator norm bound $|||L|||_{\mathbb{X}} \leq \gamma_{\max}$ all be fixed. Table 2.1 presents a combination of our results from Theorems 2.1, 2.2, and 2.4; our results suggest that the optimal approximation factor exhibits a "slow" phase transition phenomenon. It is an interesting open question whether the phase transition is sharp, and to identify the asymptotically optimal approximation factor in the regime $\lim_{n\to\infty} \frac{\log D_n}{\log n} = 1$ since our lower bounds do not apply in this linear regime.

2.4 Consequences for specific models

We now discuss the consequences of our main theorems for the three examples introduced in Section 2.2.2. For brevity, we state only upper bounds for the first two examples; our third example for temporal difference learning methods includes both upper and lower bounds.

2.4.1 Linear regression

Recall the setting of linear regression[5] from Section 2.2.2.1, including our i.i.d. observation model (2.14). We assume bounds on the second moment of ε and fourth moment of X—namely, the existence of some $\varsigma > 0$ such that

$$\mathbb{E}\langle u,\, X\rangle^4 \leq \varsigma^4, \quad \text{and} \quad \mathbb{E}(\varepsilon^2) \leq \varsigma^2 \qquad \text{for all } u \in \mathbb{S}^{D-1}. \tag{2.38}$$

These conditions ensure that Assumption 2.1(W) is satisfied with $(\sigma_L, \sigma_b) = (\beta^{-1}\varsigma^2, \beta^{-1}\varsigma^2)$. Recall that the (unprojected) covariance matrix satisfies the PSD relations $\mu I \preceq \mathbb{E}[XX^\top] \preceq \beta I$, and define the d-dimensional covariance matrix $\Sigma := \mathbb{E}\Big[(\Phi_d X)(\Phi_d X)^\top\Big]$ for convenience.

In this case, our stochastic approximation iterates (2.24a) take the form

$$v_{t+1} = v_t - \eta\Big(\Pi_{\mathbb{S}} X_{t+1} X_{t+1}^\top \Pi_{\mathbb{S}} v_t + Y_{t+1} \Pi_{\mathbb{S}} X_{t+1}\Big), \quad \text{for all } t = 0, 1, 2, \ldots, \tag{2.39}$$

and we take the averaged iterates $\widehat{v}_n := \frac{2}{n}\sum_{t=n/2}^{n-1} v_t$. For this procedure, we have the following guarantee:

Corollary 2.3. *Suppose that we have n i.i.d. observations $\{(X_i, Y_i)\}_{i=1}^n$ from the model (2.14) satisfying the moment conditions (2.38). Then there are universal positive constants (c, c_0) such that given a sample size $n \geq \frac{c_0 \varsigma^4 d}{\lambda_{\min}^2(\Sigma)} \log^2\Big(\frac{\beta}{\mu}\|v_0 - \bar{v}\|_2^2 d\Big)$, if the stochastic approximation scheme (2.39) is run with step size $\eta = \frac{1}{c_0 \varsigma^2 \sqrt{dn}}$, then the averaged iterate satisfies the bound*

$$\mathbb{E}\|\widehat{v}_n - v^*\|_2^2 \leq (1+\omega)\cdot\alpha\big(I_d - \tfrac{\Sigma}{\beta}, 1 - \tfrac{\mu}{\beta}\big)\mathcal{A}(\mathbb{S}, v^*) + c\cdot\frac{\mathrm{trace}(\Sigma^{-1})\cdot\mathbb{E}(\varepsilon^2)}{\omega n} + \frac{c}{\omega}\Big(\frac{\varsigma^2}{\lambda_{\min}(\Sigma)}\cdot\sqrt{\tfrac{d}{n}}\Big)^3$$

for any $\omega > 0$.

This result is a direct consequence of Theorem 2.1 in application to this model.

Note that the statistical error term $\frac{\mathrm{trace}(\Sigma^{-1})\cdot\mathbb{E}(\varepsilon^2)}{n}$ in this case corresponds to the classical statistical rates for linear regression in this low-dimensional subspace. The approximation factor, by Lemma 2.2, admits the closed form expression

$$\alpha\Big(I_d - \tfrac{\Sigma}{\beta}, 1 - \tfrac{\mu}{\beta}\Big) = \max_{j\in[d]} \tfrac{\mu^2 + 2\beta(\lambda_j - \mu)}{\lambda_j^2},$$

where $\{\lambda_j\}_{j=1}^d$ denote the eigenvalues of the matrix Σ. Since $\lambda_j \in [\mu, \beta]$ for each $j \in [d]$, the approximation factor is at most of the order $O\Big(\frac{\beta}{\lambda_{\min}(\Sigma)}\Big)$.

[5] Note that the stochastic approximation iterates are invariant under translation, and consequently we can assume without loss of generality that $\bar{v} = 0$.

Compared to known sharp oracle inequalities for linear regression (e.g., [173]), the approximation factor in our bound is not 1 but rather a problem-dependent quantity. This is because we study the *estimation error* under the standard Euclidean metric $\|\cdot\|_2$, as opposed to the *prediction error* under the data-dependent metric $\|\cdot\|_{L^2(P_X)}$. When the covariance matrix $\mathbb{E}[XX^\top]$ is identity, the approximation factor $\alpha\big(I_d - \frac{\Sigma}{\beta}, 1 - \frac{\mu}{\beta}\big)$ is equal to 1, recovering classical results. Another error metric of interest, motivated by applications such as transfer learning [124], is the prediction error when the covariates X follow a different distribution Q. For such a problem, the result above can be modified straightforwardly by choosing the Hilbert space $\mathbb{X}$ to be $\mathbb{R}^D$, equipped with the inner product $\langle u, v\rangle := u^\top \big(\mathbb{E}_Q[XX^\top]\big)^{-1} v$.

2.4.2 Galerkin methods

We now return to the example of Galerkin methods, as previously introduced in Section 2.2.2.2, with the i.i.d. observation model (2.18). We assume the basis functions $\phi_1, \ldots, \phi_d$ to have uniformly bounded function value and gradient, and define the scalars

$$\sigma_L := \Big(1 + \frac{2}{\beta}\Big) \max_{j \in [d]} \sup_{x \in \Omega} \|\nabla \phi_j(x)\|_2, \quad \text{and} \quad \sigma_b := \frac{\|f\|_{\mathbb{L}^2} + 1}{\beta} \max_{j \in [d]} \sup_{x \in \Omega} |\phi_j(x)|. \tag{2.40}$$

These boundedness conditions are naturally satisfied by many interesting basis functions such as the Fourier basis[6], and ensure—we verify this concretely in the proof of Corollary 2.4 to follow—that our observation model satisfies Assumption 2.1(W) with parameters (σ_L, σ_b).

Taking the finite-dimensional representation $v = \vartheta^\top \phi$, the stochastic approximation estimator for solving equation (2.19) is given by

$$\begin{aligned} \vartheta_{t+1} &= \vartheta_t - \beta^{-1}\eta\Big(\nabla\phi(x_{t+1})^\top a_{t+1} \nabla\phi(x_{t+1})\vartheta_t - f_{t+1}\phi(y_{t+1})\Big), \quad \text{for } t = 0, 1, \cdots \\ \widehat{\vartheta}_n &:= \frac{2}{n}\sum_{t=n/2}^{n-1} \vartheta_t, \quad \text{and} \quad \widehat{v}_n := \widehat{\vartheta}_n^\top \phi. \end{aligned}$$

[6] In the typical application of finite-element methods, basis functions based on local interpolation are widely used [28]. These basis functions can have large sup-norm, but via application of the Walsh–Hadamard transform, a new basis can be obtained satisfying condition (2.40) with dimension-independent constants. Since the stochastic approximation algorithm is invariant under orthogonal transformation, this modification is only for the convenience of analysis and does not change the algorithm itself.

In order to state our statistical guarantees for $\widehat{v}_n$, we define the following matrices:

$$\begin{aligned} M &:= I_d - \beta^{-1} \int_\Omega \nabla \phi(x)^\top a(x) \nabla \phi(x) dx, \\ \Sigma_L &:= \frac{1}{\beta^2} \int_\Omega \big(\nabla \phi\big)^\top a \nabla \bar{v} (\nabla \bar{v})^\top a \nabla \phi \; dx - \frac{1}{\beta^2} \Big(\int_\Omega \big(\nabla \phi\big)^\top a \nabla \bar{v} \; dx \Big) \Big(\int_\Omega \big(\nabla \phi\big)^\top a \nabla \bar{v} \; dx \Big)^\top \\ &\quad + \frac{1}{\beta^2} \int_\Omega (\nabla \phi)^\top \Big[(\nabla \bar{v})(\nabla \bar{v})^\top + \mathrm{diag}\big(\|\nabla \bar{v}\|_2^2 - (\partial_j \bar{v})^2 \big)_{j=1}^m \Big] (\nabla \phi) dx, \\ \Sigma_b &:= \frac{1}{\beta^2} \int_\Omega \big(f(x)^2 + 1 \big) \phi(x) \phi(x)^\top \; dx - \frac{1}{\beta^2} \Big(\int_\Omega f(x) \phi(x) dx \Big) \Big(\int_\Omega f(x) \phi(x) dx \Big)^\top . \end{aligned}$$

With these definitions in hand, we are ready to state the consequence of our main theorems to the estimation problem of elliptic equations.

Corollary 2.4. *Under the setup above, there are universal positive constants* (c, c_0) *such that if* $n \geq \frac{c_0 \sigma_L^2 d}{(1-\kappa(M))^2} \log^2 \left(\frac{\|v_0 - \bar{v}\|^2 \beta d}{\mu} \right)$ *and the stochastic approximation scheme is run with step size* $\eta = \frac{1}{c_0 \sigma_L \sqrt{dn}}$, *then the averaged iterates satisfy*

$$\begin{aligned} \mathbb{E} \|\widehat{v}_n - v^*\|_{\mathbb{X}}^2 \leq (1+\omega) \alpha\big(M, 1 - \tfrac{\mu}{\beta}\big) \inf_{v \in \mathbb{S}} \|v - v^*\|_{\mathbb{X}}^2 \\ + c\big(1 + \tfrac{1}{\omega}\big) \cdot \big(\mathcal{E}_n(M, \Sigma_L + \Sigma_b) + \mathcal{H}_n(\sigma_L, \sigma_b, \bar{v}) \big) \end{aligned}$$

for any $\omega > 0$.

See Appendix A.8.3.2 for the proof of this corollary.

Note that the approximation factor $\alpha\Big(M, 1 - \frac{\mu}{\beta}\Big)$ scales as $\mathcal{O}(\beta/\mu)$, which recovers Céa's energy estimates in the symmetric and uniform elliptic case [35]. On the other hand, for a suitable choice of basis vectors, the bound in Corollary 2.4 can often be much smaller: the parameter μ corresponding to a global coercive condition can be replaced by the smallest eigenvalue of the *projected* operator M. Furthermore, note that our analysis does not require the symmetry and contraction condition of the operator L, and so applies also to the case where the operator A is not uniformly elliptic.

It is also worth noting that the bound in Corollary 2.4 is given in terms of Sobolev norm $\| \cdot \|_{\mathbb{X}} = \| \cdot \|_{\dot{\mathbb{H}}^1}$, as opposed to standard $\mathbb{L}^2$-norm used in the nonparametric estimation literature. By the Poincaré inequality, a Sobolev $\dot{\mathbb{H}}^1$-norm bound implies an $\mathbb{L}^2$-norm bound, and ensures stronger error guarantees on the gradient of the estimated function.

2.4.3 Temporal difference learning

We now turn to the final example previously introduced in Section 2.2.2.3, namely that of the TD algorithm in reinforcement learning. Recall the i.i.d. observation model (2.21).

Also recall the equivalent form of the projected fixed point equation (2.23), and note that the population-level operator L satisfies the norm bound

$$|||L|||_{\mathbb{X}} = \gamma \cdot \sup_{\|v\| \leq 1} \|Pv\| \leq \gamma := \gamma_{\max},$$

since ξ is the stationary distribution of the transition kernel P.

2.4.3.1 Upper bounds on stochastic approximation with averaging

As mentioned before, this example is somewhat non-standard in that the basis functions ψ_i are not necessarily orthonormal; indeed the classical *temporal difference* (TD) learning update in $\mathbb{R}^d$ involves the stochastic approximation algorithm

$$\vartheta_{t+1} = \vartheta_t - \eta\Big(\psi(s_{t+1})\psi(s_{t+1})^\top \vartheta_t - \gamma\psi(s_{t+1})\psi(s^+_{t+1})^\top \vartheta_t - R_{t+1}(s_{t+1})\psi(s_{t+1})\Big). \tag{2.41a}$$

The Polyak–Ruppert averaged estimator is then given by the relations

$$\widehat{\vartheta}_n = \frac{2}{n}\sum_{t=n/2}^{n-1} \vartheta_t, \quad \text{and} \quad \widehat{v}_n := \widehat{\vartheta}_n^\top \psi. \tag{2.41b}$$

Note that the updates (2.22) are, strictly speaking, different from the canonical iterates (2.25), but this should not be viewed as a fundamental difference since we are ultimately interested in the value function iterates $\widehat{v}_n$; these are obtained from the iterates $\widehat{\vartheta}_n$ by passing back to the original Hilbert space.

Nevertheless, this cosmetic difference necessitates some natural basis transformations before stating our results. Define the matrix[7] $B \in \mathbb{R}^{d \times d}$ by $B_{ij} := \langle \psi_i, \psi_j \rangle$ for $i, j \in [d]$; this defines an orthonormal basis given by

$$\begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_d \end{bmatrix} := \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_d \end{bmatrix} B^{-1/2}.$$

We define the *min/max eigenvalues* $\beta := \lambda_{\max}(B)$, and $\mu := \lambda_{\min}\Big(B\Big)$, so that β/μ is the condition number of the covariance matrix of the features.

Having set up this transformation, we are now ready to state the implication of our main theorem to the case of LSTD problems. We assume the following fourth-moment condition:

$$\mathbb{E}_\xi\Big[R^4(s)\Big] \leq \varsigma^4, \quad \text{and} \quad \mathbb{E}_\xi\Big(u^\top B^{-1/2}\psi(s)\Big)^4 \leq \varsigma^4 \quad \text{for all } u \in \mathbb{S}^{d-1}. \tag{2.42}$$

As verified in the proof of Corollary 2.5 to follow, equation (2.42) suffices to guarantee that Assumption 2.1(W) is satisfied with parameters $(\sigma_L, \sigma_b) = (2\varsigma^2, \varsigma^2/\sqrt{\beta})$. We also

[7] Since the functions ψ_i are linearly independent, we have $B \succ 0$.

define the matrices

$$\begin{aligned} M &:= \gamma B^{-1/2}\mathbb{E}_\xi[\psi(s)\psi(s^+)^\top]B^{-1/2}, \quad \text{and} \\ \Sigma^* &:= \operatorname{cov}_\xi\Big[B^{-1/2}\psi(s)\Big(\psi(s) - \gamma\psi(s^+) - R(s)\Big)^\top \bar{\vartheta}\Big]. \end{aligned}$$

The following corollary then provides a guarantee on the Polyak–Ruppert averaged TD(0) iterates (2.41).

Corollary 2.5. *Under the set-up above, there are universal positive constants* (c, c_0) *such that given a sample size* $n \geq \frac{c_0 \varsigma^4 \beta^2 d}{\mu^2(1-\kappa(M))^2} \log^2\Big(\frac{\|v_0 - \bar{v}\|_2^2 \beta d}{\mu(1-\kappa(M))}\Big)$, *and when the stochastic approximation scheme* (2.41a) *is run with step size* $\eta = \frac{1}{c_0\varsigma^2\beta\sqrt{dn}}$, *the averaged iterates satisfy the bound*

$$\begin{aligned} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \leq (1+\omega)&\alpha(M,\gamma)\mathcal{A}(\mathbb{S}, v^*) \\ &+ c\big(1 + \tfrac{1}{\omega}\big)\Big[\mathcal{E}_n(M, \Sigma^*) + \big(1 + \|\bar{v}\|^2\big)\big(\tfrac{\varsigma^2\beta}{(1-\kappa(M))\mu}\sqrt{\tfrac{d}{n}}\big)^3\Big] \qquad (2.43) \end{aligned}$$

for any $\omega > 0$.

See Appendix A.8.1 for the proof of this corollary.

In the worst case, the approximation factor $\alpha(M,\gamma)$ scales as $\frac{1}{1-\gamma^2}$, recovering the classical result (2.6). More generally, it gives a fine-grained characterization of the approximation factor depending on the one-step auto-covariance matrix for the feature vectors. By Lemma 2.1, we have $\alpha(M,\gamma) \leq O\big(\frac{1}{1-\kappa(M)}\big)$, so intuitively, the approximation factor is large when there are feature space directions in which the Markov chain transitions slowly. On the other hand, if the one-step-transitions move rapidly in all directions of feature space, then the approximation factor is much smaller.

The statistical error term $\mathcal{E}_n(M, \Sigma^*)$ matches the Cramér–Rao lower bound, and gives a finer characterization than both worst-case upper bounds [16], as well as existing instance-dependent upper bounds [115]. Note that the final higher-order term depends on the condition number $\frac{\beta}{\mu}$ of the covariance matrix B. This ratio is 1 when the basis vectors are orthonormal, but in general, the speed of algorithmic convergence depends on this parameter.

2.4.3.2 Approximation factor lower bounds for MRPs

We conclude our discussion of discounted MRPs with an information-theoretic lower bound for policy evaluation. This bound involves technical effort beyond that in the proof of Theorem 2.2, since any valid construction for MRPs must make use only of operators L that are constructed using a valid transition kernel. To set the stage, we say that a Markov reward process (P, γ, r) and associated basis functions $\{\psi_j\}_{j=1}^d$ are in the *canonical set-up* if

- The stationary distribution ξ of P exists and is unique.

- The reward function and its observations are uniformly bounded. In particular, we have $\|r\|_\infty \leq 1$, and $\|R\|_\infty \leq 1$ almost surely.
- The basis functions are orthonormal, i.e., $\mathbb{E}_\xi[\psi(s)\psi(s)^\top] = I_d$.

The three conditions are standard assumptions in Markov reward processes.

Now given scalars $\nu \in (0, 1]$ and $\gamma \in (0, 1)$, integer $D > 0$ and scalar $\delta \in (0, 1/2)$, we consider the following class of MRPs and associated feature vectors:

$$\mathbb{C}_{\mathsf{MRP}}\Big(\nu, \gamma, D, \delta\Big) := \Big\{(P, \gamma, r, \psi) \;\Big|\; \begin{array}{c} (P, \gamma, r, \psi) \text{ is in the canonical setup}, \quad |\mathcal{S}| = D, \\ \mathcal{A}(\mathbb{S}, v^*) \leq \delta^2, \quad \kappa\Big(\mathbb{E}_\xi[\psi(s)\psi(s^+)^\top]\Big) \leq \nu. \end{array} \Big\}.$$

Note that under the canonical set-up, we have $M = \gamma\mathbb{E}_\xi[\psi(s)\psi(s^+)^\top]$, and consequently, a problem instance in the class $\mathbb{C}_{\mathsf{MRP}}(\nu, \gamma, D, \delta)$ satisfies $\kappa(M) \leq \nu\gamma$ in the set-up of Corollary 2.5. The condition $\kappa\Big(\mathbb{E}_\xi[\psi(s)\psi(s^+)^\top]\Big) \leq \nu$ can be seen as a "mixing" condition in the projected space: when ν is bounded away from 1, the feature vector cannot have too large a correlation with its next-step transition in any direction.

We have the following minimax lower bound for this class, where we use the shorthand $\mathbb{C}_{\mathsf{MRP}} \equiv \mathbb{C}_{\mathsf{MRP}}\Big(\nu, \gamma, D, \delta\Big)$ for convenience.

Proposition 2.1. *There are universal positive constants (c, c_1) such that if $D \geq c_1(n^2 + d)$, then for all scalars $\nu \in (0, 1]$ and $\gamma \in (0, 1)$, we have*

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{(P,\gamma,r,\psi) \in \mathbb{C}_{\mathsf{MRP}}} \|\widehat{v}_n - v^*\|^2 \geq \frac{c}{1 - \nu\gamma}\delta^2 \wedge 1. \tag{2.44}$$

See Appendix A.8.2 for the proof of this proposition.

A few remarks are in order. First, in conjunction with Corollary 2.5 and the second upper bound in Lemma 2.1, we can conclude that the TD algorithm for policy evaluation with linear function approximation attains the minimax-optimal approximation factor over the class $\mathbb{C}_{\mathsf{MRP}}$ up to universal constants, in the regime where the optimal error is bounded by $O(1)$. It is also worth noting that Proposition 2.1 also shows that the worst-case upper bound (2.6) due to Tsitsiklis and Van Roy [204] is indeed sharp up to a universal constant; indeed, note that for all $\gamma \in (0, 1)$, we have $\frac{1}{1-\gamma^2} \asymp \frac{1}{1-\gamma}$, and that the latter factor can be obtained from the lower bound (2.44) by taking $\nu = 1$.

Second, note that the class $\mathbb{C}_{\mathsf{MRP}}$ is defined in a more "global" sense, as opposed to the "local" class $\mathbb{C}_{\mathsf{approx}}$ used in Theorem 2.2. This class contains all the MRP instances satisfying the approximation error bound and the constraint on $\kappa(M)$, and a minimax lower bound over this larger class is weaker than the lower bound over the local class that imposes restrictions on the projected matrix. That being said, Proposition 2.1 still captures more structure in the Markov transition kernel than the fact that it is contractive in the ξ-norm. For example, when the Markov chain makes "local moves" in the feature space, the correlation between feature vectors can be large, leading to large value of ν and larger values of optimal approximation factor. On the other hand, if the one-step transition of the feature vector jumps a large distance in all directions, the optimal approximation factor will be small.

Finally, it is worth noticing that Proposition 2.1 holds only for the i.i.d. observation models. If we are given the entire trajectory of the Markov reward process, the approximation factor can be made arbitrarily close to 1, using TD(λ) methods [204]. The trade-off inherent to the Markov observation model is an important direction of future work.

2.5 Proofs

We now turn to the proofs of our main results. The main proofs of Theorems 2.1 and 2.2 are given in this section, with some technical lemmas deferred to Appendix A. The proofs of Theorems 2.3 and 2.4, Corollaries 2.1 and 2.2, as well as associated lemmas, are presented in Appendix A

2.5.1 Proof of Theorem 2.1

We divide the proof into two parts, corresponding to the two components in the mean-squared error of the estimator $\widehat{v}_n$. The first term is the *approximation error* $\|\bar{v} - v^*\|^2$ that arises from the difference between the exact solution v^* to the original fixed point equation, and the exact solution $\bar{v}$ to the projected set of equations. The second term is the *estimation error* $\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2$, measuring the difficulty of estimating $\bar{v}$ on the basis of n noisy samples.

In particular, under the conditions of the theorem, we prove that the approximation error is upper bounded as

$$\|\bar{v} - v^*\|^2 \leq \alpha(M, \vert\!\vert\!\vert L \vert\!\vert\!\vert_{\mathbb{X}}) \inf_{v \in \mathbb{S}} \|v - v^*\|^2, \tag{2.45a}$$

whereas the estimation error is bounded as

$$\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \leq c \frac{\text{trace}\left((I - M)^{-1}\Sigma^*(I - M)^{-\top}\right)}{n} + c\frac{\sigma_L}{(1-\kappa)^3}\left(\frac{d}{n}\right)^{3/2}\left(\|\bar{v}\|^2\sigma_L^2 + \sigma_b^2\right). \tag{2.45b}$$

Given these two inequalities, it is straightforward to prove the bound (2.28) stated in the theorem. By expanding the square, we have

$$\begin{aligned}
\mathbb{E}\|\widehat{v}_n - v^*\|^2 &= \mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 + \|\bar{v} - v^*\|^2 + 2\mathbb{E}\langle \widehat{v}_n - \bar{v}, \, \bar{v} - v^* \rangle \\
&\overset{(i)}{\leq} \mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 + \|\bar{v} - v^*\|^2 + 2\sqrt{\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \cdot \|\bar{v} - v^*\|^2} \\
&\overset{(ii)}{\leq} \mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 + \|\bar{v} - v^*\|^2 + \tfrac{1}{\omega}\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 + \omega\|\bar{v} - v^*\|^2 \\
&= (1 + \omega)\|\bar{v} - v^*\|^2 + (1 + \tfrac{1}{\omega})\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2
\end{aligned}$$

where step (i) follows from the Cauchy–Schwarz inequality; and step (ii) follows from the arithmetic-geometric mean inequality, and is valid for any $\omega > 0$. Substituting the bounds from equations (2.45a) and (2.45b) yields the claim of the theorem.

The remainder of our argument is devoted to the proofs of the bounds (2.45a) and (2.45b).

2.5.1.1 Proof of approximation error bound (2.45a)

We begin with some decomposition relations for vectors and operators. Note that $\mathbb{S}$ is a finite-dimensional subspace, and therefore is closed. We use

$$\mathbb{S}^\perp := \{u \in \mathbb{X} \mid \langle u, \, v\rangle = 0 \mid \text{for all } v \in \mathbb{S}.\}$$

to denote its orthogonal complement. The pair $(\mathbb{S}, \mathbb{S}^\perp)$ forms a direct product decomposition of $\mathbb{X}$, and the projection operators satisfy $\Pi_{\mathbb{S}} + \Pi_{\mathbb{S}^\perp} = I$. Also define the operators $L_{\mathbb{S},\mathbb{S}} = \Pi_{\mathbb{S}} L \Pi_{\mathbb{S}}$ and $L_{\mathbb{S},\perp} = \Pi_{\mathbb{S}} L \Pi_{\mathbb{S}^\perp}$. With this notation, our proof can be broken down into two auxiliary lemmas, which we state here:

Lemma 2.3. *The error $\|\bar{v} - v^*\|$ between the projected fixed point $\bar{v}$ and the original fixed point v^* is bounded as*

$$\|\bar{v} - v^*\|^2 \leq \Big(1 + |\!|\!|(I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp}|\!|\!|_{\mathbb{X}}^2\Big) \inf_{v \in \mathbb{S}} \|v - v^*\|^2. \tag{2.46}$$

Lemma 2.4. *Under the set-up above, we have*

$$|\!|\!|(I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp}|\!|\!|_{\mathbb{X}}^2 \leq \lambda_{max}\Big((I_d - M)^{-1}\Big(|\!|\!|L|\!|\!|_{\mathbb{X}}^2 I_d - MM^\top\Big)(I_d - M)^{-\top}\Big).$$

The claimed bound (2.45a) on the approximation error follows by combining these two lemmas, and recalling our definition of $\alpha(M, L)$. We now prove these two lemmas in turn.

2.5.1.2 Proof of Lemma 2.3

For any vector $v \in \mathbb{X}$, we perform the orthogonal decomposition $v = v_{\mathbb{S}} + v_\perp$, where $v_{\mathbb{S}} := \Pi_{\mathbb{S}}(v)$ is a member of the set $\mathbb{S}$, and $v_\perp := \Pi_{\mathbb{S}^\perp, \xi}$ is a member of the set $\mathbb{S}^\perp$. With this notation, the operator L can be decomposed as

$$L = (\Pi_{\mathbb{S}} + \Pi_{\mathbb{S}^\perp}) L (\Pi_{\mathbb{S}} + \Pi_{\mathbb{S}^\perp}) \;=\; \underbrace{\Pi_{\mathbb{S}} L \Pi_{\mathbb{S}}}_{=:L_{\mathbb{S},\mathbb{S}}} + \underbrace{\Pi_{\mathbb{S}} L \Pi_{\mathbb{S}^\perp}}_{=:L_{\mathbb{S},\perp}} + \underbrace{\Pi_{\mathbb{S}^\perp} L \Pi_{\mathbb{S}}}_{=:L_{\perp,\mathbb{S}}} + \underbrace{\Pi_{\mathbb{S}^\perp} L \Pi_{\mathbb{S}^\perp}}_{=:L_{\perp,\perp}}.$$

The four operators $L_{\mathbb{S},\mathbb{S}}, L_{\mathbb{S},\perp}, L_{\perp,\mathbb{S}}, L_{\perp,\perp}$ defined in the equation above are also bounded linear operators. By the properties of projection operators, we note that $L_{\mathbb{S},\mathbb{S}}$ and $L_{\perp,\mathbb{S}}$ both map each element of $\mathbb{S}^\perp$ to 0, and $L_{\mathbb{S},\perp}$ and $L_{\perp,\perp}$ both map each element of $\mathbb{S}$ to 0.

Decomposing the target vector v^* in an analogous manner yields the two components

$$\widetilde{v} := \Pi_{\mathbb{S}}(v^*), \quad \text{and} \quad v^\perp := v^* - \widetilde{v}.$$

The fixed point equation $v^* = Lv^* + b$ can then be written using $\mathbb{S}$ and its orthogonal complement as

$$\widetilde{v} \overset{(a)}{=} L_{\mathbb{S},\mathbb{S}}\widetilde{v} + L_{\mathbb{S},\perp}v^\perp + b_{\mathbb{S}}, \quad \text{and} \quad v^\perp \overset{(b)}{=} L_{\perp,\mathbb{S}}\widetilde{v} + L_{\perp,\perp}v^\perp + b_\perp. \tag{2.47}$$

For the projected solution $\bar{v}$, we have the defining equation

$$\bar{v} = L_{\mathbb{S},\mathbb{S}}\bar{v} + b_{\mathbb{S}}. \tag{2.48}$$

Subtracting equation (2.47)(a) from equation (2.48) yields

$$(I - L_{\mathbb{S},\mathbb{S}})(\widetilde{v} - \bar{v}) = L_{\mathbb{S},\perp}v^\perp.$$

Recall the quantity $M = \Phi_d L \Phi_d^*$, and our assumption that $\kappa(M) = \frac{1}{2}\lambda_{\max}(M + M^T) < 1$. This condition implies that $I - L_{\mathbb{S},\mathbb{S}}$ is invertible on the subspace $\mathbb{S}$. Since this operator also maps each element of $\mathbb{S}^\perp$ to itself, it is invertible on all of $\mathbb{X}$, and we have $\widetilde{v} - \bar{v} = (I - L_{\mathbb{S},\mathbb{S}})^{-1}L_{\mathbb{S},\perp}v^\perp$.

Applying the Pythagorean theorem then yields

$$\begin{aligned}\|\bar{v} - v^*\|^2 \;&=\; \|\bar{v} - \widetilde{v}\|^2 + \|\widetilde{v} - v^*\|^2 = \|(I - L_{\mathbb{S},\mathbb{S}})^{-1}L_{\mathbb{S},\perp}v^\perp\|^2 + \|v^\perp\|^2 \\ &\leq \Big(1 + |\!|\!|(I - L_{\mathbb{S},\mathbb{S}})^{-1}L_{\mathbb{S},\perp}|\!|\!|_{\mathbb{X}}^2\Big) \cdot \|v^\perp\|^2,\end{aligned} \tag{2.49}$$

as claimed.

2.5.1.3 Proof of Lemma 2.4

By the definition of operator norm for any vector $v \in \mathbb{X}$ such that $\|v\| = 1$, we have

$$|\!|\!|L|\!|\!|_{\mathbb{X}}^2 \geq \|Lv\|^2 = \|L_{\mathbb{S},\mathbb{S}}v_{\mathbb{S}} + L_{\mathbb{S},\perp}v_\perp\|^2 + \|L_{\perp,\mathbb{S}}v_{\mathbb{S}} + L_{\perp,\perp}v_\perp\|^2 \geq \|L_{\mathbb{S},\mathbb{S}}v_{\mathbb{S}} + L_{\mathbb{S},\perp}v_\perp\|^2.$$

Noting the fact that $L_{\mathbb{S},\mathbb{S}}v_\perp = 0 = L_{\mathbb{S},\perp}v_{\mathbb{S}}$, we have the following norm bound on the linear operator $L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp}$:

$$\begin{aligned}|\!|\!|L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp}|\!|\!|_{\mathbb{X}} &= \sup_{\|v\|=1} \|(L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp})v\| \\ &= \sup_{\|v\|=1} \|L_{\mathbb{S},\mathbb{S}}v_{\mathbb{S}} + L_{\mathbb{S},\perp}v_\perp\| \leq |\!|\!|L|\!|\!|_{\mathbb{X}}.\end{aligned}$$

By definition, the operator $L_{\mathbb{S},\perp}^* = \Pi_{\mathbb{S}^\perp}L^*\Pi_{\mathbb{S}}$ maps any vector to $\mathbb{S}^\perp$, and the operator $L_{\mathbb{S},\mathbb{S}}$ maps any element of $\mathbb{S}^\perp$ to 0. Therefore, we have the identity $L_{\mathbb{S},\mathbb{S}}L_{\mathbb{S},\perp}^* = 0$. A similar argument yields that $L_{\mathbb{S},\perp}L_{\mathbb{S},\mathbb{S}}^* = 0$. Consequently, we have

$$\begin{aligned}|\!|\!|L|\!|\!|_{\mathbb{X}}^2 \geq |\!|\!|L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp}|\!|\!|_{\mathbb{X}}^2 &= |\!|\!|(L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp})(L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp})^*|\!|\!|_{\mathbb{X}} \\ &= |\!|\!|\underbrace{L_{\mathbb{S},\mathbb{S}}L_{\mathbb{S},\mathbb{S}}^* + L_{\mathbb{S},\perp}L_{\mathbb{S},\perp}^*}_{=:G}|\!|\!|_{\mathbb{X}}.\end{aligned} \tag{2.50}$$

Note that the operator G can be expressed as $G = \Pi_{\mathbb{S}}\Big(L\Pi_{\mathbb{S}}L^* + L\Pi_{\mathbb{S}^\perp}L^*\Big)\Pi_{\mathbb{S}}$. From this representation, we see that:

- For any vector $x \in \mathbb{X}$, we have $Gx \in \mathbb{S}$.
- For any vector $y \in \mathbb{S}^\perp$, we have $Gy = 0$.

Consequently, there exists a matrix $\widetilde{G} \in \mathbb{R}^{d\times d}$ such that $G = \Phi_d^*\widetilde{G}\Phi_d$. Since G is a positive semi-definite operator, the matrix $\widetilde{G}$ is positive semi-definite. Equation (2.50) implies that

$$\lambda_{\max}(\widetilde{G}) = \|\widetilde{G}\|_{\mathrm{op}} = \|G\|_{\mathbb{X}} \leq \|L\|_{\mathbb{X}}^2. \tag{2.51a}$$

Now defining $\tau := \|(I - L_{\mathbb{S},\mathbb{S}})^{-1}L_{\mathbb{S},\perp}\|_{\mathbb{X}}$, note that

$$\tau^2 = \| \underbrace{(I - L_{\mathbb{S},\mathbb{S}})^{-1}L_{\mathbb{S},\perp}L_{\mathbb{S},\perp}^*(I - L_{\mathbb{S},\mathbb{S}}^*)^{-1}}_{=:H} \|_{\mathbb{X}}. \tag{2.51b}$$

Moreover, the operator H is self-adjoint, and we have the following properties:

- The operator $L_{\mathbb{S},\perp}$ maps any vector to $\mathbb{S}$, and $(I - L_{\mathbb{S},\mathbb{S}})^{-1}$ maps $\mathbb{S}$ to itself. Consequently, for any $x \in \mathbb{X}$, the vector $Hx = (I - L_{\mathbb{S},\mathbb{S}})^{-1}L_{\mathbb{S},\perp}\Big(L_{\mathbb{S},\perp}^*(I - L_{\mathbb{S},\mathbb{S}}^*)^{-1}\Big)x$ is a member of the set $\mathbb{S}$.
- The operator $L_{\mathbb{S},\perp}^* = \Pi_{\mathbb{S}^\perp}L^*\Pi_{\mathbb{S}}$ maps any vector from $\mathbb{S}^\perp$ to 0. Consequently, for any $y \in \mathbb{S}^\perp$, we have $Hy = (I - L_{\mathbb{S},\mathbb{S}})^{-1}L_{\mathbb{S},\perp}\Big(L_{\mathbb{S},\perp}^*(I - L_{\mathbb{S},\mathbb{S}}^*)^{-1}\Big)y = 0$.

Given the facts above, there exists a matrix $\widetilde{H} \in \mathbb{R}^{d\times d}$ such that $H = \Phi_d^*\widetilde{H}\Phi_d$. Since the operator H is positive semi-definite, so is the matrix $\widetilde{H}$. Consequently, by equation (2.51b), we obtain the identity $\tau^2 = \|H\|_{\mathbb{X}} = \|\widetilde{H}\|_{\mathrm{op}} = \lambda_{\max}(H)$. In particular, letting $u \in \mathbb{S}^{d-1}$ be a maximal eigenvector of $\widetilde{H}$, we have

$$\widetilde{H} \succeq \tau^2 uu^\top. \tag{2.52}$$

Since $M = \Phi_d L_{\mathbb{S},\mathbb{S}}\Phi_d^*$ by definition, combining the above matrix inequalities (2.51a) and (2.52), we arrive at the bound:

$$\begin{aligned}
&\|L\|_{\mathbb{X}}^2 I_d \succeq \widetilde{G}\\
&= \Phi_d\Big(L_{\mathbb{S},\mathbb{S}}L_{\mathbb{S},\mathbb{S}}^* + L_{\mathbb{S},\perp}L_{\mathbb{S},\perp}^*\Big)\Phi_d^*\\
&= \Phi_d L_{\mathbb{S},\mathbb{S}}L_{\mathbb{S},\mathbb{S}}^*\Phi_d^*\\
&\qquad + \Big(\Phi_d(I - L_{\mathbb{S},\mathbb{S}})\Phi_d^*\Big)\cdot\Big(\Phi_d(I - L_{\mathbb{S},\mathbb{S}})^{-1}L_{\mathbb{S},\perp}L_{\mathbb{S},\perp}^*(I - L_{\mathbb{S},\mathbb{S}}^*)^{-1}\Phi_d^*\Big)\cdot\Big(\Phi_d(I - L_{\mathbb{S},\mathbb{S}}^*)\Phi_d^*\Big)\\
&= MM^\top + (I - M)\widetilde{H}(I - M^\top)\\
&\succeq MM^\top + \tau^2(I - M)uu^\top(I - M^\top).
\end{aligned}$$

Re-arranging and noting that $u \in \mathbb{S}^{d-1}$, we arrive at the inequality

$$\begin{aligned}\tau^2 &\leq u^\top \Big[(I-M)^{-1}(|\!|\!| L |\!|\!|_{\mathbb{X}}^2 I_d - MM^\top)(I-M)^{-\top}\Big] u \\ &\leq \lambda_{\max}\Big((I-M)^{-1}(|\!|\!| L |\!|\!|_{\mathbb{X}}^2 I_d - MM^\top)(I-M)^{-\top}\Big),\end{aligned}$$

which completes the proof of Lemma 2.4.

2.5.1.4 Proof of estimation error bound (2.45b)

We now turn to the proof of our claimed bound on the estimation error. Our analysis relies on two auxiliary lemmas. The first lemma provides bounds on the mean-squared error of the standard iterates $\{v_t\}_{t \geq 0}$—that is, without the averaging step:

Lemma 2.5. *Suppose that the noise conditions in Assumption 2.1(W) hold. Then for any stepsize $\eta \in \big(0, \frac{1-\kappa}{4\sigma_L^2 d + 1 + |\!|\!| L |\!|\!|_{\mathbb{X}}^2}\big)$, we have the bound*

$$\mathbb{E}\|v_t - \bar{v}\|^2 \leq e^{-(1-\kappa)\eta t/2}\mathbb{E}\|v_0 - \bar{v}\|^2 + \frac{8\eta}{1-\kappa}(\|\bar{v}\|^2 \sigma_L^2 d + \sigma_b^2 d) \qquad \textit{valid for } t = 1, 2, \ldots. \tag{2.53}$$

See Appendix A.1.1 for the proof of this claim.

Our second lemma provides a bound on the PR-averaged estimate $\widehat{v}_n$ based on n observations in terms of a covariance term, along with the error of the non-averaged sequences $\{v_t\}_{t \geq 1}$:

Lemma 2.6. *Under the setup above, we have the bound*

$$\begin{aligned}\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 &\leq \frac{3}{n-n_0}\operatorname{trace}\Big((I-M)^{-1}\Sigma^*(I-M)^{-\top}\Big) \\ &+ \frac{3}{(n-n_0)^2}\sum_{t=n_0}^{n}\mathbb{E}\|(I-M)^{-1}\Phi_d(L_{t+1}-L)(v_t - \bar{v})\|_2^2 + \frac{3\mathbb{E}\|v_n - v_{n_0}\|^2}{\eta^2(n-n_0)^2(1-\kappa)^2}.\end{aligned} \tag{2.54}$$

See Appendix A.1.2 for the proof of this claim.

Equipped with these two lemmas, we can now complete the proof of the claimed bound (2.45b) on the estimation error. Recalling that $n_0 = n/2$, we see that the first term in the bound (2.54) matches a term in the bound (2.45b). As for the remaining two terms in equation (2.54), the second moment bounds from Assumption 2.1(W) combined with the assumption that $\kappa(M) < 1$ imply that

$$\begin{aligned}\mathbb{E}\|(I-M)^{-1}\Phi_d(L_{t+1}-L)(v_t-\bar{v})\|_2^2 &\leq \frac{1}{(1-\kappa)^2}\mathbb{E}\|\Phi_d(L_{t+1}-L)(v_t-\bar{v})\|_2^2 \\ &\leq \frac{1}{(1-\kappa)^2}\sum_{j=1}^{d}\mathbb{E}\langle \phi_j,\, (L_{t+1}-L)(v_t-\bar{v})\rangle^2 \leq \frac{\sigma_L^2 d\|v_t-\bar{v}\|^2}{(1-\kappa)^2}.\end{aligned}$$

On the other hand, we can use Lemma 2.5 to control the third term in the bound (2.54). We begin by observing that

$$\|v_n - v_{n_0}\|^2 \leq 2\|v_n - \bar{v}\|^2 + 2\|v_{n_0} - \bar{v}\|^2 \leq 4 \sup_{n_0 \leq t \leq n} \mathbb{E}\|v_t - \bar{v}\|^2.$$

If we choose a burn-in time $n_0 > \frac{c_0}{(1-\kappa)\eta} \log \left(\frac{\|v_0 - \bar{v}\|^2 d}{1-\kappa} \right)$, then Lemma 2.5 ensures that

$$\sup_{n_0 \leq t \leq n} \mathbb{E}\|v_t - \bar{v}\|^2 \leq \frac{16\eta}{1-\kappa} \Big(\|\bar{v}\|^2 \sigma_L^2 d + \sigma_b^2 d \Big).$$

Finally, taking the step size $\eta = \frac{1}{24\sigma_L\sqrt{dn}}$, recalling that $n_0 = n/2$, and putting together the pieces yields

$$\begin{aligned}
&\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \\
&\leq \frac{12}{n} \operatorname{trace} \Big((I - M)^{-1} \Sigma^* (I - M)^{-\top} \Big) + \frac{1}{(1-\kappa)^2} \Big(\frac{12\sigma_L^2 d}{n} + \frac{48}{\eta^2 n^2} \Big) \sup_{n_0 \leq t \leq n} \mathbb{E}\|v_t - \bar{v}\|^2 \\
&\leq \frac{12}{n} \operatorname{trace} \Big((I - M)^{-1} \Sigma^* (I - M)^{-\top} \Big) + \frac{48\sigma_L}{(1-\kappa)^3} \Big(\frac{d}{n} \Big)^{3/2} \Big(\|\bar{v}\|^2 \sigma_L^2 + \sigma_b^2 \Big),
\end{aligned}$$

as claimed.

2.5.2 Proof of Theorem 2.2

At a high level, our proof of the lower bound proceeds by constructing two ensembles of problem instances that are hard to distinguish from each other, and such that the approximation error on at least one of them is large. The two instances are indexed by values of a bit $z \in \{-1, 1\}$, and each instance is, in turn, obtained as a mixture over 2^{D-d} centers; each center is indexed by a binary string $\varepsilon \in \{-1, 1\}^{D-d}$. The problem is then phrased as one of estimating the value of z from the observations; this is effectively a reduction to testing and the use of Le Cam's mixture-vs-mixture method.

Specifically, let $u \in \mathbb{S}^{d-1}$ be an eigenvector associated to the largest eigenvalue of the matrix $(I - M_0)^{-1} \big(\gamma_{\max}^2 I - M_0 M_0^\top \big) (I - M_0)^{-\top}$. By the definition of the approximation factor $\alpha(M_0, \gamma_{\max})$, we have:

$$\big(\alpha(M_0, \gamma_{\max}) - 1 \big) \cdot (I - M_0) u u^\top (I - M_0)^\top \preceq \gamma_{\max}^2 I - M_0 M_0^\top.$$

Based on the eigenvector u, we further define the d-dimensional vectors:

$$w := \sqrt{\alpha(M_0, \gamma_{\max}) - 1} \cdot (I - M_0) u, \quad \text{and} \quad y := \sqrt{\alpha(M_0, \gamma_{\max}) - 1} \cdot \delta u. \tag{2.55}$$

Substituting into the above PSD domination relation yields that

$$w w^\top + M_0 M_0^\top \preceq \gamma_{\max}^2 I. \tag{2.56}$$

Now consider the following class of (population) problem instances $(L^{(\varepsilon,z)}, b^{(\varepsilon,z)}, v^*_{\varepsilon,z})$ indexed by a binary string $\varepsilon \in \{-1,1\}^{D-d}$ and a bit $z \in \{-1,1\}$:

$$L^{(\varepsilon,z)} := \begin{bmatrix} M_0 & \frac{\sqrt{d}}{D-d}\varepsilon_{d+1}w & \cdots & \frac{\sqrt{d}}{D-d}\varepsilon_D w \\ 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \; v^*_{\varepsilon,z} := \begin{bmatrix} \sqrt{2d}\Big(zy + (I - M_0)^{-1}h_0\Big) \\ \sqrt{2}z\delta\varepsilon_{d+1} \\ \vdots \\ \sqrt{2}z\delta\varepsilon_D \end{bmatrix},$$

$$b^{(\varepsilon,z)} := (I - L^{(\varepsilon,z)})v^*_{\varepsilon,z} = \begin{bmatrix} \sqrt{2d}h_0 \\ \sqrt{2}z\delta\varepsilon_{d+1} \\ \vdots \\ \sqrt{2}z\delta\varepsilon_D \end{bmatrix}. \tag{2.57}$$

We take the weight vector ξ to be

$$\xi = \Big[\underbrace{\tfrac{1}{2d} \; \cdots \; \tfrac{1}{2d}}_{d} \;\; \underbrace{\tfrac{1}{2(D-d)} \; \cdots \; \tfrac{1}{2(D-d)}}_{(D-d)}\Big],$$

and the weighted inner product $\langle \cdot, \cdot \rangle$ on the space $\mathbb{X} = \mathbb{R}^D$ is defined via

$$\langle p,\, q \rangle := \sum_{j=1}^{D} p_j \xi_j q_j \qquad \text{for each pair } p, q \in \mathbb{R}^D.$$

This choice of inner product then induces the vector norm $\|\cdot\|$ and operator norm $|\!|\!|\cdot|\!|\!|_{\mathbb{X}}$.

Next, we define the basis vectors via

$$\phi_i = \begin{cases} \sqrt{2d}e_i & \text{for } i = 1, 2, \cdots, d, \text{ and} \\ \sqrt{2(D-d)}e_i & \text{for } i = d+1, \cdots D. \end{cases}$$

By construction, we have ensured that $\|\phi_i\| = 1$ for each $i \in [D]$. We let the subspace $\mathbb{S}$ be the span of the first d standard basis vectors, i.e., $\mathbb{S} := \text{span}(e_1, e_2, \cdots, e_d)$.

For each binary string $\varepsilon \in \{-1,1\}^{D-d}$ and signed bit $z \in \{-1,1\}$, a straightforward calculation reveals that the projected problem instance satisfies the identities

$$\Phi_d L^{(\varepsilon,z)} \Phi_d^* = M_0, \quad \text{and} \quad \Phi_d b^{(\varepsilon,z)} = h_0. \tag{2.58a}$$

Also note that for any pair (ε, z), we have by construction that

$$\inf_{v \in \mathbb{S}} \|v^*_{\varepsilon,z} - v\|^2 = \frac{1}{2(D-d)} \sum_{j=d+1}^{D} (\sqrt{2}z\delta\varepsilon_j)^2 = \delta^2. \tag{2.58b}$$

In words, this shows that the $\|\cdot\|$-error of approximating $v^*_{\varepsilon,z}$ with the linear subspace $\mathbb{S}$ is always δ, irrespective of which $\varepsilon \in \{-1,1\}^{D-d}$ and $z \in \{-1,1\}$ are chosen.

Next, we construct the random observation models for the i.i.d. observations, which are also indexed by the pair (ϵ, z). In particular, we construct the random matrix $L_i^{(\varepsilon,z)}$ and random vector $b_i^{(\varepsilon,z)}$ via

$$L_i^{(\varepsilon,z)} := \begin{bmatrix} M_0 & 0 & \cdots & 0 & \sqrt{d}\varepsilon_{\tau_L^{(i)}} w & 0 & \cdots & 0 \\ 0 & 0 & & & \cdots & & & 0 \\ & & \vdots & & & \vdots & & \\ 0 & 0 & & & \cdots & & & 0 \end{bmatrix}, \quad b_i^{(\varepsilon,z)} := \begin{bmatrix} \sqrt{2d}h_0 \\ 0 \\ \vdots \\ 0 \\ \sqrt{2}(D-d)z\delta\varepsilon_{\tau_b^{(i)}} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.59}$$

where the random indices $\tau_L^{(i)}$ and $\tau_b^{(i)}$ are chosen independently and uniformly at random from the set $\{d+1, d+2, \cdots, D\}$. By construction, we have ensured that for each $\varepsilon \in \{-1,1\}^{D-d}$ and $z \in \{-1,1\}$, the observations have mean

$$\mathbb{E}\Big[L_i^{(\varepsilon,z)}\Big] = L^{(\varepsilon,z)}, \quad \text{and} \quad \mathbb{E}\Big[b_i^{(\varepsilon,z)}\Big] = b^{(\varepsilon,z)}.$$

This concludes our description of the problem instances themselves. Since our proof proceeds via Le Cam's lemma, we require some more notation for product distributions and mixtures under this observation model. Let $\mathbb{P}^{(n)}_{\varepsilon,z}$ denote the n-fold product of the probability laws of the pair $\big(L_i^{(\varepsilon,z)}, b_i^{(\varepsilon,z)}\big)$. We also define the following mixture of product measures for each $z \in \{-1,1\}$:

$$\mathbb{P}^{(n)}_z := \frac{1}{2^{D-d}} \sum_{\varepsilon \in \{\pm 1\}^{D-d}} \mathbb{P}^{(n)}_{\varepsilon,z}.$$

We seek bounds on the total variation distance $d_{\mathrm{TV}}\Big(\mathbb{P}^{(n)}_1, \mathbb{P}^{(n)}_{-1}\Big)$.

With this setup, the following lemmas assert that (a) Our construction satisfies the conditions in Assumption 2.1(S), and (b) The total variation distance is small provided $n \lesssim \sqrt{D-d}$.

Lemma 2.7. *For each binary string $\varepsilon \in \{-1,1\}^{D-d}$ and bit $z \in \{-1,1\}$:*
(a) The population-level matrix $L^{(\varepsilon,z)}$ defined in equation (2.57) satisfies $\|\!|L^{(\varepsilon,z)}\|\!|_{\mathbb{X}} \leq \gamma_{\max}$.
(b) The random observations $\big(L_i^{(\varepsilon,z)}, b_i^{(\varepsilon,z)}\big)$ defined in equation (2.59) satisfies Assumption 2.1(S), for any scalar pair (σ_L, σ_b) such that $\sigma_L \geq \gamma_{\max}$ and $\sigma_b \geq \delta$.

Lemma 2.8. *Under the set-up above, we have* $d_{\mathrm{TV}}\left(\mathbb{P}_1^{(n)}, \mathbb{P}_{-1}^{(n)}\right) \leq \frac{12n^2}{D-d}$.

See Appendices A.2.1 and A.2.2 in the supplementary fileAppendix A for the proofs of the two lemmas, respectively.

Part (a) of Lemma 2.7 and equations (2.58a)–(2.58b) together ensure that population-level problem instance (L, b) we constructed belongs to the class $\mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max})$. Part (b) of Lemma 2.7 further ensures the probability distribution $\mathbb{P}_{L,b}$ belongs to the class $\mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)$. Lemma 2.8 ensures that the two mixture distributions corresponding to different choices of the bit z are close provided n is not too large. The final step in applying Le Cam's mixture-vs-mixture result is to show that the approximation error is large for at least one of the choices of the bit z. We carry out this step by splitting the rest of the proof into two cases, depending on whether or not we enforce that our estimator $\widehat{v}$ is constrained to lie in the subspace $\mathbb{S}$. Throughout, we use the decomposition $\widehat{v} = \begin{bmatrix} \widehat{v}_1 \\ \widehat{v}_2 \end{bmatrix}$, where $\widehat{v}_1 \in \mathbb{R}^d$ and $\widehat{v}_2 \in \mathbb{R}^{D-d}$. Also recall the definition of the vector y from equation (2.55).

Case I: $\widehat{v} \in \mathbb{S}$. This corresponds to the "proper learning" case where the estimator is restricted to take values in the subspace $\mathbb{S}$ and $\widehat{v}_2 = 0$. Note that for any $\varepsilon \in \{-1, 1\}^{D-d}$, we have

$$\|v^*_{\varepsilon,z} - \widehat{v}\|^2 = \|v^*_{\varepsilon,z} - \Pi_{\mathbb{S}}(v^*_{\varepsilon,z})\|^2 + \|v^*_{\varepsilon,z} - \widehat{v}\|^2 = \delta^2 + \frac{1}{2d}\|\widehat{v}_1 - \sqrt{2d}zy\|_2^2.$$

Therefore, for any $\varepsilon, \varepsilon' \in \{-1, 1\}^{D-d}$, the following chain of inequalities holds:

$$\begin{aligned} \frac{1}{2}\Big(\|v^*_{\varepsilon,1} - \widehat{v}\|^2 + \|v^*_{\varepsilon',-1} - \widehat{v}\|^2\Big) &= \delta^2 + \frac{1}{4d}\Big(\|\widehat{v}_1 - \sqrt{2d}y\|_2^2 + \|\widehat{v}_1 + \sqrt{2d}y\|_2^2\Big) \\ &= \delta^2 + \frac{1}{2d}\Big(\|\widehat{v}_1\|_2^2 + 2d\|y\|_2^2\Big) \\ &\geq \delta^2 + \|y\|_2^2 = \alpha(M_0, \gamma_{\max}) \cdot \delta^2. \end{aligned}$$

By Le Cam's lemma, we thus have

$$\begin{aligned} \inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{S}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 &\geq \alpha(M_0, \gamma_{\max})\delta^2 \cdot \Big(1 - d_{\mathrm{TV}}(\mathbb{P}_{-1}^{(n)}, \mathbb{P}_1^{(n)})\Big) \\ &\overset{(i)}{\geq} (1 - \omega) \cdot \alpha(M_0, \gamma_{\max}) \cdot \delta^2, \end{aligned}$$

where in step (i), we have applied Lemma 2.8 in conjunction with the inequality $D \geq d + \frac{12n^2}{\omega}$.

Case II: $\widehat{v} \notin \mathbb{S}$. This corresponds to the case of "improper learning" where the estimator can take values in the entire space $\mathbb{X}$. In this case, for any pair $\varepsilon, \varepsilon' \in \{-1, 1\}^{D-d}$, we obtain

$$\|v^*_{\varepsilon,1} - v^*_{\varepsilon',-1}\| \geq \left\| \begin{bmatrix} 2\sqrt{2d} y^\top & 0 & \cdots & 0 \end{bmatrix}^\top \right\| = 2\|y\|_2 = 2\delta\sqrt{\alpha(M_0, \gamma_{\max}) - 1}.$$

Applying triangle inequality and Young's inequality yields the bound

$$\begin{aligned} \frac{1}{2}(\|\widehat{v} - v^*_{\varepsilon,1}\|^2 + \|\widehat{v} - v^*_{\varepsilon',1}\|^2) &\geq \frac{1}{4}(\|\widehat{v} - v^*_{\varepsilon,1}\| + \|\widehat{v} - v^*_{\varepsilon',1}\|)^2 \\ &\geq \frac{1}{4}\|v^*_{\varepsilon,1} - v^*_{\varepsilon',-1}\|^2 \geq (\alpha(M_0, \gamma_{\max}) - 1) \cdot \delta^2. \end{aligned}$$

By Le Cam's lemma, we once again have

$$\begin{aligned} \inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 &\geq \big(\alpha(M_0, \gamma_{\max}) - 1\big) \cdot \delta^2 \cdot \Big(1 - d_{\mathrm{TV}}(\mathbb{P}^{(n)}_{-1}, \mathbb{P}^{(n)}_{1})\Big) \\ &\geq (1 - \omega) \cdot \big(\alpha(M_0, \gamma_{\max}) - 1\big) \cdot \delta^2. \end{aligned}$$

Putting together the two cases completes the proof.

2.6 Discussion

In this chapter, we studied methods for computing approximate solutions to fixed point equations in Hilbert spaces, using methods that search over low-dimensional subspaces of the Hilbert space, and operate on stochastic observations of the problem data. We analyzed a standard stochastic approximation scheme involving Polyak–Ruppert averaging, and proved non-asymptotic instance-dependent upper bounds on its mean-squared error. This upper bound involved a pure approximation error term, reflecting the discrepancy induced by searching over a finite-dimensional subspace as opposed to the Hilbert space, and an estimation error term, induced by the noisiness in the observations. We complemented this upper bound with an information-theoretic analysis, that established instance-dependent lower bounds for both the approximation error and the estimation error. A noteworthy consequence of our analysis is that the optimal approximation factor in the oracle inequality is neither unity nor constant, but a quantity depending on the projected population-level operator. As direct consequences of our general theorems, we showed oracle inequalities for three specific examples in statistical estimation: linear regression on a linear subspace, Galerkin methods for elliptic PDEs, and value function estimation via temporal difference methods in Markov reward processes.

The results of this chapter leave open a number of directions for future work:

- This chapter focused on the case of independently drawn observations. Another observation model, one which arises naturally in the context of reinforcement learning, is the Markov observation model. As discussed in Section 2.2.2.3, consider the problem with $L = \gamma P$ and $b = r$, where P is a Markov transition kernel, γ is the discount factor and r is the reward function. The observed states and rewards in this setup are given by a single trajectory of the Markov chain P, as opposed to being drawn i.i.d. from the stationary distribution. It is known [204] that the resolvent formalism (a.k.a. TD(λ)) leads to an improved approximation factor with larger $\lambda \in [0, 1)$. On the other hand, larger choices of λ may lead to larger variance and slower convergence for the stochastic approximation estimator, and a model selection problem exists (See Section 2.2 in the monograph [200] for a detailed discussion). It is an important future work to extend our fine-grained risk bounds to the case of TD(λ) methods with Markov data. Leveraging the instance-dependent upper and lower bounds, one can also design and analyze estimators that achieve the optimal trade-off.

- This chapter focused purely on oracle inequalities defined with respect to a subspace. However, the framework of oracle inequalities is far more general; in the context of statistical estimation, one can prove oracle inequalities for any star-shaped set with bounds on its metric entropy. (See Section 13.3 in the monograph [213] for the general mechanism and examples.) For all the three examples considered in Section 2.2.2, one might imagine approximating solutions using sets with nonlinear structure, such as those defined by ℓ_1-constraints, Sobolev ellipses, or the function class representable by a given family neural networks. An interesting direction for future work is to understand the complexity of projected fixed point equations defined by such approximating classes.

Chapter 3

Optimal and instance-dependent guarantees for Markovian stochastic approximations

Continuing the discussion in Chapter 2, we move on to the Markovian observation model. In this chapter, we study stochastic approximation procedures for approximately solving a d-dimensional linear fixed point equation based on observing a trajectory of length n from an ergodic Markov chain. We first exhibit a non-asymptotic bound of the order $t_{\mathrm{mix}} \frac{d}{n}$ on the squared error of the last iterate of a standard scheme, where t_{mix} is a mixing time. We then prove a non-asymptotic instance-dependent bound on a suitably averaged sequence of iterates, with a leading term that matches the local asymptotic minimax limit, including sharp dependence on the parameters (d, t_{mix}) in the higher order terms. We complement these upper bounds with a non-asymptotic minimax lower bound that establishes the instance-optimality of the averaged SA estimator. We derive corollaries of these results for policy evaluation with Markov noise—covering the TD(λ) family of algorithms for all $\lambda \in [0, 1)$—and linear autoregressive models. In combination with the optimal oracle inequalities in Chapter 2, our instance-dependent characterizations open the door to the design of fine-grained model selection procedures for hyperparameter tuning (e.g., choosing the value of λ when running the TD(λ) algorithm).

3.1 Introduction

Linear Z-estimation problems—in which we are interested in computing the fixed point of a linear system of equations—are widely used in many application domains, including reinforcement learning and approximate dynamic programming [14, 200], stochastic control and filtering [11, 20, 111], and time-series analysis [73]. In many of these applications, the data-generating mechanism is modeled using an underlying Markov chain. The resulting dependency among the observations presents challenges for algorithm design as well as statistical analysis. In this chapter, our goal is to provide an instance-dependent statistical analysis—one that captures the difficulty of the particular Z-estimation problem at hand—and to develop computationally efficient algorithms that match these fundamental limits.

A linear Z-estimation problem in $\mathbb{R}^d$ is specified by a fixed point equation of the form

$$\theta = \bar{L}\theta + \bar{b}, \tag{3.1}$$

where the matrix $\bar{L} \in \mathbb{R}^{d\times d}$ and the vector $\bar{b} \in \mathbb{R}^d$ are parameters of the problem. In settings of interest in this chapter, the problem parameters $(\bar{L}, \bar{b})$ are unknown, and we observe only a sequence $(L_t, b_t)_{t\geq 1}$ of noisy observations, generated according to a Markov process in the following manner. The Markov process generates a sequence $(s_t)_{t\geq 0}$ of states taking values in some underlying state space $\mathcal{S}$. This chain is assumed to be ergodic, with a unique stationary distribution ξ. The observed pair (L_{t+1}, b_{t+1}) at each time t depends on the current state s_t, and moreover, their expectations under the stationary distribution ξ are equal to their population-level counterparts $(\bar{L}, \bar{b})$.

This general formulation includes a number of special cases of interest. In the simplest setting, at each time t, we observe a matrix-vector pair of the form $L_{t+1} = \boldsymbol{L}(s_t)$ and $b_{t+1} = \boldsymbol{b}(s_t)$, where $\boldsymbol{L} : \mathcal{S} \to \mathbb{R}^{d\times d}$ and $\boldsymbol{b} : \mathcal{S} \to \mathbb{R}^d$ are deterministic mappings such that

$$\mathbb{E}_\xi\big[\boldsymbol{L}(s)\big] = \bar{L}, \quad \text{and} \quad \mathbb{E}_\xi\big[\boldsymbol{b}(s)\big] = \bar{b}. \tag{3.2a}$$

Many applications involve additional sources of randomness beyond that naturally associated with the Markov chain itself. In order to accommodate this possibility, we can consider observations of the form

$$L_{t+1} = \boldsymbol{L}_{t+1}(s_t), \quad \text{and} \quad b_{t+1} = \boldsymbol{b}_{t+1}(s_t). \tag{3.2b}$$

Here the mappings $\boldsymbol{L}_{t+1}$ and $\boldsymbol{b}_{t+1}$ are now allowed to be i.i.d. random, independent of s_t, but are required to be related to the deterministic mappings $\boldsymbol{L}$ and $\boldsymbol{b}$ via the relation

$$\mathbb{E}\big[\boldsymbol{L}_{t+1}(s)\big] = \boldsymbol{L}(s), \quad \mathbb{E}\big[\boldsymbol{b}_{t+1}(s)\big] = \boldsymbol{b}(s), \quad \text{for all } s \in \mathcal{S}. \tag{3.2c}$$

By the tower property of conditional expectation, equations (3.2a) and (3.2c) imply that $\boldsymbol{L}_{t+1}(s_t)$ and $\boldsymbol{b}_{t+1}(s_t)$ are unbiased estimates of $\bar{L}$ and $\bar{b}$, respectively.[1]

Stochastic approximation (SA) methods, dating back to the seminal work of [174], are standard iterative procedures for using data to approximately compute θ. These algorithms proceed in a streaming fashion: upon receiving each data point, an incremental update is made and the (averaged or) final iterate is returned in a single pass. In this way, each iteration of stochastic approximation incurs only mild computational and storage costs. Given these attractive computational properties, it is natural to ask if there are SA methods that also enjoy optimal statistical performance.

[1] However, equation (3.2c) does not require the observations to be conditionally unbiased.

In this chapter, we analyze the SA procedure based on the updates

$$\theta_{t+1} := (1-\eta)\theta_t + \eta(L_{t+1}\theta_t + b_{t+1}), \quad \text{for } t = 0, 1, \ldots \tag{3.3a}$$

$$\widehat{\theta}_n := \frac{1}{n-n_0}\sum_{t=n_0}^{n-1} \theta_t \quad \text{for } n = n_0+1, n_0+2, \ldots. \tag{3.3b}$$

Equation (3.3a) describes a standard stochastic approximation update with constant stepsize $\eta > 0$, whereas equation (3.3b) corresponds to an application of the Polyak–Ruppert averaging procedure [169, 186] to the iterates, with burn-in period n_0. When each matrix observation L_{t+1} has a constant rank independent of the dimension d—as is the case for temporal difference learning methods in reinforcement learning (see Section 3.2.2)—the SA method (3.3) can be implemented with $\mathcal{O}(d)$ computational and storage cost per iteration.

There is an extensive body of past work on stochastic approximation methods with Markov data. Here we provide an overview of the literature most germane to our contributions, and defer a more detailed review to Appendix B.1. Asymptotic convergence of SA procedures with Markovian data can be established using either the ODE method [20] or the Poisson equation method [11]. [204] analyzes the asymptotic convergence of SA in the specific context of temporal difference methods in reinforcement learning. Although asymptotic guarantees provide helpful guidance, it is often most useful to have non-asymptotic guarantees that account both for limited sample size and scale of modern problems, and for these reasons, non-asymptotic analysis of Markovian SA procedures has attracted much recent attention.

Assuming a mixing time bound on the Markov chain, a projected variant of linear SA was analyzed by [16], who established non-asymptotic rates that are near-optimal in their dependence on the sample size n. [194] analyzed the standard SA scheme without the projection step used in [16], and obtained the same convergence rate in both mean-squared error and higher moments. Under an appropriate Lyapunov function assumption on the Markov chain, [56] proved finite-time bounds for linear SA using stability properties of random matrix products. Variants and special cases of SA procedures with Markov data have also been studied, including two-time-scale algorithms [85], gradient-based optimization under Markov data [52], and estimation in auto-regressive models [159, 109].

Despite this encouraging progress to date, two important questions still remain open, and form the focus of this chapter:

- **Sample complexity in high dimensions:** The primary goal of non-asymptotic analysis is to provide guarantees on the estimation error that have an explicit dependence on the problem at hand, and that hold true for a reasonable range of values of the sample size n. For instance, suppose the linear Z-estimation problem in $\mathbb{R}^d$ is driven by an underlying Markov chain of mixing time t_{mix}. Then under natural noise assumptions, one should expect the mean-squared error to scale as $\mathcal{O}(t_{\mathrm{mix}} d/n)$, with this being the dominant term whenever $n \gtrsim t_{\mathrm{mix}} d$. Such an error bound is particularly important for sieve estimators, where the problem dimension d is adaptively chosen based on the sample size n. However, existing analyses of linear SA do not provide such tight dimension-dependence. Using the notation of this chapter, the estimation error bounds in the papers [194, 16] rely on a uniform upper bound on the operator norm of the stochastic matrix $L_{t+1}(s_t)$; this quantity scales linearly with dimension d in many applications. Consequently, the resulting bounds on the MSE have a sub-optimal dependence on dimension, which is unsatisfactory for high-dimensional problems. Similarly, the bounds in the papers [56, 108, 40] also exhibit a sub-optimal dependence on dimension. To the best of our knowledge, the question of whether linear SA succeeds under the minimal conditions on sample size—in particular, with n mildly larger than $d \cdot t_{\mathrm{mix}}$—remains open.

- **Instance-dependent optimality:** While many estimators may exhibit near-optimal statistical performance in the globally minimax (i.e., worst-case) sense, some of them perform significantly better than others when applied to practical problem instances. This phenomenon motivates the study of local (i.e., instance-dependent) performance in the non-asymptotic regime. Such results have recently been established for linear Z-estimation in the i.i.d. setting by existing literature [164, 122, 98] and Chapter 2 of this dissertation. In particular, Theorem 2.3 in Chapter 2 provides a non-asymptotic analogs of classical theory on local asymptotic minimaxity (c.f. [209]), which establishes lower bounds by looking at the worst-case family of instances in a local neighborhood of a given problem. In the Markov setting, two questions naturally arise: (1) What does it mean for an estimator to be locally optimal in a non-asymptotic sense? (2) Does the linear SA estimator (3.3) match the local lower bound for every problem instance?

3.1.1 Contributions and organization

The primary goal of this chapter is to resolve these challenges, and provide a sharp analysis of (averaged) linear SA algorithms. These answers are not merely of theoretical interest: they also provide important guidance for practice, such as in choosing algorithm parameters including the burn-in period and stepsize. In more detail:

- We perform a fine-grained analysis of linear SA and produce an upper bound on its statistical error that transparently tracks the dependence on problem-specific complexity as well as step-size. Furthermore, our bound holds true provided $n \gtrsim t_{\mathrm{mix}} \cdot d$, establishing that the algorithm does indeed attain a sharp sample complexity guarantee in high dimensions.

- In a complementary direction to our upper bounds, we show a local minimax lower bound with an appropriately defined notion of local neighborhood of Markov chains. This lower bound certifies the statistical optimality of the linear SA estimator, again in an instance-dependent sense.

- We derive consequences of our general analysis for temporal difference methods in reinforcement learning, demonstrating a key problem-dependent quantity in matching upper and lower bounds.

One technical aspect of our analysis is noteworthy. En route to establishing bounds with sharp dimension dependence, we introduce a careful "bootstrapping" argument: starting with a loose bound, we progressively refine it via the repeated application of certain self-bounding inequalities. We suspect that this method may be of independent interest in providing sharp analyses of other stochastic approximation methods.

The remainder of this chapter is organized as follows. We complete this section by introducing additional notation to be used throughout the chapter, and then providing a more detailed discussion of related work. In Section 3.2, we provide the basic problem set-up, discuss the underlying assumptions, and give some illustrative examples. Section 3.3 is devoted to the presentation of our main results, which include upper bounds on the estimation error of stochastic approximation procedures, along with local minimax lower bounds that apply to any estimator. In Section 3.4, we develop some consequences of these results for specific models, including policy evaluation in reinforcement learning and estimation in autoregressive models. Sections 3.5, 3.6 are devoted to the proofs of Proposition 3.1, Theorem 3.1, respectively. We conclude with a discussion in Section 3.7. The proof of Theorem 3.2 and some auxiliary results, as well as some corollaries, are postponed to the appendix.

Additional notations: Throughout the chapter, we use $\mathcal{F}_t := \sigma\big((b_i, L_i, s_i)_{i \leq t}\big)$ to denote the natural filtration induced by the Markov observations.

3.2 Problem set-up

Recall from our earlier set-up (cf. equation (3.1)) that we are interested in solving a fixed point equation of the form $\theta = \bar{L}\theta + \bar{b}$, based on noisy observations of the pair $(\bar{L}, \bar{b})$, as defined by the Markov observation model (3.2). We require that the matrix $\bar{L}$ satisfies the conditions

$$\kappa := \frac{1}{2}\lambda_{\max}\big(\bar{L} + \bar{L}^\top\big) < 1, \quad \text{and} \quad |\!|\!|\bar{L}|\!|\!|_{\mathrm{op}} \leq \gamma_{\max}. \tag{3.4}$$

3.2.1 Assumptions

We now introduce and discuss the remaining four assumptions that underlie our analysis.

3.2.1.1 Conditions on Markov chain

We first describe the conditions imposed on the underlying Markov chain in our observation model. Let $\{s_t\}_{t \geq 0}$ denote a trajectory drawn from a Markov chain with transition kernel P. We assume that this chain has a unique stationary distribution ξ, and impose the following mixing condition in Wasserstein-1 distance:

Assumption 3.1. *There exists a natural number t_{mix} and a universal constant $c_0 \geq 1$ such that for any $x, y \in \mathcal{S}$, for all $t = 0, 1, 2, \ldots$, we have the bounds*

$$\mathcal{W}_{1,\rho}(\delta_x P^{t_{\mathrm{mix}}}, \delta_y P^{t_{\mathrm{mix}}}) \overset{(a)}{\leq} \frac{1}{2}\rho(x, y), \quad \textit{and} \quad \mathcal{W}_{1,\rho}(\delta_x P^{t}, \delta_y P^{t}) \overset{(b)}{\leq} c_0 \rho(x, y). \tag{3.5}$$

We assume throughout that the chain is initialized with a sample $s_0 \sim \xi$ from the stationary distribution. Given that our mixing time bound guarantees exponential decay of the Wasserstein distance, this condition is mild: it can be removed by waiting $\mathcal{O}(t_{\mathrm{mix}})$ iterations for the process to mix.

3.2.1.2 Tail conditions on noise

In our observation model, the "noise" terms correspond to the differences $\boldsymbol{L}_{t+1}(s_t) - \boldsymbol{L}(s_t)$ and $\boldsymbol{L}(s_t) - \bar{L}$, along with analogous quantities for the vector b. Our second assumption imposes conditions on these noise variables. We consider separate conditions on these martingale $\boldsymbol{L}_{t+1}(s_t) - \boldsymbol{L}(s_t)$ and Markov $\boldsymbol{L}(s_t) - \bar{L}$ parts of the noise, as well as the b-noise analogues.

Assumption 3.2. *There exists an even integer $\bar{p} \in [2, +\infty]$ and non-negative constants σ_L and σ_b, such that for any positive even integer $p \leq \bar{p}$, scalar $t \geq 0$, vector $u \in \mathbb{S}^{d-1}$, and index $j \in \{1, \ldots, d\}$, we have*

$$\mathbb{E}\big[\langle e_j,\, \big(\boldsymbol{L}_{t+1}(s) - \boldsymbol{L}(s)\big)u\rangle^p \mid s\big] \leq p!\sigma_L^p, \quad \textit{and} \quad \mathbb{E}\big[\langle e_j,\, \boldsymbol{b}_{t+1}(s) - \boldsymbol{b}(s)\rangle^p \mid s\big] \leq p!\sigma_b^p, \tag{3.6a}$$

as well as

$$\mathbb{E}_{s\sim\xi}\big[\langle e_j,\, \big(\boldsymbol{L}(s) - \bar{L}\big)u\rangle^p\big] \leq p!\sigma_L^p, \quad \textit{and} \quad \mathbb{E}_{s\sim\xi}\big[\langle e_j,\, \boldsymbol{b}(s_t) - \bar{b}\rangle^p\big] \leq p!\sigma_b^p. \tag{3.6b}$$

Note that this assumption is mildest for $\bar{p} = 2$, and strongest for $\bar{p} = \infty$. In the latter case, when $\bar{p} = \infty$, the assumption requires L_{t+1} and b_{t+1} to be sub-exponential random variables in the standard coordinate directions (since $\log(p!) \leq p \log(p/2)$ by concavity of the log function). This condition covers, for instance, the case where L_{t+1} is the outer product of sub-Gaussian random vectors, as in temporal difference learning methods. In addition to accommodating this case, Assumption 3.2 also covers the heavier-tailed setting in which only finitely many moments exist. In particular, when $\bar{p} = 2$, the second moment assumption coincides with Assumption 2.1(W) made in Chapter 2.

An important quantity in our analysis is the *effective noise level* given by

$$\bar{\sigma} := \sup_{p\in[2,\bar{p}]} \sup_{j\in[d]} p^{-1}\Big(\mathbb{E}\big[\langle e_j,\, (\boldsymbol{L}_{t+1}(s_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}_{t+1}(s_t) - \bar{b})\rangle^p\big]\Big)^{1/p}. \tag{3.7}$$

Note that under Assumption 3.2, we have the upper bound $\bar{\sigma} \leq \sigma_L \|\bar{\theta}\|_2 + \sigma_b$.

3.2.1.3 Metric space conditions

For most of our analysis, we impose the following condition:

Assumption 3.3. *The metric space $(\mathcal{S}, \rho)$ has diameter at most one.*

Note that our assumption of unit diameter is arbitrary; boundedness suffices. In order to accommodate the general case, it suffices to rescale the parameters σ_L and σ_b.

When applying our theory to unbounded spaces (e.g., $\mathcal{S} = \mathbb{R}^d$), we use a truncation argument to show that there is an event over a reduced state space on which this condition holds with probability tending exponentially to 1. (See Appendix B.2 for the details of this argument.)

3.2.1.4 Lipschitz condition

Finally, we place a Lipschitz assumption—under the metric ρ—on the mapping from the metric space $\mathcal{S}$ to the stochastic operators. Given the Markov chain setup in the metric space $(\mathcal{S}, \rho)$, it is alluring to assume a dimension-free Lipschitz bound on the mappings $(\boldsymbol{L}_t, \boldsymbol{b}_t)$. However, as the space $\mathcal{S}$ has diameter bounded by 1, such Lipschitz constants typically depend on dimension for practical problems. Concretely, view the $\bar{L}$-scale parameters $(\kappa, \gamma_{\max})$ as constants and assume that the observations $\boldsymbol{L}_{t+1}(s_t)$ each have rank at most r. We then have

$$\mathbb{E}\big[\|\boldsymbol{L}_{t+1}(s_t)\|_{\mathrm{op}}\big] \geq \frac{\mathbb{E}\big[\|\boldsymbol{L}_{t+1}(s_t)\|_{\mathrm{nuc}}\big]}{r} \geq \frac{\mathrm{trace}\big(\mathbb{E}\big[\boldsymbol{L}_{t+1}(s_t)\big]\big)}{r} = \frac{\mathrm{trace}(\bar{L})}{r}. \tag{3.8}$$

The trace $\mathrm{trace}(\bar{L})$ typically scale as $\Theta(d)$, even in the “easy case” when $\bar{L}$ is a constant multiple of identity matrix.

So the Lipschitz constant for the mapping $\boldsymbol{L}_t : \mathcal{S} \to \mathbb{R}^{d \times d}$ is at least $\Omega(d)$. On the other hand, as a d-dimensional standard Gaussian random variable has norm $\Omega(\sqrt{d})$ with high probability, it is natural to assume the Lipschitz constant for the vector-valued mapping $\boldsymbol{b}_t : \mathcal{S} \to \mathbb{R}^d$ to be of order at least $\Omega(\sqrt{d})$. We therefore make the following assumption:

Assumption 3.4. *There exist constants $\sigma_L, \sigma_b > 0$ such that, almost surely for any $x, y \in \mathcal{S}$, we have*

$$\|\boldsymbol{L}_t(x) - \boldsymbol{L}_t(y)\|_{op} \leq \sigma_L d \cdot \rho(x, y) \quad \textit{and} \quad \|\boldsymbol{b}_t(x) - \boldsymbol{b}_t(y)\|_2 \leq \sigma_b \sqrt{d} \cdot \rho(x, y) \tag{3.9}$$

for all $t = 1, 2, \ldots$.

Note that in Assumption 3.4, we explicitly rescale the RHS of the inequalities with factors that depend on the problem dimension d, so that the pair (σ_L, σ_b) should indeed be viewed as dimension-free. The notation (σ_L, σ_b) is actually overloaded in Assumptions 3.2 and 3.4. In practice, we can take the maximum of the bounds in the two assumptions. Besides, as shown in Appendix B.2, for certain natural problem classes, Assumption 3.2 indeed implies Assumption 3.4 with discrete metric, up to logarithmic factors.

3.2.2 Some illustrative examples

Our assumptions cover a broad range of ergodic Markov chains, and the fixed-point equation (3.1) associated with their stationary distribution naturally arises from several problems. In this section, we describe a few concrete examples of our general setup. We first discuss the class of Markov chains satisfying our assumptions, and then describe the linear Z-estimators associated with such problems.

3.2.2.1 Examples of Markov chains

By varying our choice of the metric ρ, we recover several important classes of Markov chains that satisfy Assumptions 3.1 and 3.3.

- Consider a Markov chain defined on a countable state space $\mathcal{S}$, and consider the discrete metric $\rho(x, y) := \mathbf{1}_{x \neq y}$. In this context, Assumption 3.1 corresponds to mixing time bound in total variation—viz.

  $$d_{\mathrm{TV}}(\delta_x P^{t_{\mathrm{mix}}}, \delta_y P^{t_{\mathrm{mix}}}) \leq \tfrac{1}{2} \qquad \text{for all pairs } x, y \in \mathcal{S}.$$

  This mixing condition is satisfied for some finite t_{mix} when the Markov chain is irreducible, aperiodic and positive recurrent. Moreover, this metric space has unit diameter, so that Assumption 3.3 holds as well.

- As another example, consider the state space $\mathcal{S} = \mathbb{B}(0, 1) \subseteq \mathbb{R}^d$ equipped with the Euclidean metric $\rho(x, y) = \|x - y\|_2$. We can define a Markov chain on this space via the random evolution $X_{k+1} = \mathcal{T}_{k+1}(X_k)$, where the random non-linear operators $\{\mathcal{T}_k\}_{k \geq 1} \subseteq \mathcal{S}^{\mathcal{S}}$ are drawn i.i.d. from some distribution. We assume that the expected operator $\bar{\mathcal{T}} := \mathbb{E}[\mathcal{T}_1]$ satisfies the contraction condition $\|\bar{\mathcal{T}}(x) - \bar{\mathcal{T}}(y)\|_2 \leq \gamma \|x - y\|_2$ with some $\gamma < 1$. Assuming the stochastic operator $\mathcal{T}$ to be Lipschitz and to satisfy a second moment bound, this dynamical system satisfies the Wasserstein contraction condition under the Euclidean metric.

3.2.2.2 Examples of linear Z-estimators

We now describe some interesting examples of linear Z-estimators, to which we will return in later sections.

Example 3.1 (Approximate policy evaluation)**.** We begin by considering the temporal difference (TD) algorithm for approximate estimation of value functions. This problem arises in the context of Markov reward processes (MRPs), which are Markov chains that are augmented with a reward function $r : \mathcal{S} \to \mathbb{R}$. A trajectory from a Markov reward process is a sequence $\{(s_t, R_t)\}_{t \geq 0}$, where $\{s_t\}_{t \geq 0}$ is the Markov trajectory of states, and R_t is a random reward, corresponding to a conditionally unbiased estimate (given s_t) of the reward function value $r(s_t)$. Given a discount factor $\gamma \in [0, 1)$, the expected discount reward defines the *value function* $v^*(s) = \mathbb{E}\big[\sum_{t=0}^{\infty} \gamma^t R_t \mid s_0 = s\big]$.

This value function is connected to linear Z-estimators via the Bellman principle. Let P denote the transition operator of the Markov chain, and let ξ denote the stationary distribution. Note that the P maps the space $\mathbb{L}^2(\mathcal{S}, \xi)$ to itself. With this notation, the value function v^* is known to be the unique fixed point of the *Bellman evaluation equation*

$$v = \gamma P v + r. \tag{3.10}$$

In general, this equation is non-trivial to solve, especially given a limited trajectory length. In practice, it is standard to compute approximate solutions using linear basis expansions [27, 204], and this approach underlies the family of TD algorithms.

Let $\{\phi_j\}_{j=1}^d$ be a collection of linearly independent real-valued functions defined on the state space, and consider the linear subspace $\mathbb{S}$ of all functions of the form $V_\theta(s) = \sum_{j=1}^d \theta_j \phi_j(s)$. This subspace defines the *projected Bellman equation*

$$\bar{v} = \Pi_{\mathbb{S}}\big(\gamma P \bar{v} + r\big), \tag{3.11}$$

where $\Pi_{\mathbb{S}}$ is the orthogonal projection operator under $\mathbb{L}^2(\mathcal{S}, \xi)$.

By definition, the projected fixed point $\bar{v}$ can be written in the form $\bar{v}(s) = \sum_{j=1}^d \bar{\theta}_j \phi_j(s)$ for some vector $\bar{\theta} \in \mathbb{R}^d$. Denote the vector-valued mapping $\phi = [\phi_j]_{j=1}^d$, some simple calculations show that this parameter vector must satisfy the linear system

$$\Sigma_0 \bar{\theta} = \gamma \Sigma_1 \bar{\theta} + \mathbb{E}_{s \sim \xi}\big[R_0(s)\phi(s)\big], \tag{3.12}$$

where $\Sigma_0 = \mathbb{E}_{s\sim\xi}\big[\phi(s)\phi(s)^\top\big]$ is the second-moment matrix of $\phi(s)$ under the stationary distribution, and $\Sigma_1 = \mathbb{E}[\phi(s)\phi(s^+)^\top]$ is the cross-moment operator of the Markov chain. In defining this cross-moment, the expectation is taken over $s \sim \xi$ and $s^+ \sim P(s, \cdot)$.

This problem can be viewed within our framework by considering a Markov chain on the augmented state space $\Omega_t = (s_t, s_{t+1})$. Equation (3.12) defines a fixed point equation under the stationary distribution of this Markov chain. Define the minimum and maximum eigenvalues $\mu := \lambda_{\min}(\Sigma_0)$ and $\beta := \lambda_{\max}(\Sigma_0)$, along with the observation functions

$$\boldsymbol{b}_{t+1}(\Omega_t) = \tfrac{1}{\beta} R_t(s_t)\phi(s_t), \quad \text{and} \quad \boldsymbol{L}_{t+1}(\Omega_t) = I_d - \tfrac{1}{\beta}\big[\phi(s_t)\phi(s_t)^\top - \gamma\phi(s_t)\phi(s_{t+1})^\top\big]. \tag{3.13}$$

With these choices, the stochastic approximation procedure (3.3) is the widely used TD(0) algorithm. On the other hand, for a stationary Markov chain $(s_t)_{t\in\mathbb{Z}}$, the fixed-point equation $\bar{\theta} = \mathbb{E}\left[\boldsymbol{L}_{t+1}(\Omega_t)\right] \cdot \bar{\theta} + \mathbb{E}\left[\boldsymbol{b}_{t+1}(\Omega_t)\right]$ is equivalent to Eq (3.12). Note that though the expression for the mappings $\boldsymbol{b}_{t+1}$ and $\boldsymbol{L}_{t+1}$ depends on unknown parameter β, they can be absorbed into the stepsize choice, and the algorithm works well without such knowledge.

Typically, the Euclidean norm $\|\phi(s)\|_2$ of the feature vectors scales as $\sqrt{d}$, and under the stationary distribution ξ, the variance of any coordinate of $\phi(s)$ is of constant order. Under these conditions, the cross-moment matrix Σ_1 has operator norm of constant order. On the other hand, as for the random observations, we have the scalings $|\!|\!| L_{t+1} |\!|\!|_{\mathrm{op}} = \mathcal{O}(d)$ and $\|b_{t+1}\|_2 = \mathcal{O}(\sqrt{d})$, so that Assumptions 3.2 and 3.4 are satisfied. ♣

In the context of TD, it is natural to consider a *sieve estimator*. Given a collection of basis functions $\{\phi_j\}_{j=1}^{\infty}$, we can define the nested family $\mathbb{S}_1 \subset \mathbb{S}_2 \subset \cdots$, where $\mathbb{S}_d$ denotes the span of the sub-collection $\{\phi_j\}_{j=1}^{d}$. Here the choice of the sieve parameter d is key: larger values reduce the approximation error at the expense of increasing the estimation error. We discuss how this can be done in Section 3.4.

Another extension of the TD(0) algorithm—one that becomes feasible under the Markovian observation model—is the TD(λ) family of procedures. A fundamental question is how well the solution of the projected fixed-point equation (3.11) approximates the true value function V^*. Theorem 2.1 in Chapter 2 analyzes this quantity, and provides matching upper and lower bounds in the i.i.d. setting. However, the Markovian observation model actually allows this approximation error to be reduced, albeit at the cost of increased estimation error, as discussed in our next example.

Example 3.2 (Policy evaluation with TD(λ))**.** The family of TD(λ) algorithms is motivated by the following observation: since the value function v^* is the fixed point of Eq (3.10), it is also the fixed point of the composition of itself. Concretely, for any $k \geq 1$, we have:

$$v^* = (\gamma P)^k v^* + \sum_{j=0}^{k-1} (\gamma P)^j r.$$

For any $\lambda \in [0, 1)$, we take the weighted average of the above (infinite) collection of equations using exponentially-decaying weight $(1, \lambda, \lambda^2, \cdots)$, and obtain the following equation.

$$v = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k (\gamma P)^{k+1} v + \sum_{k=0}^{\infty} \lambda^k (\gamma P)^k r. \tag{3.14a}$$

The solution v^* to the equation (3.10) also solves Eq (3.14a).

Following the same route as TD(0), for a given subspace $\mathbb{S}$ of functions, we seek a solution $\bar{v}^{(\lambda)}$ to the projected fixed equation equation

$$\bar{v}^{(\lambda)} = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k \Pi_{\mathbb{S}} (\gamma P)^{k+1} \bar{v}^{(\lambda)} + \sum_{k=0}^{\infty} \lambda^k \Pi_{\mathbb{S}} (\gamma P)^k r, \tag{3.14b}$$

in which the operator P has been replaced by the projection $\Pi_{\mathbb{S}} P$. Although the fixed points of equation (3.14a) and the Bellman equation (3.10) coincide, the projected version (3.14b) has a different set of fixed points.

Since the value function $\bar{v}^{(\lambda)}$ lies in the linear space $\mathbb{S}$, it has a representation of the form $\bar{v}^{(\lambda)}(s) = \sum_{j=1}^{d} \bar{\theta}_j^{(\lambda)} \phi_j(s)$ for some coefficient vector $\bar{\theta}^{(\lambda)} \in \mathbb{R}^d$. From equation (3.14b), this vector must satisfy a linear system of the form

$$\Big[\sum_{k=0}^{\infty} (\lambda\gamma)^k \Sigma_k \Big] \bar{\theta}^{(\lambda)} = \Big[\sum_{k=0}^{\infty} (\lambda\gamma)^k \gamma \Sigma_{k+1} \Big] \bar{\theta}^{(\lambda)} + \sum_{k=0}^{\infty} (\lambda\gamma)^k \mathbb{E}\big[R_0(s_0) \phi(s_{-k}) \big], \tag{3.15}$$

where $\{s_k\}_{k=-\infty}^{\infty}$ is a stationary Markov chain following the transition kernel P, and we define $\Sigma_k = \mathbb{E}[\phi(s_{-k})\phi(s_0)^\top]$ for each integer k. As it should, when we set $\lambda = 0$, equation (3.15) reduces to the TD(0) update from equation (3.12).

In order to use stochastic approximation methods to solve this equation, we consider an augmented Markov process $(s_{t+1}, s_t, g_t)_{t \in \mathbb{Z}}$ in the space $\mathcal{S}^2 \times \mathbb{R}^d$, which evolves as

$$s_{t+1} \sim P(s_t, \cdot), \quad \text{and} \quad g_t = \phi(s_t) + \gamma\lambda g_{t-1}. \tag{3.16a}$$

If feature vectors $\phi(s_t)$ lie in a compact set almost surely, we have $g_t = \sum_{k=0}^{+\infty} (\gamma\lambda)^k \phi(s_{t-k})$. Let $\widetilde{\xi}$ be the stationary distribution of this augmented Markov chain.[2] In terms of an element $\Omega = (s, s^+, g)$ drawn according this stationary distribution, the fixed-point equation (3.14b) admits the succinct representation

$$\mathbb{E}_{\widetilde{\xi}}\big[g\phi(s)^\top\big]\bar{\theta}^{(\lambda)} = \gamma \mathbb{E}_{\widetilde{\xi}}\big[g\phi(s^+)^\top\big]\bar{\theta}^{(\lambda)} + \mathbb{E}_{\widetilde{\xi}}\big[R_0(s)g\big]. \tag{3.16b}$$

By choosing the observation functions

$$\boldsymbol{L}_{t+1}(\Omega_t) = I_d - \nu \cdot \big(g_t\phi(s_t)^\top - \gamma g_t \phi(s_{t+1})^\top\big), \quad \boldsymbol{b}_{t+1}(\Omega_t) = \nu \cdot R_t(s_t)\phi(s_t), \tag{3.16c}$$

for a scalar $\nu > 0$, this algorithm is a special case of our general set-up. In particular, by substituting the infinite-sum expression for the random variable g_t into Eq (3.16b), we obtain the projected linear equation (3.15) under the low-dimensional representation. See Section 3.4 for a more detailed verification of the assumptions needed to apply our main results for this problem. ♣

For our last example, we turn to a different class of problems involving vector autoregressive (VAR) models for time series [131].

Example 3.3 (Parameter estimation in autoregressive models)**.** An m-dimensional VAR model of order k describes the evolution of a random vector X_t as a k^{th}-order Markov process. The model is specified by a collection of $m \times m$ matrices $\{A_j^*\}_{j=1}^k$, and the random vector evolves according to the recursion

$$X_{t+1} = \sum_{j=1}^{k} A_j^* X_{t-j+1} + \varepsilon_{t+1}, \tag{3.17}$$

where the noise sequence $(\varepsilon_t)_{t \geq 0}$ is i.i.d. and zero-mean and supported on a bounded set.

Considering the $(k+1)$-fold tuple $\Omega_t = (X_{t+1}, X_t, \cdots, X_{t-k+1})$, the process $(\Omega_t)_{t \geq 0}$ is Markovian. Under appropriate stability assumptions on the model parameter, the process mixes rapidly under the $(k+1)m$-dimensional Euclidean metric. Let $\widetilde{\xi}$ denote

[2] Such a stationary distribution exists and is unique under suitable assumptions. See 3.4.2 for details.

its stationary distribution, and suppose for convenience that the chain is observed at stationarity.

In order to estimate the model parameters, we consider the following set of Yule–Walker estimation equations:

$$\mathbb{E}\big[X_{t+1}X_{t-\ell}^\top\big] = A_1^*\mathbb{E}\big[X_t X_{t-\ell}^\top\big] + A_2^*\mathbb{E}\big[X_{t-1}X_{t-\ell}^\top\big] + \cdots + A_k^*\mathbb{E}\big[X_{t-k+1}X_{t-\ell}^\top\big], \tag{3.18}$$

for $\ell = 0, 1, \cdots, k-1$.

These equations form a km^2-dimensional linear system for estimating km^2-dimensional parameters. Note that the parameters live in the space of matrix sequences, and so we slightly abuse our notation for simplicity: L denotes a linear operator from $\mathbb{R}^{k\times m\times m}$ to itself, and b is an element in $\mathbb{R}^{k\times m\times m}$. At the sample level, for any collection $A := \{A_j\}_{j=1}^k \in \mathbb{R}^{k\times m\times m}$ of system matrices, the stochastic observations are given by

$$\big[\boldsymbol{b}_{t+1}(\Omega_t)\big]_\ell = \nu\, X_{t+1}X_{t-\ell}^\top \quad \text{for } \ell = 0, 1, \ldots, k-1, \text{ and}$$
$$\big(\boldsymbol{L}_{t+1}(\Omega_t)\big)[A]_\ell = A_\ell - \nu \sum_{j=0}^{k-1} A_j X_{t-j} X_{t-\ell}^\top, \quad \text{for } \ell = 0, 1, \ldots, k-1.$$

Once again, the parameter ν is a scaling constant needed to fit into the fixed-point equation framework, and is absorbed into the stepsize choice of the algorithm. ♣

3.3 Main results

We now turn to the statement of our main results, beginning with our upper bounds in Section 3.3.1, followed by lower bounds in Section 3.3.2.

3.3.1 Instance-dependent upper bounds

In this section, we begin by stating some upper bounds (Theorem 3.1) on the behavior of the Polyak–Ruppert averaged SA scheme (3.3b). These bounds are instance-dependent, in the sense that they are specified in terms of an explicit function of the operator $\bar{L}$ and the fixed point $\bar{\theta}$. We then state a second result (Proposition 3.1) on the non-averaged iterates, which plays a key role in proving Theorem 3.1.

3.3.1.1 Instance-dependent bounds on the averaged iterates

For any state $s \in \mathcal{S}$, define the functions

$$\varepsilon_{\mathrm{MG}}(s) := (\boldsymbol{b}_1(s) - \boldsymbol{b}(s)) + (\boldsymbol{L}_1(s) - \boldsymbol{L}(s))\bar{\theta}, \quad \text{and} \quad \varepsilon_{\mathrm{Mkv}}(s) := \boldsymbol{b}(s) + \boldsymbol{L}(s)\bar{\theta} - \bar{\theta}.$$

Note that for a fixed state s, the quantity $\varepsilon_{\mathrm{MG}}(s)$ depends on the random variables $\boldsymbol{b}_1(s)$ and $\boldsymbol{L}_1(s)$, and so is a random vector, whereas by contrast, the quantity $\varepsilon_{\mathrm{Mkv}}(s)$

is deterministic. Letting $(\widetilde{s}_t)_{t=-\infty}^{\infty}$ be a stationary Markov chain under the transition kernel P, we then define the matrices

$$\Sigma^*{}_{\mathrm{MG}} := \mathbb{E}_{\xi}\big[\operatorname{cov}\big(\varepsilon_{\mathrm{MG}}(s) \mid s\big)\big], \quad \text{and} \quad \Sigma^*{}_{\mathrm{Mkv}} := \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(\widetilde{s}_t)\varepsilon_{\mathrm{Mkv}}(\widetilde{s}_0)^\top\big]. \tag{3.19}$$

Overall, the performance of our algorithm depends on the *matrix sum* $\Sigma^* := \Sigma^*{}_{\mathrm{MG}} + \Sigma^*{}_{\mathrm{Mkv}}$, as well as the effective noise variance $\bar{\sigma}^2$ defined in Eq (3.7). In terms of these quantities, we have the following guarantee:

Theorem 3.1. *Under Assumptions 3.1–3.3, suppose that we set the stepsize η and burn-in parameter n_0 as $\eta = \big(c(\sigma_L^2 d + \gamma_{\max}^2)(1-\kappa)n^2 t_{\mathrm{mix}}\big)^{-1/3}$ and $n_0 = \frac{1}{2}n$, where c is a suitably chosen universal constant. Then for any sample size n satisfying $\frac{n}{\log^2 n} \geq \frac{2t_{\mathrm{mix}}(\sigma_L^2 d + \gamma_{\max}^2)}{(1-\kappa)^2}\log(c_0 d)$, the Polyak–Ruppert estimate (3.3b) has MSE bounded as*

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_2^2\big] \leq \tfrac{c'}{n}\mathrm{Tr}\big((I-\bar{L})^{-1}(\Sigma^*{}_{\mathrm{MG}} + \Sigma^*{}_{\mathrm{Mkv}})(I-\bar{L})^{-\top}\big) + c'\big(\tfrac{\bar{\sigma}^2 d t_{\mathrm{mix}}}{(1-\kappa)^2 n}\big)^{4/3}\log^2 n. \tag{3.20}$$

See Section 3.6 for the proof of this theorem.

A few remarks are in order. First, and as shown in the next section, the first term $n^{-1}\mathrm{Tr}\big((I-\bar{L})^{-1}\Sigma^*(I-\bar{L})^{-1}\big)$ is optimal for the Markovian stochastic approximation problem in an instance-dependent sense. This term appears in existing central limit results for Markovian stochastic approximation [61], whereas our bound captures this dependence in a non-asymptotic manner. When the Markov chain is uniformly geometrically ergodic, a central limit theorem for the averaged iterate $\widehat{\theta}_n$ directly follow from classical Markovian CLT (see [143], Chapter 17).

The first term can always be upper bounded by $c'\frac{\bar{\sigma}^2}{(1-\kappa)^2 n}t_{\mathrm{mix}}d \cdot \log^2(c_0 d)$.[3] On the other hand, disregarding dependence on (σ_L, σ_b) and logarithmic factors in the sample size, the second term in the bound scales as $\mathcal{O}\big(\big(\frac{t_{\mathrm{mix}}d}{(1-\kappa)^2 n}\big)^{4/3}\big)$. Consequently, up to polylogarithmic factors, we have

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_2^2\big] \lesssim \frac{\bar{\sigma}^2 t_{\mathrm{mix}} d}{(1-\kappa)^2 n}. \tag{3.21}$$

Thus, at least in a worst-case sense, the second term is always dominated by the first term.

We note that Theorem 3.1 makes two types of tail assumptions on the random observations: Assumption 3.2 with $\bar{p} = 2$ requires dimension-free second moment bounds

[3] This can be easily seen from exponential decay of the correlation; in particular, see equation (3.75) in the proof of the theorem.

in any coordinate direction, whereas the Lipschitz condition (Assumption 3.4) together with Assumption 3.3 (boundedness of the domain) imply a (dimension-dependent) uniform upper bound on the noise. The two assumptions play very different roles in the analysis of high-dimensional problems. As we will see in Proposition 3.3, such assumptions are naturally satisfied in the context of sieve estimators, for which dimension d of the problem is selected adaptively based on sample size n.

Finally, we also note that the requirement on the sample size n is nearly optimal, since we require $n = \widetilde{\Omega}\big(\frac{t_{\mathrm{mix}} d}{(1-\kappa)^2}\big)$ to make the estimation error (3.21) less than a constant (by seeing σ_L and $\gamma_{\max}$ as constants). Up to an additional $\mathcal{O}\left(t_{\mathrm{mix}}\right)$ factor, the sample size requirement in Theorem 3.1 also matches that of linear stochastic approximation in the i.i.d. setting established in existing literature [115, 150] and 2. This additional $\mathcal{O}\left(t_{\mathrm{mix}}\right)$ factor is unavoidable, which can be seen from the following reduction from the Markov to the i.i.d. setting. Consider a problem instance in the i.i.d. setup, given by a probability distribution $\mathbb{P}$ over $\mathbb{R}^{d \times d} \times \mathbb{R}^d$. Defining the state (L_t, b_t), consider a lazy Markov chain that remains at the same state with probability $1 - \frac{1}{t_{\mathrm{mix}}}$, and jumps to an independent state drawn from $\mathbb{P}$ with probability $\frac{1}{t_{\mathrm{mix}}}$. A Markov trajectory of size n in this lazy Markov chain is approximately equivalent to $\mathcal{O}\left(n/t_{\mathrm{mix}}\right)$ samples in the i.i.d. model, and results in a multiplicative blow-up of $\mathcal{O}\left(t_{\mathrm{mix}}\right)$ in the sample complexity requirement for the Markov case.

3.3.1.2 Bounds on the non-averaged iterates

The proof of Theorem 3.1 involves first analyzing the non-averaged iterates. Since the upper bound established in this step is of independent interest, we state and discuss it here:

Proposition 3.1. *Under Assumptions 3.1—3.3, there are universal positive constants* (c_0, c_1) *such that for any integer* $p \in \{1\} \cup [\log n, \bar{p}/2]$, *scalar* $\tau \geq 2p t_{\mathrm{mix}} \log(c_0 d)$, *and positive stepsize* $\eta \in \big(0, \frac{1-\kappa}{2cp^3(\sigma_L^2 d + \gamma_{\max}^2)\tau}\big]$, *we have*

$$\left(\mathbb{E}\|\theta_t - \bar{\theta}\|_2^{2p}\right)^{1/p} \leq e^{-\frac{1}{2}\eta(1-\kappa)t}\left(\mathbb{E}\|\theta_0 - \bar{\theta}\|_2^{2p}\right)^{1/p} + \frac{cp^3\eta}{1-\kappa}\bar{\sigma}^2\tau d \tag{3.22}$$

for all $t = 1, \ldots, n$.

See Section 3.5 for the proof of this proposition.

Note that the guarantees on the unaveraged iterates in Proposition 3.1—unlike those of Theorem 3.1 for the averaged iterates—do not match the optimal instance-dependent behavior. This is to be expected, since at least asymptotically, the unaveraged sequence converges to a Gaussian random vector with covariance specified by the solution of a Riccati equation. (For details, see Section 4.5.3 of [11]). This covariance term need not match the optimal statistical error.

On the other hand, by choosing $\eta \asymp \frac{\log n}{(1-\kappa)n}$, the bound in Proposition 3.1 matches the worst-case bound in equation (3.21), up to log factors. We also note that in

Proposition 3.1, the exponent p can take values in two ranges: regardless of the value of $\bar{p} \in [2, \infty]$, one can always take $p = 1$ and obtain an upper bound on the mean-squared error $\mathbb{E}\big[\|\theta_t - \bar{\theta}\|_2^2\big]$. This bound only requires Assumption 3.2 to hold true with $\bar{p} \geq 2$, which covers many important examples (see Section 3.4). On the other hand, when Assumption 3.2 is satisfied with $\bar{p} \geq 2 \log n$ and a stronger moment assumption is imposed, one can obtain a p-th moment bound for any $p \geq [2 \log n, \bar{p}]$. This bound can be readily converted into a high-probability bound for the last iterate of stochastic approximation. It is worth noting that we study these two cases separately, using slightly different proof techniques.

It is worthwhile making some comparisons between Proposition 3.1 and existing results on the unaveraged forms of Markovian stochastic approximation. As we have noted in our examples, in many cases, the quantities $(\sigma_L, \sigma_b, \bar{\sigma})$ do not depend on the dimension, in which case the error bound in Proposition 3.1 grows linearly with dimension d. In comparison, in terms of our notation, the error bounds in the papers [16, 194] both exhibit quadratic dependency on the quantity $\frac{\max_{s \in \mathcal{S}} \|\boldsymbol{L}_t(s)\|_{\mathrm{op}}}{1-\kappa}$. As we noted previously in equation (3.8), this quantity scales linearly in dimension when the observations have a constant rank (independent of dimension), so that (even after optimal parameter tuning), the bounds from these parameters scale at least proportionally to $\frac{d^2}{n}$. This scaling should be contrasted with the $\mathcal{O}(d/n)$ guarantees from our bounds. On the other hand, the analysis in [56] involves a different mixing assumption, and so is not directly comparable to our results. However, it is worth noting that their bound $\|\theta_t - \bar{\theta}\|_2$ also has an explicit $\mathcal{O}\left(d/\sqrt{n}\right)$ term (cf. equation (32) in their paper), showing that the MSE bound grows quadratically with dimension.

3.3.2 Local minimax lower bounds

Thus far, we established instance-dependent upper bounds for the averaged SA scheme with Markov noise. It is natural to wonder whether these bounds can be improved. Answering this question requires the development of local minimax lower bounds, which we describe in this section.

3.3.2.1 Set-up and local neighborhoods

We begin with the set-up and the definition of local neighborhoods for our lower bounds. Let P be an irreducible Markov transition kernel on a finite state space $\mathcal{S}$ with associated stationary measure ξ_P. Consider the solution $\bar{\theta}(P)$ to the following fixed-point equation

$$\bar{\theta}(P) = \mathbb{E}_{\xi_P}\big[\boldsymbol{L}(s)\big] \cdot \bar{\theta}(P) + \mathbb{E}_{\xi_P}\big[\boldsymbol{b}(s)\big]. \tag{3.23}$$

where the maps $\boldsymbol{b}$ and $\boldsymbol{L}$ are known to the estimator, whereas the Markov transition kernel is unknown. For some fixed P_0 with stationary measure ξ_0, we would like to lower bound the number of observations required to estimate $\bar{\theta}(P_0)$ to a given accuracy. In order to obtain such a lower bound, we consider the fixed point problem (3.23) over a local neighborhood[4] of the pair (P_0, ξ_0). We assume that the estimator is based on a Markov trajectory $\{s_t\}_{t=0}^n$, with initial state s_0 drawn according to the original[5] stationary distribution ξ_0, and successive states evolving according to the transition kernel P.

In order to quantify the complexity of estimation localized around the Markov transition kernel P_0, we define the following two notions of local neighborhood:

$$\mathfrak{N}_{\mathrm{Prob}}\big(P_0, \varepsilon\big) := \big\{P : \sum_{x\in\mathcal{S}} \xi_0(x)\cdot\chi^2\left(P(x,\cdot)\ ||\ P_0(x,\cdot)\right) \le \varepsilon^2\big\}, \tag{3.24a}$$

$$\mathfrak{N}_{\mathrm{Est}}\big(P_0, \varepsilon\big) := \big\{P : \|\bar{\theta}(P) - \bar{\theta}(P_0)\|_2 \le \varepsilon\big\}. \tag{3.24b}$$

The two notions of neighborhood focus on different types of locality restrictions on the model class: the local problem class $\mathfrak{N}_{\mathrm{Prob}}$ contains all the Markov transition kernels that are "globally close" to a given kernel P_0, measured by a weighted χ^2 divergence. It is worth noting that this weighted χ^2 divergence has an operational interpretation. Suppose we draw $x \sim \xi_0$, and then draw the next state $y \sim P_0(x,\cdot)$ accordingly the original Markov kernel P_0, as well as $y' \sim P(x,\cdot)$ under the kernel P. Then the weighted χ^2 divergence is the χ^2 divergence between the joint laws of (x, y) and (x, y').

On the other hand, the local class $\mathfrak{N}_{\mathrm{Est}}$ contains Markov transition kernels P such that the solution $\bar{\theta}(P)$ to the fixed-point equation (3.23) lies in a local neighborhood of the given solution $\bar{\theta}(P_0)$, measured by the Euclidean distance. This problem class captures the complexity specifically for solving the fixed-point equation, without the need to estimate the entire transition kernel. In particular, it is easy to construct a Markov kernel P such that the solution $\bar{\theta}(P)$ is very close to $\bar{\theta}(P_0)$, but the distance between the transition kernels P and P_0 (e.g. measured in weighted χ^2 divergence) is arbitrarily large.

3.3.2.2 Instance-dependent lower bound

Our lower bound is proved on the smallest worst-case risk attainable over the intersection of $\mathfrak{N}_{\mathrm{Prob}}$ and $\mathfrak{N}_{\mathrm{Est}}$. We use the shorthand notation $\bar{L}^{(0)} := \mathbb{E}_{\xi_0}\big[\boldsymbol{L}(s)\big]$. Also recall the covariance matrix $\Sigma^*{}_{\mathrm{Mkv}} = \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(\widetilde{s}_t)\varepsilon_{\mathrm{Mkv}}(\widetilde{s}_0)^\top\big]$, as previously defined in equation (3.19), for a stationary trajectory $(\widetilde{s}_t)_{t\in\mathbb{Z}}$ under the transition kernel P_0. Our bound depends on the *local radius*

$$\varepsilon_n = n^{-1/2}\sqrt{\mathrm{trace}\left((I - \bar{L}^{(0)})^{-1}\Sigma^*{}_{\mathrm{Mkv}}(I - \bar{L}^{(0)})^{-\top}\right)}, \tag{3.25}$$

which is the contribution of Markovian noise to the upper bound stated in Theorem 3.1.

We are now ready to state our lower bound. Recall that we have assumed that the kernel P_0 is irreducible and aperiodic. We also assume the mixing condition

[4] Doing so is necessary to rule out trivial estimators, and the possibility of super-efficiency.

[5] In our construction, both kernels P_0 and P are rapidly mixing and their stationary measure are sufficiently close in TV distance that the choice of initial distribution does not affect the result. Drawing $s_0 \sim \xi_0$ is made for theoretical convenience.

(Assumption 3.1) holds with the discrete metric $\rho(x,y) = \mathbf{1}_{\{x \neq y\}}$ and mixing time t_{mix}, and that $\text{supp}\big(P_0(s,\cdot)\big) \geq 2$ for all $s \in \mathcal{S}$.

Theorem 3.2. *Under the assumptions stated above, there exist universal positive constants (c, c_1, c_2) such that for any sample size n lower bounded as*

$$n \geq \tfrac{c t_{\text{mix}}^2 \sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2}, \quad \textit{and} \quad n^2 \varepsilon_n^2 \geq \tfrac{2c(1+\sigma_L^2)\bar{\sigma}^2 t_{\text{mix}}^4 d^2}{(1-\kappa)^4} \log^6 \big(\tfrac{d}{\min_s \xi_0(s)}\big), \tag{3.26a}$$

we have the minimax lower bound

$$\inf_{\widehat{\theta}_n} \sup_{P \in \mathfrak{N}'} \mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}(P)\|_2^2\big] \geq c_2 \varepsilon_n^2, \tag{3.26b}$$

where $\mathfrak{N}' := \mathfrak{N}_{\text{Prob}}\big(P_0, c_1\sqrt{\tfrac{d}{n}}\big) \cap \mathfrak{N}_{\text{Est}}(P_0, c_1\varepsilon_n)$.

See Appendix B.5 for the proof of this theorem.

A few remarks are in order. First, note that the minimax lower bound is with respect to the problem class $\mathfrak{N}_{\text{Prob}}\big(P_0, c_1\sqrt{\tfrac{d}{n}}\big) \cap \mathfrak{N}_{\text{Est}}(P_0, c_1\varepsilon_n)$, which requires both the transition kernel P and the solution $\bar{\theta}(P)$ to be close to the given problem instance $(P_0, \bar{\theta}(P_0))$. The size of the weighted χ^2 neighborhood scales with the standard parametric rate $\sqrt{d/n}$, as desired in such problems. On the other hand, the size of the neighborhood around $\bar{\theta}(P_0)$ is proportional to the local radius ε_n that appears in the lower bound. Operationally, this result indicates that even if the estimator knows in advance that $\bar{\theta}(P)$ lies in the ball $\mathbb{B}(\bar{\theta}(P_0), c_1\varepsilon_n)$, one cannot do much better than simply outputting an arbitrary point in this ball without looking at the data.

Second, it should be noted that quantity ε_n^2 matches (up to a constant factor) the optimal mean-squared error given by the local asymptotic minimax theorem [209, 70]. In contrast to such asymptotic theory, however, Theorem 3.2 applies when n is finite, and does not impose any regularity assumptions on the estimator. Furthermore, the radius ε_n that is used to define the local neighborhood $\mathfrak{N}_{\text{Est}}(P_0, \varepsilon_n)$ is optimal in the following sense. On the one hand, since the plug-in estimator is asymptotically normal [70], for any decreasing sequence ε_n' such that $\varepsilon_n' > \varepsilon_n$ and $\varepsilon_n' \to 0^+$, the minimax risk within the neighborhood $\mathfrak{N}_{\text{Est}}(P_0, \varepsilon_n')$ behaves asymptotically as ε_n^2 up to constant factors. On the other hand, for any decreasing sequence ε_n' such that $\varepsilon_n' < \varepsilon_n$, the minimax risk in the neighborhood $\mathfrak{N}_{\text{Est}}(P_0, \varepsilon_n')$ is at most ε_n'. In the latter case, the neighborhood is so small that it provides more information than the data provides.

Theorem 3.2 matches the Markov noise term in Theorem 3.1, establishing its optimality when the martingale part of the noise vanishes, i.e., $\boldsymbol{L}_t(s) = \boldsymbol{L}(s)$ and $\boldsymbol{b}_t(s) = \boldsymbol{b}(s)$. The lower bound does not capture the martingale part of the noise because we assume that the functions $L : \mathcal{S} \to \mathbb{R}^{d\times d}$ and $b : \mathcal{S} \to \mathbb{R}^d$ are known to the estimator. In the setting where these functions are also observed only through noisy i.i.d. data (L_t, b_t), Theorem 2.3 in Chapter 2 implies a lower bound of the form $c_2 n^{-1}\,\text{trace}\,\big((I - \bar{L}^{(0)})^{-1}{\Sigma^*}_{\text{MG}}(I - \bar{L}^{(0)})^{-\top}\big)$.

Combining it with Theorem 3.2 implies a minimax lower bound involving the term $c_2' n^{-1}$ trace $\big((I-\bar{L}^{(0)})^{-1}({\Sigma^*}_{\mathrm{Mkv}}+{\Sigma^*}_{\mathrm{MG}})(I-\bar{L}^{(0)})^{-\top}\big)$ in a properly defined local neighborhood, thus establishing the optimality of Theorem 3.1. Finally, we note that Theorem 3.2 requires the sample size to be at least $t_{\mathrm{mix}}^2 d^2$, which is more stringent than the $\mathcal{O}\left(t_{\mathrm{mix}} d\right)$ requirement in the upper bound. While Theorem 3.1 holds true with a linear sample-size $n = \mathcal{O}\left(d\right)$, it is only shown to be instance-optimal for larger $n = \Omega(d^2)$. This mismatch is due to the fact that small perturbations of the Markov transition kernel in certain directions can destroy its fast mixing property. That being said, Theorem 3.2 is still a finite-sample result, with polynomial dependency on the quantities $\big(t_{\mathrm{mix}}, d, \frac{1}{1-\kappa}\big)$, and poly-logarithmic dependency on the smallest stationary probability.

3.4 Some consequences for specific problems

In this section, we specialize our analysis to the examples described in Section 3.2.2, namely approximate policy evaluation using TD algorithms, and estimation in autoregressive time series models. By verifying the conditions needed to apply Theorem 3.1 and Proposition 3.1, we obtain some more concrete corollaries of our general theory.

3.4.1 TD(0) method

Recall the TD(0) algorithm for policy evaluation, as previously described in Example 3.1. We are interested in estimating the solution v^* of the Bellman equation (3.10), and an approximation scheme is employed using the basis functions $(\phi_j)_{j=1}^d$. Using the shorthand $\langle \theta,\, \phi(s)\rangle = \sum_{j=1}^d \theta_j \phi_j(s)$ for the Euclidean inner product in $\mathbb{R}^d$, with observation model $(\boldsymbol{L}_{t+1}(\Omega_t), \boldsymbol{b}_{t+1}(\Omega_t))$ defined in Eq (3.13), the averaged SA procedure (3.3) is given by:

$$\theta_{t+1} \overset{(a)}{=} \theta_t - \eta\big\{\langle \phi(s_t) - \gamma\phi(s_{t+1}),\, \theta_t\rangle - R_{t+1}(s_t)\big\}\phi(s_t), \quad \text{and} \quad \widehat{\theta}_n \overset{(b)}{=} \tfrac{1}{n-n_0}\sum_{t=n_0}^{n-1}\theta_t. \tag{3.27}$$

To be clear, the update (3.27)(a) is the standard TD(0) algorithm with stepsize η, whereas the addition of the averaging step (3.27)(b) yields the Polyak–Ruppert averaged version of the scheme. Note that we re-scale the stepsize η by a factor of β for notational convenience. In the following subsections, we derive corollaries of our general theory for the averaged scheme under different mixing conditions on the underlying Markov chain.

3.4.1.1 Markov chains with mixing in total variation distance

We first assume that the Markov chain satisfies a mixing condition (cf. Assumption 3.1) in the discrete metric: i.e., after t_{mix} steps, we have $d_{\mathrm{TV}}(\delta_s P^{t_{\mathrm{mix}}}, \delta_{s'} P^{t_{\mathrm{mix}}}) \leq \frac{1}{2}$ for any pair $s, s' \in \mathcal{S}$. Let ξ denote the stationary distribution of the Markov chain that generates the trajectory $\{s_t\}_{t\geq 0}$, and let P denote its transition kernel. Note that the augmented state vector $\Omega_t = (s_t, s_{t+1})$ evolves according to a Markov process with mixing time $t_{\text{mix}} + 1$. Moreover, the stationary distribution of the pair $\Omega = (s, s^+)$ has the form $s \sim \xi,\; s^+ \sim P(\cdot \mid s)$. We denote the stationary covariance of the feature vectors as $B := \mathbb{E}_{s\sim\xi}\big[\phi(s)\phi(s)^\top\big]$, and also define the minimum and maximum eigenvalues $\mu := \lambda_{\min}(B)$ and $\beta := \lambda_{\max}(B)$. We assume that

$$\|B^{-1/2}\phi(s)\|_2 \overset{(a)}{\leq} \varsigma\sqrt{d} \quad \text{and} \quad |R_t(s)| \overset{(b)}{\leq} \varsigma \quad \text{for all } s \in \mathcal{S}, \text{ and} \tag{3.28a}$$

$$\mathbb{E}_\xi\big[\langle B^{-1/2}\phi(s),\, u\rangle^4\big] \leq \varsigma^4 \quad \text{for all } u \in \mathbb{S}^{d-1}. \tag{3.28b}$$

In order to state our result, we define the following quantities:

$$\begin{aligned} M &:= \gamma B^{-1/2} \cdot \mathbb{E}_{s\sim\xi, s^+\sim P(s,\cdot)}\big[\phi(s)\phi(s^+)^\top\big] \cdot B^{-1/2}, \\ \varepsilon_{\text{Mkv}}(s, s^+) &:= B^{-1/2}\big(\phi(s)^\top\bar{\theta} - \gamma\phi(s^+)^\top\bar{\theta} - r(s)\big)\phi(s), \\ \varepsilon_{\text{MG}}(s) &:= B^{-1/2}(R(s) - r(s))\phi(s) \end{aligned}$$

We also define the following covariance matrices according to Eq (3.19):

$$\begin{aligned} {\Sigma^*}_{\text{Mkv}} &:= \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\text{Mkv}}(s_t, s_{t+1})\varepsilon_{\text{Mkv}}(s_0, s_1)^\top\big], \\ {\Sigma^*}_{\text{MG}} &:= \mathbb{E}_{s\sim\xi}\big[\mathbb{E}\big[\varepsilon_{\text{MG}}(s)\varepsilon_{\text{MG}}(s)^\top \mid s\big]\big]. \end{aligned}$$

Finally, we define the quantity

$$\bar{\sigma}^2 := \varsigma^2 \cdot \sqrt{\mathbb{E}\Big[\big(\phi(s_t)^\top\bar{\theta} - \gamma\phi(s_{t+1})\bar{\theta} - R_t(s_t)\big)^4\Big]}, \tag{3.29}$$

and let $\kappa := \frac{1}{2}\lambda_{\max}(M + M^\top)$. It is easy to see that $\kappa \leq \gamma < 1$. Assuming that $\mu > 0$, we are then ready to state our main result for the TD(0) method.

Corollary 3.1. *Under the setup above, take the stepsize η and burn-in period n_0 as*

$$\eta = \tfrac{1}{c\beta((\varsigma^4+1)d(1-\kappa)n^2 t_{\text{mix}})^{1/3}}, \quad \textit{and} \quad n_0 = \tfrac{1}{2}n, \tag{3.30}$$

and suppose that $\frac{n}{\log^3 n} \geq \frac{2t_{\text{mix}}(\varsigma^4+1)d\beta^2}{(1-\kappa)^2\mu^2}$. The estimator $\widehat{v}_n := \widehat{\theta}_n\phi$ obtained from the Polyak–Ruppert procedure (3.27) satisfies the bound

$$\begin{aligned} \mathbb{E}\big[\|\widehat{v}_n - \bar{v}\|^2_{\mathbb{L}^2(\mathcal{S},\xi)}\big] \leq \tfrac{c}{n}\mathrm{Tr}\big\{(I_d - M)^{-1}({\Sigma^*}_{\text{Mkv}} + {\Sigma^*}_{\text{MG}})(I_d - M)^{-\top}\big\} \\ + c\Big(\tfrac{\beta^2\bar{\sigma}^2 d t_{\text{mix}}}{\mu^2(1-\kappa)^2 n}\Big)^{4/3}\log^2 n, \end{aligned} \tag{3.31}$$

where $\bar{v}$ is the solution to the projected fixed-point equation (3.11) and $c > 0$ is a universal constant.

See Appendix B.6.1.1 for the proof of this corollary.

A few remarks are in order. First, we measure the estimation error in the canonical $\| \cdot \|_{\mathbb{L}^2(\mathcal{S},\xi)}$ norm, instead of the Euclidean distance in $\mathbb{R}^d$. Consequently, the proof of this corollary actually uses a generalized version of Theorem 3.1 proved for weighted ℓ^2 norms. On the other hand, we note that the error bound (3.31) is with respect to the solution $\bar{v}$ to the projected fixed-point equation. In the well-specified case where $v^* \in \mathbb{S}$, this solution coincides with the value function v^*. In general, the approximation error needs to be taken into account, and was the focus of Chapter 2. In conjunction with this result, Corollary 3.1 implies the error bound

$$\begin{aligned}
&\mathbb{E}\big[\|\widehat{v}_n - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2\big] \\
&\quad \leq c\big[1 + \lambda_{\max}\big((I_d - M)^{-1}(\gamma^2 I_d - MM^\top)(I_d - M)^{-\top}\big)\big] \inf_{v \in \mathbb{S}} \|v - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2 \\
&\qquad + \tfrac{c}{n}\mathrm{Tr}\big\{(I_d - M)^{-1}({\Sigma^*}_{\mathrm{Mkv}} + {\Sigma^*}_{\mathrm{MG}}))(I_d - M)^{-\top}\big\} + c\big(\tfrac{\beta^2 \bar{\sigma}^2 d t_{\mathrm{mix}}}{\mu^2(1-\kappa)^2 n}\big)^{4/3} \log^2 n. \qquad (3.32)
\end{aligned}$$

In Section 3.4.2 to follow, we provide a general recipe to trade off approximation and estimation errors to choose the value of λ in the class of TD(λ) algorithms. Before that, we discuss two extensions of Corollary 3.1.

3.4.1.2 Markov chains with mixing in Wasserstein metric

Note that for Corollary 3.1, the mixing time condition is imposed with total variation distance. When the state space $\mathcal{S}$ is continuous, e.g., the set $\mathcal{S}$ is a subset of $\mathbb{R}^m$, mixing in Wasserstein distance could capture the geometry of the underlying metric better. In this section, we extend our analysis to such settings, highlighting the dimension dependency in the sample complexity.

Concretely, we consider a Markov chain $(s_t)_{t \geq 0}$ on a compact domain $\mathcal{S} \subseteq \mathbb{R}^m$, and a feature mapping $\phi : \mathcal{S} \to \mathbb{R}^d$. We assume that the Markov chain admits a unique stationary measure ξ, and the mixing time assumption holds in Wasserstein-1 distance, so that $\mathcal{W}_1\big(\delta_x P^{t_{\mathrm{mix}}}, \delta_y P^{t_{\mathrm{mix}}}\big) \leq \frac{1}{2}\|x - y\|_2$ for all $x, y \in \mathcal{S}$. For the sake of normalization, we assume that $\mathcal{S} \subseteq \mathbb{B}(0,1)$ and $\phi(0) = 0$. On the feature mapping ϕ, we assume the following:

$$\exists \mu, \beta > 0, \quad \mu I_d \preceq B := \mathbb{E}_{s \sim \xi}\big[\phi(s)\phi(s)^\top\big] \preceq \beta I_d, \tag{3.33a}$$

$$\forall x, y \in \mathcal{S}, \quad \|B^{-1/2}\big(\phi(x) - \phi(y)\big)\|_2 \leq \varsigma\sqrt{d}\|x - y\|_2, \tag{3.33b}$$

$$\forall u \in \mathbb{S}^{d-1}, \quad \mathbb{E}_{s \sim \xi}\big[\langle u,\, B^{-1/2}\phi(s)\rangle^4\big] \leq \varsigma^4, \tag{3.33c}$$

$$\forall s, s' \in \mathcal{S},\ t \geq 1, \quad |R_t(s) - R_t(s')| \leq \varsigma\|s - s'\|_2,\ |R_t(s)| \leq \varsigma \quad \text{a.s.} \tag{3.33d}$$

Here, we regard the parameters (ς, μ, β) as dimension-independent positive constants. Since the state space $\mathcal{S}$ has diameter bounded by 2, the feature mapping ϕ satisfying equation (3.33a) necessarily has Lipschitz constant of order $\mathcal{O}\left(\sqrt{d}\right)$. For a simple example, take the state x itself as the feature vector (after appropriate re-scaling), which corresponds to the case of $m = d$ and $\phi(x) = \sqrt{d} \cdot x$.

With this set-up, we have the following guarantee:

Corollary 3.2. *Assuming the conditions in equation (3.33), taking stepsize and burn-in period as equation (3.30), for the Polyak–Ruppert averaged stochastic approximation procedure (3.27), the bound (3.31) holds.*

See Appendix B.6.1.2 for the proof.

Corollary 3.2 shows that the same instance-dependent bound holds true for a continuous state space setting. Such a bound is useful for many applications, one of which is the case of quadratic value functions, where the dimension satisfies the relation $d = m^2$ the mapping ϕ takes the form $\phi : x \mapsto m \cdot xx^\top$. Assuming that the process $(s_t)_{t \geq 0}$ is supported in a unit ball $\mathbb{B}(0, 1)$ and has well-conditioned stationary covariance, it is easy to verify that Assumptions (3.33) are satisfied with dimension-free constants (ς, μ, β). This example is particularly useful for policy evaluation in Linear Quadratic Regulators (LQR). Nevertheless, our results hold more generally for any random dynamical system that is rapidly mixing in the $\mathcal{W}_1$ distance.

3.4.1.3 Analysis of a sieve estimator

The optimal dimension dependency in Theorem 3.1 allows us to obtain optimal estimators for various classes of non-parametric problems, in which the dimension is a parameter to be chosen. In particular, sieve methods are a class of non-parametric estimators based on nested sequences of finite-dimensional approximations. In this section, we analyze the behavior of a stochastic approximation sieve estimator in the Markovian setting. The optimal dimension dependence in our theorem recovers the minimax optimal rates for estimation, while our instance-dependent bounds help in capturing more refined structure in the problem instance.

Concretely, assuming that the Hilbert space $\mathbb{L}^2(\mathcal{S}, \xi)$ is separable, let $(\phi_j)_{j=1}^\infty$ be a set of (not necessarily orthogonal) basis functions. We consider the case where the mixing condition holds true with total variation distance[6]. The following assumptions are imposed on the basis functions:

$$\forall j \in \mathbb{N}^+, \quad \sup_{x \in \mathcal{S}} |\phi_j(x)| \leq \varsigma, \tag{3.34a}$$

$$\forall d \in \mathbb{N}^+, \quad \mu I_d \leq \Big[\mathbb{E}_{s \sim \xi}\big(\phi_j(s)\phi_\ell(s)\big)\Big]_{j,\ell \in [d]} \leq \beta I_d, \tag{3.34b}$$

$$\forall t \geq 1, \quad \sup_{x \in \mathcal{S}} |R_t(x)| \leq \varsigma. \tag{3.34c}$$

The first assumption is standard in nonparametric regression, and satisfied by many useful basis functions such as the Fourier basis and Walsh-Hadamard basis. The second

[6] By following the approach in the previous subsection, the analysis can also be extended to the case of mixing in Wasserstein distance.

assumption relaxes the orthogonality requirement on the bases, by only requiring the Gram matrix to be well-conditioned.

We define the noise level $\bar{\sigma}$ using the second moment:

$$\bar{\sigma}^2 := \varsigma^2 \cdot \sqrt{\mathbb{E}\big[\big(\bar{v}(s_t) - \gamma \bar{v}(s_{t+1}) - R_t(s_t)\big)^2\big]}. \tag{3.35}$$

Once again, we run the averaged stochastic approximation procedure (3.27) on this problem. A crucial point of departure from the parametric models discussed above is that the number of basis functions d_n in sieve estimators is chosen based on the problem structure and sample size. Let $\mathbb{S}(d_n) := \mathrm{span}(\phi_1, \phi_2, \cdots, \phi_{d_n})$ denote the subspace spanned by the first d_n basis functions. The following result is a direct corollary of our theorem, and covers the case of fixed d_n; we discuss the trade-off between approximation and estimation error in the choice of d_n presently.

Corollary 3.3. *Assuming the conditions in equation (3.34), take the stepsize and burn-in period as in equation (3.30). Assuming that $\mu, \beta, \varsigma \asymp 1$, the Polyak–Ruppert averaged stochastic approximation procedure (3.27) satisfies the bound (3.31) with $d = d_n$.*

See Appendix B.6.1.3 for the proof.

Recall that by taking into account the approximation error, the error for estimating the true value function v^* takes the following form:

$$\begin{aligned}
&\mathbb{E}\big[\|\widehat{v}_n - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2\big] \\
&\leq c\big[1 + \lambda_{\max}\big((I-M)^{-1}(\gamma^2 I_d - MM^\top)(I-M)^{-\top}\big)\big] \inf_{v \in \mathbb{S}} \|v - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2 \\
&\quad + \tfrac{c}{n}\mathrm{Tr}\big((I-M)^{-1}({\Sigma^*}_{\mathrm{Mkv}} + {\Sigma^*}_{\mathrm{MG}})(I-M)^{-\top}\big) + c\big(\tfrac{\bar{\sigma}^2 t_{\mathrm{mix}} d_n}{(1-\kappa)^2 n}\big)^{4/3} \log^2 n.
\end{aligned}$$

Let $\{\psi_j\}_{j=1}^{+\infty}$ be an orthonormal basis of $\mathbb{L}^2(\mathcal{S}, \xi)$ such that $\mathrm{span}(\psi_1, \cdots, \psi_d) = \mathrm{span}(\phi_1, \cdots, \phi_d)$ for any $d \geq 1$. (For instance, one can let $\{\psi_j\}_{j=1}^{+\infty}$ be the Gram-Schmidt orthonormalization of the original basis functions). Given a non-increasing sequence $\{\alpha_j\}_{j=1}^{\infty}$ of positive reals such that $\lim_{j \to +\infty} \alpha_j = 0$, we first let $\mathcal{H}_0$ be a linear subspace of $\mathbb{L}^2(\mathcal{S}, \xi)$, consisting of all the finite linear combination of basis vectors $\{\psi_j\}_{j=1}^{+\infty}$, equipped with the following inner product:

$$\forall u, v \in \mathcal{H}_0, \quad \langle u,\, v \rangle_{\mathcal{H}_0} := \sum_{j=1}^{\infty} \alpha_j^{-1} \cdot \langle u,\, \psi_j \rangle \cdot \langle v,\, \psi_j \rangle.$$

Note that the summation shown above is actually finite, since both both sequences $(\langle u,\, \psi_j \rangle)_{j=1}^{+\infty}$, $(\langle v,\, \psi_j \rangle)_{j=1}^{+\infty}$ only have finite non-zero entries. We then define the inner product space $(\mathcal{H}, \langle \cdot,\, \cdot \rangle_{\mathcal{H}})$ as the completion of $(\mathcal{H}_0, \langle \cdot,\, \cdot \rangle_{\mathcal{H}_0})$. It is easy to see that $\mathcal{H}$ is a Hilbert space, and a linear subspace of $\mathbb{L}^2(\mathcal{S}, \xi)$.

For any $v^* \in \mathcal{H}$, the estimation error is at most (in the worst-case)

$$\mathbb{E}\big[\|\widehat{v}_n - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2\big] \leq \tfrac{c}{1-\gamma} \cdot \alpha_{d_n} \|v^*\|_{\mathcal{H}}^2 + \tfrac{c\bar{\sigma}^2 d_n t_{\mathrm{mix}}}{(1-\gamma)^2 n}. \tag{3.36}$$

For example, when the eigenvalues of Hilbert space $\mathcal{H}$ decay as $\alpha_j \asymp j^{-2s}$ for some $s > 0$, the estimator achieves a rate of $\mathcal{O}\big((t_{\mathrm{mix}}/n)^{\frac{2s}{2s+1}}\big)$, which matches the minimax optimal rate proved by [54] in the i.i.d. setting, but with a multiplicative correction to the effective sample size by a factor t_{mix} to accommodate Markovian observations. Furthermore, since one can estimate the quantities $(M, {\Sigma^*}_{\mathrm{Mkv}}, {\Sigma^*}_{\mathrm{MG}})$ in the bound (3.31) using $\mathcal{O}\left(d\right)$ samples, instance-dependent model selection can in principle be conducted. Bounds of the form (3.36) thus open the door to asking important questions of this type.

3.4.2 TD(λ) methods

Now we turn to stochastic approximation methods for the TD(λ) projected fixed-point equation (3.14b), with some given $\lambda \in [0, 1)$. With observation model $(\boldsymbol{L}_{t+1}(\Omega_t), \boldsymbol{b}_{t+1}(\Omega_t))$ given by Eq (3.16c), the averaged SA procedure (3.3) can be written as

$$\theta_{t+1} = \theta_t - \eta \Big\{ \langle \phi(s_t) - \gamma \phi(s_{t+1})^\top, \, \theta_t \rangle - R_t(s_t) \Big\} g_t, \quad \text{where} \tag{3.37a}$$

$$g_t = \gamma \lambda g_{t-1} + \phi(s_t) \quad \text{and,} \tag{3.37b}$$

$$\widehat{\theta}_n = \tfrac{1}{n-n_0} \sum_{t=n_0}^{n-1} \theta_t. \tag{3.37c}$$

The update on g_t is the so-called “eligibility trace” in the TD(λ) algorithm. As before, we assume the two bounds in equation (3.28a), and assume that the mixing time condition 3.1 holds true for the chain $(s_t)_{t \geq 1}$, with discrete metric and mixing time t_{mix}. We consider the augmented Markov chain $\Omega_t := \big(s_t, s_{t+1}, \frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}} g_t\big) \in \mathcal{S}^2 \times \mathbb{B}(0, 1)$ and begin by establishing mixing conditions on this augmented chain.

Proposition 3.2. *Under the setup above, consider the metric*

$$\rho\big((s_1, s_2, h), (s_1', s_2', h')\big) := \tfrac{1}{4}\big(\mathbf{1}_{s_1 \neq s_1'} + \mathbf{1}_{s_2 \neq s_2'} + \|h - h'\|_2\big). \tag{3.38a}$$

Taking $\tau = 4\big(t_{\mathrm{mix}} + \frac{1}{1-\gamma\lambda}\big)$, the augmented chain $\big\{\Omega_t = (s_t, s_{t+1}, \frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}} g_t)\big\}_{t \geq 0}$ satisfies the mixing bound

$$\mathcal{W}_{1,\rho}\big(\mathcal{L}(\Omega_\tau), \mathcal{L}(\Omega_\tau')\big) \leq \frac{1}{2} \rho\big(\Omega_0, \Omega_0'\big) \tag{3.38b}$$

for two chains $(\Omega_t)_{t \geq 0}$ and $(\Omega_t')_{t \geq 0}$ starting from Ω_0 and Ω_0', respectively. In particular, the stationary distribution $\widetilde{\xi}$ of the chain $(\Omega_t)_{t \geq 0}$ exists and is unique.

See Appendix B.6.2.1 for the proof of this proposition.

Taking this proposition as given, we are now ready to present our main corollary for TD(λ) procedures. We consider the following instantiation of quantities in Theorem 3.1:

The projected linear operator $(1-\lambda)\sum_{k=0}^{+\infty}\lambda^k(\gamma\Pi_{\mathbb{S}}P)^{k+1}$ in the equation (3.14b) can be represented in the orthonormal basis of the subspace $\mathbb{S}$ as

$$\begin{aligned} M_\lambda &:= I_d - B^{-1/2}\mathbb{E}_{(s,s^+,\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g)\sim\widetilde{\xi}}\big[g\phi(s)^\top - \gamma g\phi(s^+)^\top\big]B^{-1/2} \\ &= (1-\lambda)B^{-1/2}\sum_{t=0}^{\infty}\lambda^t\gamma^{t+1}\mathbb{E}\big[\phi(s_0)\phi(s_{t+1})\big]B^{-1/2}. \end{aligned}$$

The Markovian and martingale part of the noise (in the low-dimensional subspace $\mathbb{S}$) takes the following form:

$$\begin{aligned} \varepsilon_{\mathrm{Mkv},\lambda}\big(s,s^+,\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g\big) &= B^{-1/2}\big(\phi(s)^\top\bar{\theta} - \gamma\phi(s^+)\bar{\theta} - r(s)\big)g, \\ \varepsilon_{\mathrm{MG},\lambda}\big(s,s^+,\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g\big) &= B^{-1/2}(R_0(s) - r(s))g \end{aligned}$$

Finally, we define the covariance matrices ${\Sigma^*}_{\mathrm{Mkv},\lambda}$ and ${\Sigma^*}_{\mathrm{MG},\lambda}$ according to Eq (3.19):

$$\begin{aligned} {\Sigma^*}_{\mathrm{Mkv},\lambda} &:= \sum_{t=-\infty}^{\infty}\mathbb{E}\Big[\varepsilon_{\mathrm{Mkv},\lambda}\big(s_t,s_{t+1},\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g_t\big)\varepsilon_{\mathrm{Mkv},\lambda}\big(s_0,s_1,\frac{1-\gamma\lambda}{\varsigma\sqrt{\beta d}}g_0\big)^\top\Big], \\ {\Sigma^*}_{\mathrm{MG},\lambda} &:= \mathbb{E}_{s\sim\xi}\big[\mathbb{E}\big[\varepsilon_{\mathrm{MG},\lambda}(s)\varepsilon_{\mathrm{MG},\lambda}(s)^\top \mid s\big]\big]. \end{aligned}$$

As before, we let $\beta := \lambda_{\max}(B)$, $\mu := \lambda_{\min}(B)$ and $\kappa_\lambda := \frac{1}{2}\lambda_{\max}(M_\lambda + M_\lambda^\top)$, and define the quantity $\bar{\sigma}$ according to equation (3.29). Note that a straightforward calculation reveals that $\kappa_\lambda \leq \frac{(1-\lambda)\gamma}{1-\lambda\gamma} < 1$. Assuming that $\mu > 0$, we are then ready to state our main result for TD(λ) methods.

Corollary 3.4. *Under the setup above, take the stepsize and burn-in period as*

$$\eta = \frac{(1-\gamma\lambda)^{2/3}}{c\beta\Big((\varsigma^4+1)d(1-\kappa_\lambda)n^2\big(t_{\mathrm{mix}}+\frac{1}{1-\gamma\lambda}\big)\Big)^{1/3}}, \quad \textit{and} \quad n_0 = \tfrac{1}{2}n, \tag{3.39a}$$

and suppose that $\frac{n}{\log^3 n} \geq \frac{2(t_{\mathrm{mix}}+\frac{1}{1-\gamma\lambda})\,(\varsigma^4 d+1)\beta^2}{(1-\kappa_\lambda)^2(1-\gamma\lambda)^2\mu^2}$. *Then the value function estimate* $\widehat{v}_n(s) := \langle\widehat{\theta}_n,\, \phi(s)\rangle$ *obtained from the Polyak–Ruppert procedure* (3.37) *has MSE bounded as*

$$\begin{aligned} \mathbb{E}\big[\|\widehat{v}_n - \bar{v}^{(\lambda)}\|^2_{\mathbb{L}^2(\mathcal{S},\xi)}\big] \leq cn^{-1}\mathrm{Tr}\big((I_d - M_\lambda)^{-1}({\Sigma^*}_{\mathrm{Mkv}} + {\Sigma^*}_{\mathrm{MG}})(I_d - M_\lambda)^{-\top}\big) \\ + c\Big(\frac{\beta^2\bar{\sigma}^2 d\big(t_{\mathrm{mix}}+\frac{1}{1-\gamma\lambda}\big)}{\mu^2(1-\kappa_\lambda)^2(1-\gamma\lambda)^2 n}\Big)^{4/3}\log^2 n, \end{aligned} \tag{3.39b}$$

where $\bar{v}^{(\lambda)}$ *is the solution to the projected fixed-point equation* (3.11).

See Appendix B.6.2.2 for the proof of this corollary.

A few remarks are in order. First, using the same argument as in Corollaries 3.2 and 3.3, one can extend the results for TD(λ) to the cases of continuous state spaces with Wasserstein mixing, as well as to nonparametric sieve estimators. As is well-known, different choices of the tuning parameter λ interpolate the "temporal difference" method, in which we aim at solving the Bellman equation, and the "Monte Carlo" method, in which the value function is estimated directly by averaging the rollout of a Markovian trajectory. For example, on the one hand, letting $\lambda = 0$ recovers the instance-dependent upper bound for TD(0) method in Corollary 3.1. On the other hand, by taking $\lambda = \gamma$, we have $\kappa_\lambda \leq \frac{\gamma}{1+\gamma} \leq \frac{1}{2}$, and the dependence on the discount factor γ appears only through the variance of the noise, instead of through the conditioning of the matrix M_λ. In the next section, we sketch a recipe for the instance-dependent selection of λ that also takes the approximation error into account.

3.4.2.1 Using instance-dependent results to select λ

Recall that the TD(λ) algorithm aims at estimating the solution $\bar{v}^{(\lambda)}$ to the projected fixed-point equation (3.14b). The linear operator in the unprojected fixed-point equation (3.14a) satisfies the norm bound

$$|\!|\!|(1-\lambda)\sum_{k=0}^{\infty}\lambda^k\gamma^{k+1}P^{k+1}|\!|\!|_{\mathbb{L}^2(\mathcal{S},\xi)\to\mathbb{L}^2(\mathcal{S},\xi)} \leq (1-\lambda)\sum_{k=0}^{\infty}\lambda^k\gamma^{k+1} = \tfrac{(1-\lambda)\gamma}{1-\lambda\gamma}.$$

Consequently, invoking Theorem 2.1 of Chapter 2, the approximation error satisfies the bound

$$\|\bar{v}^{(\lambda)} - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2 \leq \alpha\big(M_\lambda, \tfrac{(1-\lambda)\gamma}{1-\lambda\gamma}\big)\cdot\inf_{v\in\mathbb{S}}\|v - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2,$$

where $\alpha(M, z) := 1 + \lambda_{\max}\big((I_d - M)^{-1}(z^2 I_d - MM^\top)(I_d - M)^{-\top}\big)$ is the approximation factor. Combining with Corollary 3.4, we obtain the following bound on the distance to the true value function:

$$\begin{aligned}\mathbb{E}\big[\|\widehat{v}_n - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2\big] \leq c\alpha\big(M_\lambda, \tfrac{(1-\lambda)\gamma}{1-\lambda\gamma}\big)\cdot\inf_{v\in\mathbb{S}}\|v - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2 + c\Big(\tfrac{\beta^2\bar{\sigma}^2 d\big(t_{\mathrm{mix}} + \frac{1}{1-\gamma\lambda}\big)}{\mu^2(1-\kappa_\lambda)^2(1-\gamma\lambda)^2 n}\Big)^{4/3}\log^2 n \\ + \tfrac{c}{n}\mathrm{Tr}\big((I_d - M_\lambda)^{-1}({\Sigma^*}_{\mathrm{Mkv}} + {\Sigma^*}_{\mathrm{MG}})(I_d - M_\lambda)^{-\top}\big) \quad (3.40)\end{aligned}$$

for a universal constant $c > 0$.

It can be seen that $\alpha\big(M_\lambda, \frac{(1-\lambda)\gamma}{1-\lambda\gamma}\big) \leq c'\frac{1-\lambda\gamma}{1-\gamma}$ for a universal constant. We also recall that $\kappa_\lambda \leq \frac{(1-\lambda)\gamma}{1-\lambda\gamma}$. If we take the parameters (μ, β, ς) to be of constant order, in the worst case, the upper bound (3.40) takes the simplified form

$$\mathbb{E}\big[\|\widehat{v}_n - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2\big] \leq c\frac{1-\lambda\gamma}{1-\gamma}\inf_{v\in\mathbb{S}}\|v - v^*\|_{\mathbb{L}^2(\mathcal{S},\xi)}^2 + c\frac{\big(t_{\mathrm{mix}} + \frac{1}{1-\gamma\lambda}\big)d}{(1-\gamma)^3 n}.$$

From such an upper bound, it may appear that the optimal choice of λ is always $\lambda = \gamma \wedge (1 - 1/t_{\mathrm{mix}})$, so that the approximation factor is minimized and the variance remains controlled. However, this choice could be overly conservative, since the actual variance with small λ can be significantly smaller, with the feature vectors still having bounded one-step cross-correlation. Choosing the parameter λ close to 1 cannot take advantage of small one-step correlation. On the other hand, a fine-grained bound of the form (3.40) can be used to perform instance-dependent model selection, as follows:

- Construct a uniform finite grid $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_m = \gamma$ for possible values of λ.
- For each $\ell \in [m]$, compute the TD(λ_ℓ) estimator, and construct empirical plug-in estimates $\big(\widehat{M}_{\lambda,n}, \widehat{\Sigma^*}_{\mathrm{Mkv},\lambda,n}, \widehat{\Sigma^*}_{\mathrm{MG},\lambda,n}\big)$ for the matrices $\big(M_\lambda, {\Sigma^*}_{\mathrm{Mkv},\lambda}, {\Sigma^*}_{\mathrm{MG},\lambda}\big)$ by replacing the expectations by empirical averages. Similarly replace $\bar{\theta}^{(\lambda)}$ by $\widehat{\theta}_n$.
- Estimate the approximation factor $\alpha\big(M_\lambda, \frac{(1-\lambda)\gamma}{1-\lambda\gamma}\big)$ and the covariance $(I_d - M_\lambda)^{-1}({\Sigma^*}_{\mathrm{Mkv}} + {\Sigma^*}_{\mathrm{MG}})(I_d - M_\lambda)^{-\top}$ by plugging in the estimated matrices described above, for each $\lambda = \lambda_\ell$ with $\ell \in [m]$. Based on prior knowledge about the scale of the optimal approximation error $\inf_{v \in \mathbb{S}} \|v - v^*\|^2_{\mathbb{L}^2(\mathcal{S},\xi)}$, select λ_ℓ in the grid that minimizes our estimate of the total error according to equation (3.40).

Note that the procedure above is simply a sketch; a formal proof of correctness would show bounds that are uniform over all m estimators. It is an important direction of future work to provide sharp non-asymptotic analysis of such a model selection procedure.

3.4.3 Autoregressive models

Next, we turn to Example 3.3, the multivariate auto-regressive model. We study the stochastic approximation procedure in which, for any $i \in [k]$, we have

$$A^{(i)}_{t+1} = A^{(i)}_t - \eta\Big(\sum_{j=0}^{k-1} A^{(j)}_t X_{t-j} X^\top_{t+1-i} - X_{t+1} X^\top_{t+1-i}\Big), \quad \text{and} \quad \widehat{A}^{(i)}_n = \frac{1}{n - n_0} \sum_{t=n_0}^{n-1} A^{(i)}_t.$$

The first step in our analysis is to establish necessary and sufficient conditions for the existence and uniqueness of the stationary distribution of the process (3.17). The following $km \times km$ matrix plays a crucial role in this context:

$$R_* = \begin{bmatrix} A^*_1 & A^*_2 & \cdots & & A^*_k \\ I_m & 0 & \cdots & & 0 \\ 0 & I_m & 0 & \cdots & 0 \\ 0 & & \ddots & & 0 \\ 0 & \cdots & 0 & I_m & 0 \end{bmatrix}.$$

In the noiseless case, the stability of the linear dynamical system is equivalent to the following *Lyapunov stability condition* (see e.g. [160], Section 3.3):

$$\exists P_* \succ 0, Q_* \succ 0, \quad \text{such that } R_*^\top P_* R_* = P_* - Q_*. \tag{3.41}$$

Clearly we have $P_* \succ Q_*$. We let $\beta := \lambda_{\max}(P_*)$ and $\mu := \lambda_{\min}(Q_*)$. Based on stability theory for discrete-time linear systems [29], condition (3.41) is necessary for the stationary distribution to exist. In the following proposition, we show that this condition is also sufficient, with a concrete mixing time bound.

Proposition 3.3. *Under the Lyapunov stability condition (3.41) and assuming that the noise has bounded first moment $\mathbb{E}\big[\|\varepsilon_t\|_2\big] < \infty$, the stationary distribution $\widetilde{\xi}$ for the sliding window $\Omega_t = (X_{t+1}, X_t, \cdots, X_{t-k+1})$ of the auto-regressive process (3.17) exists and is unique. Furthermore, the mixing assumption 3.1 is satisfied with Wasserstein distance in $\mathbb{R}^{(k+1)m}$ and a mixing time bound $t_{\mathrm{mix}} = ck + c\frac{\beta}{\mu}\big(1 + \log\frac{\beta}{\mu}\big)$.*

See Section B.6.3.1 for the proof of this claim.

In addition to this mixing guarantee, we also make the following assumptions on the noise:

$$\mathbb{E}\big[\varepsilon_t\big] = 0, \quad \sup_{u \in \mathbb{S}^{d-1}} \mathbb{E}\big[\langle u,\, \varepsilon_t\rangle^4\big] \le \varsigma^4, \quad \text{and} \quad \|\varepsilon_t\|_2 \le \varsigma\sqrt{m}, \text{ a.s.} \tag{3.42}$$

We are now in a position to consider the problem of parameter estimation using stochastic approximation. Consider the vectorized version of the parameter $\theta = \mathrm{vec}\big(\big[A^{(1)}; A^{(2)}; \cdots; A^{(k)}\big]\big) \in \mathbb{R}^{km^2}$. The population-level Yule–Walker estimation equation (3.18) can be written as

$$\big(\underbrace{\big[\Gamma_{j-i}\big]_{i,j\in[k]}}_{H^*} \otimes I_m\big)\theta = \mathrm{vec}\big(\big[\Gamma_1; \Gamma_2; \cdots; \Gamma_k\big]\big), \tag{3.43}$$

where $\Gamma_i := \mathbb{E}\big[X_i X_0^\top\big] \in \mathbb{R}^{m\times m}$, for $i \in \mathbb{Z}$. We assume that

$$\frac{1}{2}\big(H^* + (H^*)^\top\big) \succeq h^* I_{km}, \quad \text{for some } h^* > 0.$$

In order to state the main corollary of Theorem 3.1 to auto-regressive models, the following quantities are relevant:

$$\varepsilon_{\mathrm{Mkv}}(\Omega_t) := \mathrm{vec}\Big(\Big(\sum_{j=0}^{k-1} A_*^{(j)} X_{t-j} - X_{t+1}\Big)\cdot\big[X_{t-1}^\top \quad X_{t-2}^\top \quad \cdots X_{t-k}^\top\big]\Big)$$

$$\Sigma^*{}_{\mathrm{Mkv}} := \sum_{t=-\infty}^{\infty} \mathbb{E}\big[\varepsilon_{\mathrm{Mkv}}(\Omega_t)\varepsilon_{\mathrm{Mkv}}(\Omega_0)^\top\big].$$

Corollary 3.5. *Under the setup above, take the stepsize and burn-in period as*

$$\eta = \frac{1}{c\Big(n^2\left(\frac{\beta}{\mu}\log\frac{\beta}{\mu}\right)(h^*)^2\varsigma^4k^3m^2\beta^8/\mu^8\Big)^{1/3}}, \quad \textit{and} \quad n_0 = \tfrac{1}{2}n, \tag{3.44a}$$

and suppose that $\frac{n}{\log^3 n} \geq \big(k + \frac{\beta}{\mu}\log\frac{\beta}{\mu}\big)\varsigma^4k^3m^2\frac{\beta^8}{\mu^8(h^*)^2}$. *Then the Polyak–Ruppert estimator* $(\widehat{A}_n^{(j)})_{j\in[k]}$ *satisfies*

$$\sum_{j=1}^{k}\mathbb{E}\big[\|\widehat{A}_n^{(j)} - A_j^*\|_F^2\big] \leq \tfrac{c}{n}\mathrm{Tr}\Big(\big(H^*\otimes I_m\big)^{-1}\Sigma_{\mathrm{Mkv}}\big(H^*\otimes I_m\big)^{-1}\Big)$$
$$+ \left\{\tfrac{km^2\cdot\lambda_{max}\left(\mathbb{E}\left[\varepsilon_{\mathrm{Mkv}}(s_0)\varepsilon_{\mathrm{Mkv}}(s_0)^\top\right]\right)}{(h^*)^2n}\Big(k + \frac{\beta}{\mu}\log\frac{\beta}{\mu}\Big)\right\}^{4/3}\log^2 n. \tag{3.44b}$$

A few remarks are in order. First, the leading-order term in the bound (3.44b) matches the variance of asymptotic efficient estimators for AR(m) models, up to a constant factor (see [29], Section 8). This simply follows from the fact that the plug-in Yule-Walker estimator is asymptotically efficient for auto-regressive models. On the other hand, Corollary 3.5 is completely non-asymptotic, holding true for any reasonably large sample size. Note that the sample complexity lower bound exhibits an $\mathcal{O}\left(\beta^9/\mu^9\right)$ dependency on the conditioning β/μ of the Lyapunov stability certificate (P_*, Q_*). In particular, a term linear in β/μ arises from the mixing time $\frac{\beta}{\mu}\log\frac{\beta}{\mu}$, and all other factors are from the almost-sure bounds on $\|X_t\|_2$ and moment bound $\sup_{u\in\mathbb{S}^{m-1}}\langle u,\, X_t\rangle^4$. If we instead assumed these quantities were bounded explicitly as in some prior work [109], the factor $\beta^8\varsigma^4k^2/\mu^8$ in the sample size requirement and stepsize choice can be replaced by such a bound.

3.5 Proof of Proposition 3.1

We begin by proving the bound on the last iterate claimed in Proposition 3.1. Define the error term $\Delta_t := \theta_t - \bar{\theta}$, as well as the noise terms

$$Z_{t+1} := L_{t+1} - \boldsymbol{L}(s_t), \qquad \zeta_{t+1} := (L_{t+1} - \boldsymbol{L}(s_t))\bar{\theta} + (b_{t+1} - \boldsymbol{b}(s_t)), \tag{3.45a}$$
$$N_t := \boldsymbol{L}(s_t) - \bar{L}, \qquad \nu_t := (\boldsymbol{L}(s_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}(s_t) - \bar{b}). \tag{3.45b}$$

Using this notation, we have the recursion

$$\Delta_{t+1} = (I - \eta(I - \bar{L}))\Delta_t + \eta\big(N_t + Z_{t+1}\big)\Delta_t + \eta(\nu_t + \zeta_{t+1}). \tag{3.46}$$

Taking squared norms on both sides yields the bound $\|\Delta_{t+1}\|_2^2 \leq \sum_{i=1}^4 T_i$, where

$$\begin{aligned}
T_1 &:= \|(I - \eta(I - \bar{L}))\Delta_t\|_2^2,\\
T_2 &:= 2\eta\langle(I - \eta(I - \bar{L}))\Delta_t,\, N_t\Delta_t + \nu_t\rangle,\\
T_3 &:= 2\eta\langle(I - \eta(I - \bar{L}))\Delta_t,\, \big(Z_{t+1}\Delta_t + \zeta_{t+1}\big)\rangle, \quad \text{and}\\
T_4 &:= 4\eta^2\big(\|N_t\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2 + \|\nu_t\|_2^2\big).
\end{aligned}$$

Beginning with the term T_1, expanding the square and then invoking the condition (3.4) yields

$$\begin{aligned} T_1 &= \|\Delta_t\|^2 - 2\eta\langle \Delta_t,\, (I-\bar{L})\Delta_t\rangle + \eta^2 \|(I-\bar{L})\Delta_t\|^2 \\ &\leq \big(1 - 2\eta(1-\kappa) + 2\eta^2(1+\gamma_{\max}^2)\big)\|\Delta_t\|^2. \end{aligned}$$

As for the cross terms involved in T_2 and T_3, we note that

$$\begin{aligned} 2\langle (I-\bar{L})\Delta_t,\, N_t\Delta_t\rangle &\leq \|(I-\bar{L})\Delta_t\|_2^2 + \|N_t\Delta_t\|_2^2 \;\leq\; 2(1+\gamma_{\max}^2)\|\Delta_t\|_2^2 + \|N_t\Delta_t\|_2^2, \\ 2\langle (I-\bar{L})\Delta_t,\, \nu_t\rangle &\leq \|(I-\bar{L})\Delta_t\|_2^2 + \|\nu_t\|_2^2 \;\leq\; 2(1+\gamma_{\max}^2)\|\Delta_t\|_2^2 + \|\nu_t\|_2^2, \\ 2\langle (I-\bar{L})\Delta_t,\, Z_{t+1}\Delta_t\rangle &\leq \|(I-\bar{L})\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2 \;\leq\; 2(1+\gamma_{\max}^2)\|\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2, \\ 2\langle (I-\bar{L})\Delta_t,\, \zeta_{t+1}\rangle &\leq \|(I-\bar{L})\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2 \;\leq\; 2(1+\gamma_{\max}^2)\|\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2. \end{aligned}$$

We collect the above bounds on the sum $\sum_{i=1}^4 T_i$ and use the stepsize bound $\eta \leq \frac{1-\kappa}{12(1+\gamma_{\max}^2)}$, which results in the recursive inequality

$$\begin{aligned} \|\Delta_{t+1}\|_2^2 \leq \big(1-\eta(1-\kappa)\big)\|\Delta_t\|_2^2 &+ 2\eta \underbrace{\big(\langle \Delta_t,\, N_t\Delta_t\rangle + \langle \Delta_t,\, \nu_t\rangle\big)}_{:=H_1(t)} \\ &+ 2\eta \underbrace{\big(\langle \Delta_t,\, Z_{t+1}\Delta_t\rangle + \langle \Delta_t,\, \zeta_{t+1}\rangle\big)}_{:=H_2(t)} \\ &+ 8\eta^2 \underbrace{\big(\|N_t\Delta_t\|_2^2 + \|Z_{t+1}\Delta_t\|_2^2 + \|\zeta_{t+1}\|_2^2 + \|\nu_t\|_2^2\big)}_{:=H_3(t)}. \end{aligned}$$

Multiplying both sides by $e^{\eta(1-\kappa)(t+1)}$ and using the fact that $\big(1-\eta(1-\kappa)\big) \leq e^{-\eta(1-\kappa)}$, we have

$$e^{\eta(1-\kappa)(t+1)}\|\Delta_{t+1}\|_2^2 \leq e^{\eta(1-\kappa)t}\|\Delta_t\|_2^2 + 2\eta e^{\eta(1-\kappa)(t+1)}\big(H_1(t)+H_2(t)\big) + 8\eta^2 e^{\eta(1-\kappa)(t+1)}H_3(t).$$

Unrolling this expression yields

$$e^{\eta(1-\kappa)n}\|\Delta_n\|_2^2 \leq \|\Delta_0\|_2^2 + 2\eta\sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\big(H_1(t)+H_2(t)\big) + 8\eta^2\sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}H_3(t), \tag{3.47}$$

which is the key recursion underlying our analysis.

3.5.1 Analyzing the recursion (3.47)

Note that the running sum $M_2(n) := \sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}H_2(t)$ is, by construction, a martingale adapted to the filtration $(F_t)_{t\geq 0}$. In contrast, the analogous quantity defined in terms of the process H_1 is *not* an adapted martingale. In order to circumvent this obstacle, our proof is based on introducing a *surrogate version* $\widetilde{H}_1$ of the process H_1, such that the running sum

$$\widetilde{M_1}(n) := \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+\tau)} \widetilde{H}_1(t+\tau)$$

can be decomposed as a sum of τ martingales. See the proof of Lemma 3.1 for the details of the construction of $\widetilde{H}_1$. This decomposition allows us to apply standard maximal inequalities for martingales. Of course, we also need the bound the moments of the differences $\widetilde{H}_1(t) - H_1(t)$; see Lemma 3.1 for the bound that we provide on this difference.

We prove the MSE bounds and higher-moment bounds using slightly different analysis tools. In order to study the mean-squared error (the case $p = 1$), we note that both $\widetilde{M_1}(t)$ and $H_2(t)$ have zero expectation for any $t \geq 0$. Taking expectations on both sides of equation (3.47), we obtain the bound

$$e^{\eta(1-\kappa)n} \mathbb{E}\big[\|\Delta_n\|_2^2\big] \leq \|\Delta_0\|_2^2 + 2\eta \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)} \mathbb{E}\Big[\Big|H_1(t) - \widetilde{H}_1(t)\Big|\Big] \\ + 8\eta^2 \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)} \mathbb{E}\big[H_3(t)\big]. \tag{3.48}$$

For higher moments, our analysis of the recursion (3.47) is based on a Lyapunov function Φ_n and auxiliary function Λ_n given by

$$\Phi_n := \Big(\mathbb{E}\big[\sup_{0 \leq t \leq n} e^{\eta(1-\kappa)tp} \|\Delta_t\|_2^{2p}\big]\Big)^{1/p}, \quad \text{and} \quad \Lambda_n = \max_{t \in \{0,1,\ldots,n\}} e^{-\frac{\eta(1-\kappa)t}{2}} \Phi_t.$$

By applying Minkowski's inequality to the recursion (3.47), we obtain the upper bound

$$\Phi_n \leq \Phi_0 + 4\eta\Big(\mathbb{E} \sup_{0 \leq t \leq n} |\widetilde{M_1}(t)|^p\Big)^{1/p} + 4\eta\Big(\mathbb{E}\Big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t} |H_1(t) - \widetilde{H}_1(t)|\Big)^p\Big)^{1/p} \\ + 4\eta\Big(\mathbb{E} \sup_{0 \leq t \leq n} |M_2(t)|^p\Big)^{1/p} + 16\eta^2\Big(\mathbb{E}\Big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t} H_3(t)\Big)^p\Big)^{1/p}. \tag{3.49}$$

In order to complete the proof, we need to control each of the terms on the right-hand side. The following auxiliary results provide the needed control; in all cases, the quantities (c, c_0) etc. denote universal constants; the number n in the following lemmas is seen as a general iteration index, instead of the total sample size in the final statement of the theorem.

Our first auxiliary result guarantees the existence of the surrogate variables $\widetilde{H}_1(t)$ with desirable properties:

Lemma 3.1. *There is a surrogate version $\{\widetilde{H}_1(t)\}_{t\geq 0}$ of the process $\{H_1(t)\}_{t\geq 0}$ such that $\mathbb{E}\big[\widetilde{H}(t)\big] = 0$ for any $t \geq 0$, and for any integer $p \in [1, \bar{p}/2]$, scalar $\tau \geq cpt_{\mathrm{mix}} \log(c_0 t_{\mathrm{mix}} d)$ and stepsize $\eta \leq \frac{1}{ct_{\mathrm{mix}}(\gamma_{\max} + p\sigma_L d)}$, we have the following bounds for any $n > 0$:*

$$\Big(\mathbb{E}\Big[\Big|H_1(n) - \widetilde{H}_1(n)\Big|^p\Big]\Big)^{1/p} \leq c\eta p^2 \tau\Big((d\sigma_L^2 + \gamma_{\max}^2) \cdot \big(\mathbb{E}\|\Delta_{n-\tau\vee 0}\|_2^{2p}\big)^{\frac{1}{p}} + \bar{\sigma}^2 d\Big), \tag{3.50a}$$

and for any $p \geq 2$, we have that

$$\big(\mathbb{E}\sup_{0\leq t\leq n} |\widetilde{M}_1(t)|^p\big)^{1/p} \leq \frac{cp^{3/2}}{\sqrt{\eta(1-\kappa)}}\big(\sigma_L\sqrt{d}\Phi_n + \bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big). \tag{3.50b}$$

See Section 3.5.2 for the proof of this claim. We note that it is especially challenging to prove the bound (3.50a).

Our second auxiliary result is a more straightforward bound on a martingale supremum:

Lemma 3.2. *The process M_2 is a martingale adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. Furthermore, for each $p \in [1, \bar{p}/2]$, $\tau \geq 2pt_{\mathrm{mix}} \log(c_0 d)$ and $\eta \leq \frac{1}{c(\gamma_{\max} + \sigma_L d)\tau}$, for any $n > 0$, we have that*

$$\big(\mathbb{E}\sup_{0\leq t\leq n} |M_2(t)|^p\big)^{1/p} \leq \frac{cp^{3/2}\tau^{1/2}}{\sqrt{\eta(1-\kappa)}}\big(\sigma_L\sqrt{d}\Phi_n + \bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big). \tag{3.51}$$

See Section 3.5.3 for the proof of this claim.

Finally, our third auxiliary result provides control on the process $H_3(t)$:

Lemma 3.3. *There is a universal constant c such that given $\tau \geq cpt_{\mathrm{mix}} \log(c_0 t_{\mathrm{mix}} d)$ and stepsize $\eta \leq \frac{1}{ct_{\mathrm{mix}}(\gamma_{\max} + \sigma_L d)}$, for any $p \in [1, \bar{p}/2]$, we have*

$$\big(\mathbb{E}\big[H_3(t)^p\big]\big)^{1/p} \leq c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p} + cp^2\bar{\sigma}^2 d. \tag{3.52}$$

See Section 3.5.4 for the proof of this claim.

We now use these three lemmas to complete the proof of Proposition 3.1. We prove the case of $\bar{p} = 2$ and $\bar{p} \geq \log n$ separately.

Proof in the case of $\bar{p} = 2$: By Lemma 3.1 with $\tau = ct_{\mathrm{mix}} \log(c_0 t_{\mathrm{mix}} d)$ and Cauchy–Schwarz inequality, we have that

$$\begin{aligned}
\mathbb{E}\Big[\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}|\widetilde{H}_1(t) - H_1(t)|\Big] &\leq c\eta\tau \sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\Big((\sigma_L^2 d + \gamma_{\max}^2)\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^2\big] + \bar{\sigma}^2 d\Big) \\
&\leq \frac{c\tau\bar{\sigma}^2 d}{1-\kappa} e^{\eta(1-\kappa)n} + ce\eta\tau(\sigma_L^2 d + \gamma_{\max}^2)\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^2\big].
\end{aligned}$$

Similarly, by applying Lemma 3.3 to the last term of equation (3.48), we obtain the bound

$$\sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+1)}\mathbb{E}\big[H_3(t)\big] \leq \frac{c\bar{\sigma}^2 d}{(1-\kappa)\eta} e^{\eta(1-\kappa)n} + ce(\sigma_L^2 d + \gamma_{\max}^2)\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^2\big].$$

Combining them with the decomposition (3.48), for any $n = 1, 2, \cdots$, we find that $e^{\eta(1-\kappa)n}\mathbb{E}\big[\|\Delta_n\|_2^2\big]$ is upper bounded by

$$\|\Delta_0\|_2^2 + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa} e^{\eta(1-\kappa)n} + c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2)\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^2\big]. \tag{3.53}$$

In order to exploit this recursive upper bound, we define the partial sum sequence $S_n := \sum_{t=0}^{n} e^{\eta(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^2\big]$. Equation (3.48) implies that

$$\begin{aligned} S_n &\leq S_0 + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa} e^{\eta(1-\kappa)n} + \big(1 + c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2)\big)S_{n-1} \\ &\leq S_0 \cdot \sum_{t=0}^{n} e^{c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2)t} + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa}\cdot\sum_{t=0}^{n} e^{c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2)t + \eta(1-\kappa)(n-t)} \\ &\leq \frac{3}{(1-\kappa)\eta} e^{\eta(1-\kappa)n/3} S_0 + \frac{3c\tau\bar{\sigma}^2 d}{(1-\kappa)^2} e^{\eta(1-\kappa)n}. \end{aligned}$$

Substituting back into the recursion (3.53) yields

$$\begin{aligned} \mathbb{E}\big[\|\Delta_n\|_2^2\big] &\leq \frac{6}{(1-\kappa)\eta} e^{-\eta(1-\kappa)n/3}\|\Delta_0\|_2^2 + c\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa} + c\eta^2\tau(\sigma_L^2 d + \gamma_{\max}^2)\cdot\frac{2c\tau\bar{\sigma}^2 d}{(1-\kappa)^2} \\ &\leq e^{-\eta(1-\kappa)n/2}\|\Delta_0\|_2^2 + c'\frac{\eta\tau\bar{\sigma}^2 d}{1-\kappa}, \end{aligned}$$

which completes the proof of the MSE bound.

Proof in the case of $\bar{p} \geq \log n$: Now we turn to prove the p-th moment bound under Assumption 3.2 with $\bar{p} \geq \log n$. Recall that we analyze the growth of the Lyapunov function Φ_n, and we start from the decomposition (3.49).

The first term in equation (3.49) is simply $\|\Delta_0\|_2^2$, and the second term is controlled using equation (3.50b) in Lemma 3.1. In order to bound the third term, we apply Hölder's inequality, and obtain the bound

$$\mathbb{E}\Big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\Big|H_1(t) - \widetilde{H}_1(t)\Big|\Big)^p \leq \Big(\sum_{t=0}^{n-1} e^{\frac{\eta(1-\kappa)pt}{2(p-1)}}\Big)^{p-1}\cdot\sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}}\mathbb{E}\Big[\Big|H_1(t) - \widetilde{H}_1(t)\Big|^p\Big].$$

By equation (3.50a) in Lemma 3.1, this quantity is at most

$$(\eta(1-\kappa))^{1-p} e^{\frac{\eta(1-\kappa)pn}{2}}\sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}}\Big(c\tau\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p} + c\tau p^2\bar{\sigma}^2 d\Big)^p.$$

We then obtain the inequality:

$$
\begin{aligned}
&\big(\mathbb{E}\big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\left|H_1(t)-\widetilde{H}_1(t)\right|\big)^p\big)^{1/p}\\
&\le cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2\tau d + c\big(p^2\sigma_L^2 d+\gamma_{\max}^2\big)\tau\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}\big(\sum_{t=0}^{n-1} e^{\frac{1}{2}\eta p(1-\kappa)t}\mathbb{E}\big[\|\Delta_t\|_2^{2p}\big]\big)^{1/p}\\
&\le cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2\tau d + c\big(p^2\sigma_L^2 d+\gamma_{\max}^2\big)\tau\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}\big(\sum_{t=0}^{n-1} e^{-\frac{1}{2}\eta p(1-\kappa)t}\Phi_t^p\big)^{1/p}\\
&\le cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2\tau d + c\big(p^2\sigma_L^2 d+\gamma_{\max}^2\big)\tau\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}n^{1/p}\Lambda_n.
\end{aligned}
$$

Similarly, the fourth term on the right hand side is controlled using Lemma 3.2, and the bounds for the last term are based on Lemma 3.3 and the same strategy as above. Concretely, combining Hölder's inequality with the bound (3.52) yields

$$
\mathbb{E}\big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}H_3(t)\big)^p \le \big(\sum_{t=0}^{n-1} e^{\frac{\eta(1-\kappa)pt}{2(p-1)}}\big)^{p-1}\cdot\sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}}\mathbb{E}[H_3(t)^p].
$$

This quantity is at most

$$
(\eta(1-\kappa))^{1-p}e^{\frac{\eta(1-\kappa)pn}{2}}\sum_{t=0}^{n-1} e^{\frac{\eta p(1-\kappa)t}{2}}\big(c\big(p^2\sigma_L^2 d+\gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p}+cp^2\bar{\sigma}^2 d\big)^p.
$$

Noting that each term satisfies the inequality $e^{\frac{\eta p(1-\kappa)t}{2}}\big(\mathbb{E}\big[\|\Delta_{t-\tau\vee 0}\|_2^{2p}\big]\big)^{1/p}\le\Lambda_n$ for $t\in[0,n]$. We conclude that the moment $\big(\mathbb{E}\big(\sum_{t=0}^{n-1}e^{\eta(1-\kappa)t}H_3(t)\big)^p\big)^{1/p}$ is upper bounded by

$$
cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2 d + c\big(p^2\sigma_L^2 d+\gamma_{\max}^2\big)\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}n^{1/p}\Lambda_n.
$$

Collecting the above bounds and substituting into the decomposition (3.49), we note that

$$
\begin{aligned}
\Phi_n &\le \Phi_0 + c\sqrt{\frac{p^3\eta}{1-\kappa}}\big(\sigma_L\sqrt{d}\Phi_n+\bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big)\\
&\qquad + cp^2\frac{e^{\eta(1-\kappa)n}}{\eta(1-\kappa)}\bar{\sigma}^2\tau d + \big(p^2\sigma_L^2 d+\gamma_{\max}^2\big)\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{\eta(1-\kappa)}\tau n^{1/p}\Lambda_n\\
&\le \Phi_0 + 4c\sigma_L\sqrt{\frac{p^3\tau\eta d}{1-\kappa}}\Phi_n + \frac{1}{4}\Phi_n + c\eta\frac{\bar{\sigma}^2p^3 d\tau}{1-\kappa}\cdot e^{\eta(1-\kappa)n}\\
&\qquad + cp^2\eta\frac{e^{\eta(1-\kappa)n}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\big(p^2\sigma_L^2 d+\gamma_{\max}^2\big)\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{1-\kappa}\tau\Lambda_n
\end{aligned}
$$

In the last step, we apply Young's inequality to the term $\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}$, and use the condition $p \geq \log n$ to the last term so that $n^{1/p} \leq e$.

Taking the stepsize $\eta \leq \frac{1-\kappa}{64c^2\sigma_L^2\tau d p^3}$, we arrive at the following bound valid for any $n \in [1, e^p]$:

$$e^{-\frac{\eta(1-\kappa)n}{2}}\Phi_n \leq 2\Phi_0 + cp^3\eta\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\frac{p^2\sigma_L^2 d + \gamma_{\max}^2}{1-\kappa}\tau\Lambda_n.$$

Note that the right-hand-side of above expression is monotonic increasing in the index n. For any integer pair (t, n) such that $0 < t \leq n \leq e^p$, we have the inequality:

$$\begin{aligned} e^{-\frac{\eta(1-\kappa)t}{2}}\Phi_t &\leq 2\Phi_0 + cp^3\eta\frac{e^{\frac{1}{2}\eta(1-\kappa)t}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\frac{p^2\sigma_L^2 d + \gamma_{\max}^2}{1-\kappa}\tau\Lambda_t \\ &\leq 2\Phi_0 + cp^3\eta\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\frac{p^2\sigma_L^2 d + \gamma_{\max}^2}{1-\kappa}\tau\Lambda_n. \end{aligned}$$

Given the value of n fixed and taking supremum over $t \in \{0, 1, 2, \cdots, n\}$ in the left-hand-side, we arrive at the conclusion:

$$\Lambda_n = \sup_{t\in\{0,1,\cdots,n\}} e^{-\frac{\eta(1-\kappa)t}{2}}\Phi_t \leq 2\Phi_0 + cp^3\eta\frac{e^{\frac{1}{2}\eta(1-\kappa)n}}{1-\kappa}\bar{\sigma}^2\tau d + c\eta\frac{p^2\sigma_L^2 d + \gamma_{\max}^2}{1-\kappa}\tau\Lambda_n.$$

Given the stepsize $\eta \leq \frac{1-\kappa}{2c(p^3\sigma_L^2 d + \gamma_{\max}^2)\tau}$, we arrive at the bound

$$\left(\mathbb{E}\|\Delta_t\|_2^p\right)^{1/p} \leq e^{-\frac{1}{2}\eta(1-\kappa)n}\Lambda_n \leq e^{-\frac{1}{2}\eta(1-\kappa)n}\left(\mathbb{E}\|\Delta_0\|_2^p\right)^{1/p} + \frac{cp^3\eta}{1-\kappa}\bar{\sigma}^2\tau d,$$

which completes the proof of the theorem.

It remains to prove our three auxiliary lemmas.

3.5.2 Proof of Lemma 3.1

We break the proof into three steps. In the first step, given in Section 3.5.2.1, we construct the surrogate process, whereas the remaining two steps are devoted to the proving the bounds (3.50b) and (3.50a), as detailed in Sections 3.5.2.2 and 3.5.2.3 respectively.

3.5.2.1 Construction of the surrogate process

We first claim that for any $t = 1, 2, \ldots$ and any $\tau \in \{0, \ldots, t\}$, there is a random variable $\widetilde{s}_t \in \mathcal{S}$ such that $\widetilde{s}_t \mid \mathcal{F}_{t-\tau} \sim \xi$, and

$$\left(\mathbb{E}\left[\rho(s_t, \widetilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\right]\right)^{1/p} \leq c_0 \exp\left(-\tfrac{\tau}{2t_{\mathrm{mix}}p}\right) \quad \text{for each } p \geq 2. \tag{3.54}$$

Here c_0 is a universal constant.

Our construction is based on the following bound on the Wasserstein distance:

Lemma 3.4. *Under Assumptions 3.1 and 3.3, the Wasserstein distance is upper bounded as*

$$\mathcal{W}_{1,\rho}\big(\delta_x P^\tau, \xi\big) \leq c_0 2^{-\lfloor \frac{\tau}{t_{\mathrm{mix}}} \rfloor},$$

valid for any $x \in \mathcal{S}$ and $\tau \geq 0$.

See Appendix B.3.1 for the proof of this claim.

We now use Lemma 3.4 to construct the desired process. We begin by constructing a coupling conditionally on the σ-field $\mathcal{F}_{t-\tau}$: let $\widetilde{s}_t$ be a state whose conditional law is ξ, satisfying the identity:

$$\mathbb{E}\big[\rho(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big] = \mathcal{W}_{1,\rho}\big(\mathcal{L}(s_t \mid \mathcal{F}_{t-\tau}), \xi\big). \tag{3.55}$$

The existence of such $\widetilde{s}_t$ is guaranteed by the definition of Wasserstein distance. We now bound the relevant quantities based on this construction.

Combining the identity (3.55) with Lemma 3.4 yields $\mathbb{E}\big[\rho(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big] \leq c_0 \cdot 2^{-\lfloor \frac{\tau}{t_{\mathrm{mix}}} \rfloor}$. Applying Cauchy–Schwarz inequality and invoking Assumption 3.3, we find that

$$\begin{aligned}
\Big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\big]\Big)^{1/p} &\leq \Big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big]\Big)^{\frac{1}{2p}} \cdot \Big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^{2p-1} \mid \mathcal{F}_{t-\tau}\big]\Big)^{\frac{1}{2p}} \\
&\leq \Big(\mathbb{E}\big[\rho(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big]\Big)^{\frac{1}{2p}} \\
&\leq c_0 \cdot 2^{1 - \frac{\tau}{2 t_{\mathrm{mix}} p}},
\end{aligned} \tag{3.56}$$

which establishes the claim.

We now use the sequence of random variables $\widetilde{s}_t$ just constructed to define the extended filtration $\widetilde{\mathcal{F}}_t := \sigma\big((s_k)_{0 \leq k \leq t}, (\widetilde{s}_k)_{0 \leq k \leq t}, \big((L_k, b_k)\big)_{0 \leq k \leq t}\big)$, as well as the surrogate quantities

$$\begin{aligned}
\widetilde{\nu}_t &:= \big(\boldsymbol{L}(\widetilde{s}_t) - \bar{L}\big)\bar{\theta} + \big(\boldsymbol{b}(\widetilde{s}_t) - \bar{b}\big), \quad \text{and} \\
\widetilde{H}_1(t) &:= \langle \Delta_{(t-\tau)\vee 0}, \widetilde{\nu}_t \rangle + \langle \Delta_{(t-\tau)\vee 0}, \big(\boldsymbol{L}(\widetilde{s}_t) - \bar{L}\big)\Delta_{(t-\tau)\vee 0} \rangle.
\end{aligned}$$

Note that by definition, we have $\mathbb{E}\big[\widetilde{H}_1(t) \mid \widetilde{\mathcal{F}}_{(t-\tau)\vee 0}\big] = 0$ for each $t = 0, 1, 2, \ldots$.

3.5.2.2 Proof of the bound (3.50b)

We first perform a decomposition on the process $\widetilde{M}_1$. In particular, for $\ell \in \{0, 1, \cdots, \tau - 1\}$, we define the stochastic process $\widetilde{M}_1^{(\ell)}(n) := \sum_{t=0}^{n-1} e^{\eta(1-\kappa)(t+\tau)} \widetilde{H}_1(t+\tau) \mathbf{1}_{\{t \bmod \tau = \ell\}}$. Clearly, we have $\widetilde{M}_1(n) = \sum_{\ell=0}^{\tau-1} \widetilde{M}_1^{(\ell)}(n)$ for any $n \geq 0$. Furthermore, we note for any $t \geq 0$, we have the relations:

$$\mathbb{E}\big[\widetilde{H}_1(t+\tau) \mid \widetilde{\mathcal{F}}_t\big] = 0, \quad \text{and} \quad \widetilde{H}_1(t) \in \widetilde{\mathcal{F}}_t.$$

So for each $\ell \in [0, \tau - 1]$, the process $\widetilde{M}_1^{(\ell)}$ is a martingale adapted to the filtration $(\widetilde{\mathcal{F}}_t)_{t \geq 0}$.

By the BDG inequality, we have the maximal inequality $\big(\mathbb{E} \sup_{0 \leq t \leq n} |\widetilde{M}_1^{(\ell)}(t)|^p\big)^{1/p} \leq cp\big(\mathbb{E}\big([\widetilde{M}_1^{(\ell)}]_n\big)^{p/2}\big)^{1/p}$, valid for all $\ell = 0, 1, \ldots, \tau - 1$. Similarly, for the quadratic variation term $[\widetilde{M}_1^{(\ell)}]_n$, we have that

$$\begin{aligned}\mathbb{E}\big[\big([\widetilde{M}_1^{(\ell)}]_n\big)^{p/2}\big] &= \mathbb{E}\big[\big(\sum_{k=0}^{\lfloor \frac{n-1}{\tau} \rfloor} e^{\eta(1-\kappa)(k\tau+\tau+\ell)} \|\widetilde{H}_1(k\tau + \ell)\|_2^2\big)^{p/2}\big] \\ &\leq \big(\sum_{k=0}^{\lfloor \frac{n-1}{\tau} \rfloor} e^{\eta(1-\kappa)p(k\tau+\tau+\ell)} \mathbb{E}\big[\|\widetilde{H}_1(k\tau + \ell)\|_2^p\big]\big) \cdot \big(\sum_{t=0}^{n-1} e^{-\frac{p^2}{2p-4}\tau\eta(1-\kappa)t}\big)^{\frac{p-2}{2}},\end{aligned}$$

which is at most

$$\begin{aligned}\big(\eta\tau(1-\kappa)\big)^{-\frac{p}{2}+1} \sum_{t=\tau}^{n-1} e^{\eta(1-\kappa)tp} \big(&\mathbb{E}\big[\,\big|2\langle \Delta_{t-\tau},\, (\bar{L}(\widetilde{s}_t) - \bar{L})\Delta_{t-\tau}\rangle\big|^p\,\big] \\ &+ \mathbb{E}\big[\,|2\langle \widetilde{\nu}_t,\, \Delta_{t-\tau}\rangle|^p\,\big]\big) \mathbf{1}_{\{t \bmod \tau = \ell\}}.\end{aligned}$$

Invoking the tail condition in Assumption 3.2 under the stationary distribution, we have that

$$\begin{aligned}\mathbb{E}\big[\,\big|2\langle \Delta_{t-\tau},\, (\bar{L}(\widetilde{s}_t) - \bar{L})\Delta_{t-\tau}\rangle\big|^p \mid \mathcal{F}_{t-\tau}\big] &\leq \big(p\sigma_L\sqrt{d} \cdot \|\Delta_{t-\tau}\|_2^2\big)^p, \quad \text{and} \\ \mathbb{E}\big[\,|\langle \widetilde{\nu}_t,\, \Delta_{t-\tau}\rangle|^p \mid \mathcal{F}_{t-\tau}\big] &\leq \big(p\bar{\sigma}\sqrt{d} \cdot \|\Delta_{t-\tau}\|_2\big)^p.\end{aligned}$$

Substituting into the moment bounds for $[\widetilde{M}_1^{(\ell)}]_n$ and combining the results for $\ell = 0, 1, \cdots, \tau - 1$ using Minkowski's inequality, we arrive at the bound

$$\begin{aligned}&\big(\mathbb{E} \sup_{0 \leq t \leq n} |\widetilde{M}_1(t)|^p\big)^{1/p} \\ &\leq \sum_{\ell=0}^{\tau-1} \big(\mathbb{E} \sup_{0 \leq t \leq n} |\widetilde{M}_1^{(\ell)}(t)|^p\big)^{1/p} \\ &\leq \frac{\tau \cdot n^{\frac{1}{p}} \sqrt{p}}{\big(\eta\tau(1-\kappa)\big)^{\frac{1}{2}+\frac{1}{p}}} \big\{p\sigma_L\sqrt{d} \cdot \max_{0 \leq t \leq n} \big[e^{\eta(1-\kappa)t} \big(\mathbb{E}\|\Delta_t\|_2^{2p}\big)^{1/p}\big] \\ &\qquad + e^{\frac{\eta(1-\kappa)n}{2}} p\bar{\sigma}\sqrt{d} \max_{0 \leq t \leq n} \big[e^{\eta(1-\kappa)t/2} \big(\mathbb{E}\|\Delta_t\|_2^{p}\big)^{1/p}\big]\big\} \\ &\leq \sqrt{\frac{\tau p}{\eta(1-\kappa)}} \big(p\sigma_L\sqrt{d}\Phi_n + p\bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big),\end{aligned}$$

which completes the proof of this lemma.

3.5.2.3 Proof of the bound (3.50a)

By Minkowski's inequality, we can upper bound the error as $\big(\mathbb{E}\big[(H_1(t) - \widetilde{H}_1(t))^p\big]\big)^{1/p} \leq \sum_{k=1}^{6} J_k$, where

$$\begin{aligned} J_1 &:= \big(\mathbb{E}\big[\langle \Delta_{t-\tau},\, \nu_t - \widetilde{\nu}_t\rangle^p\big]\big)^{1/p}, & J_3 &:= \big(\mathbb{E}\big[\langle \Delta_{t-\tau},\, \big(\boldsymbol{L}(\widetilde{s}_t) - \boldsymbol{L}(s_t)\big)\Delta_{t-\tau}\rangle^p\big]\big)^{1/p}, \\ J_2 &:= \big(\mathbb{E}\big[\langle \Delta_t - \Delta_{t-\tau},\, \nu_t\rangle^p\big]\big)^{1/p} & J_4 &:= \big(\mathbb{E}\big[\langle \Delta_t - \Delta_{t-\tau},\, N_t\Delta_{t-\tau}\rangle^p\big]\big)^{1/p} \\ J_5 &:= \big(\mathbb{E}\big[\langle \Delta_t,\, N_t(\Delta_t - \Delta_{t-\tau})\rangle^p\big]\big)^{1/p}, & J_6 &:= \big(\mathbb{E}\big[\langle \Delta_t - \Delta_{t-\tau},\, N_t(\Delta_t - \Delta_{t-\tau})\rangle^p\big]\big)^{1/p} \end{aligned}$$

The terms J_1 and J_3 can be controlled using the bound on $\rho(s_t, \widetilde{s}_t)$ and the Lipschitz condition (3.4); doing so yields the bound

$$\begin{aligned} J_1 &\leq \bar{\sigma} d\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^p \cdot \mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\big]\big]\big)^{1/p} \leq 2c_0\bar{\sigma} d\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^p\big)^{1/p} \cdot 2^{-\frac{\tau}{2pt_{\text{mix}}}}, \quad \text{and} \\ J_3 &\leq \sigma_L d\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^{2p} \cdot \mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^p \mid \mathcal{F}_{t-\tau}\big]\big]\big)^{1/p} \leq 2c_0\sigma_L d\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{1/p} \cdot 2^{-\frac{\tau}{2pt_{\text{mix}}}}. \end{aligned}$$

Given the time lag parameter $\tau \geq cpt_{\text{mix}} \log(c_0 t_{\text{mix}} d) \geq 2pt_{\text{mix}} \log\left(\frac{d}{\eta}\right)$, we have the bound

$$J_1 \leq \eta\bar{\sigma}\sqrt{d}\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^p\big)^{1/p}, \quad \text{and} \quad J_3 \leq \eta\eta\sigma_L\sqrt{d}\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{1/p}. \tag{3.57}$$

Turning to the J_2 term, applying the Cauchy–Schwarz inequality yields

$$J_2 \leq \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \big(\mathbb{E}\|\nu_t\|_2^{2p}\big)^{\frac{1}{2p}} \overset{(i)}{\leq} \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot p\bar{\sigma}\sqrt{d}. \tag{3.58}$$

where step (i) follows from Assumption 3.2.

The terms J_4 and J_5 can be controlled via once again replacing s_t with its surrogate $\widetilde{s}_t$. First, by Cauchy–Schwarz inequality, we note that

$$\begin{aligned} J_4 &\leq \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \big(\mathbb{E}\|N_t\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}, \\ J_5 &\leq \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \big(\mathbb{E}\|N_t^\top\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}. \end{aligned}$$

Using the decomposition $N_t = (\boldsymbol{L}(\widetilde{s}_t) - \bar{L}) + (\boldsymbol{L}(s_t) - \boldsymbol{L}(\widetilde{s}_t))$, we note that

$$\big(\mathbb{E}\|N_t\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \leq \big(\mathbb{E}\|(\boldsymbol{L}(\widetilde{s}_t) - \bar{L})\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} + \big(\mathbb{E}\|(\boldsymbol{L}(s_t) - \boldsymbol{L}(\widetilde{s}_t))\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}.$$

We bound the conditional expectations of the quantities above. The first term can be controlled via Assumption 3.2:

$$\mathbb{E}\big[\|(\boldsymbol{L}(\widetilde{s}_t) - \bar{L})\Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau}\big] \leq (\sigma_L p\sqrt{d})^{2p}\|\Delta_{t-\tau}\|_2^{2p},$$

and the second term is controlled using the Lipschitz condition 3.4:

$$\begin{aligned} \mathbb{E}\big[\|(\boldsymbol{L}(s_t) - \boldsymbol{L}(\widetilde{s}_t))\Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau}\big] &\leq (\sigma_L d)^{2p} \cdot \mathbb{E}\big[\rho(s_t, \widetilde{s}_t)^{2p} \mid \mathcal{F}_{t-\tau}\big] \cdot \|\Delta_{t-\tau}\|_2^{2p} \\ &\leq (\sigma_L d)^{2p} \cdot c_0 \cdot 2^{1-\frac{\tau}{t_{\text{mix}}}} \cdot \|\Delta_{t-\tau}\|_2^{2p}. \end{aligned}$$

Consequently, taking $\tau \geq 2t_{\mathrm{mix}} p \log(c_0 d)$, we have the bounds

$$\big(\mathbb{E}\|N_t \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \leq \sigma_L p\sqrt{d} \cdot \big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}, \quad \text{and}$$
$$\big(\mathbb{E}\|N_t^\top \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \leq \sigma_L p\sqrt{d} \cdot \big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}.$$

Putting together the pieces, we arrive at the bound

$$J_4 + J_5 \leq 2\big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}} \cdot \sigma_L p\sqrt{d} \cdot \big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{2p}}. \tag{3.59}$$

By the Lipschitz condition (3.4) and the assumed boundedness (3.3) of the metric space, the term J_6 admits the simple upper bound

$$J_6 \leq \big(\mathbb{E}\big[\|\!|N_t|\!\|_{\mathrm{op}}^p \|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big]\big)^{\frac{1}{p}} \leq \sigma_L d \big(\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{p}} \tag{3.60}$$

From all of these bounds, we see that the remaining crucial piece is to bound $\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}$. In order to do so, we require the following two helper lemmas

Lemma 3.5. *Given $p \geq 2$ and $\ell > 0$, the iterates (3.3a) with stepsize $\eta \leq \big(6(\gamma_{\max} + \sigma_L d)\ell\big)^{-1}$ satisfy the bound*

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq e\eta\ell(\gamma_{\max} + \sigma_L d)\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 3\eta p\ell\sqrt{d}\bar{\sigma}, \tag{3.61a}$$

and consequently,

$$\frac{1}{2}\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} - 6\eta p\ell\sqrt{d}\bar{\sigma} \leq \big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \leq e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\bar{\sigma}. \tag{3.61b}$$

See Appendix B.3.2 for the proof of this claim.

Our second auxiliary result is of a bootstrap nature: it is based on assuming that for some given an integer $p \geq 2$, fix any integer $\tau \geq 2t_{\mathrm{mix}} p \log(c_0 d)$, there exist positive scalars $\omega_p, \beta_p > 0$ such that

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \eta\omega_p \cdot \big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + \eta\beta_p\bar{\sigma} \tag{3.62}$$

for any $t \geq 0$, $\eta \leq \frac{1}{48(\gamma_{\max} + \sigma_L d)\tau}$ and $\ell \in [0, \tau]$. We then have the following guarantee:

Lemma 3.6. *When the condition (3.62) holds, then, for any $t \geq 0$, $\eta \leq \frac{1}{48(\gamma_{\max} + \sigma_L d)\tau}$, and $\ell \in [0, \tau]$, we have*

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \eta\Big(12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\ell + \frac{\omega_p}{2}\Big)\Big(\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + \eta p(\tau + \ell)\sqrt{d}\bar{\sigma}\Big)$$
$$+ \eta\Big(2p\ell\sqrt{d} + \frac{1}{2}\beta_p\Big)\bar{\sigma}. \tag{3.63}$$

See Appendix B.3.3 for the proof of this claim.

We now complete the proof of the bound (3.50a) by using a bootstrapping argument in order to obtain a sharp bound on $\mathbb{E}\|\Delta_t - \Delta_{t-\tau}\|_2^p$. Let $\omega_p^{(0)} := e\tau(\gamma_{\max} + \sigma_L d)$ and $\beta_p^{(0)} := p\tau\sqrt{d}$, and define the following recursion:

$$\begin{cases} \omega_p^{(i+1)} = \frac{1}{2}\omega_p^{(i)} + 12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\tau, \\ \beta_p^{(i+1)} = \frac{1}{2}\beta_p^{(i)} + 2p\tau\sqrt{d} + 2\eta\big(12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\tau + \frac{1}{2}\omega_p^{(i)}\big)p\tau\sqrt{d}. \end{cases}$$

It can be seen that as $i \to \infty$, the sequence $(\omega_p^{(i)}, \beta_p^{(i)})$ converges to a unique limit (ω_p^*, β_p^*); this limit is the unique fixed point of the iterates defined above.

By Lemma 3.6, if the iterates satisfy the bound (3.62) with constants $\big(\omega_p^{(i)}, \beta_p^{(i)}\big)$, then it also satisfy the bound with constants $\big(\omega_p^{(i+1)}, \beta_p^{(i+1)}\big)$. By Lemma 3.5, the iterates satisfy bound with constants $\big(\omega_p^{(0)}, \beta_p^{(0)}\big)$. An induction argument then yields the bound for any $\big(\omega_p^{(i)}, \beta_p^{(i)}\big)$. In particular, the bound is satisfied by the fixed point $\big(\omega_p^*, \beta_p^*\big)$.

Solving directly for the fixed-point equation, we find that

$$\omega_p^* = 24\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\tau, \quad \text{and} \quad \beta_p^* = 4p\tau\sqrt{d} + 96\eta\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)p\tau^2\sqrt{d}.$$

Taking the stepsize $\eta \leq \frac{1}{48(\gamma_{\max}+p\sigma_L d)\tau}$, we arrive at the bound

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq 24\eta\tau\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + 6\eta p\tau\sqrt{d}\bar{\sigma}, \tag{3.64}$$

for any $t \geq 0$ and $\ell \in [0, \tau]$.

Collecting the bounds (3.57), (3.58), (3.59), (3.60) and (3.64) and taking the stepsize $\eta \leq \frac{1}{c(\gamma_{\max}+p\sigma_L d)\tau}$, we arrive at the bound

$$\big(\mathbb{E}\big[(H_1(t) - \widetilde{H}_1(t))^p\big]\big)^{1/p} \leq c\eta p^2\tau\big((d\sigma_L^2 + \gamma_{\max}^2)\cdot\big(\mathbb{E}\|\Delta_{t-\tau}\|_2^{2p}\big)^{\frac{1}{p}} + \bar{\sigma}^2 d\big),$$

thereby completing the proof of the bound (3.50a).

3.5.3 Proof of Lemma 3.2

By the BDG inequality, we have $\big(\mathbb{E}\sup_{0\leq t\leq n}|M_2(t)|^p\big)^{1/p} \leq cp\big(\mathbb{E}\big([M_2]_n\big)^{p/2}\big)^{1/p}$, valid for all $\ell = 0, 1, \ldots, \tau - 1$.

As for the quadratic variation $[M_2]_n$, applying Hölder's inequality yields

$$\begin{aligned} \mathbb{E}\big[\big([M_2]_n\big)^{p/2}\big] &= \mathbb{E}\big[\big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)t}\|H_2(t)\|_2^2\big)^{p/2}\big] \\ &\leq \big(\sum_{t=0}^{n-1} e^{\eta(1-\kappa)tp}\mathbb{E}\big[\|H_2(t)\|_2^p\big]\big)\cdot\big(\sum_{t=0}^{n-1} e^{-\frac{p^2}{2p-4}\eta(1-\kappa)t}\big)^{\frac{p-2}{2}} \\ &\leq \big(\eta(1-\kappa)\big)^{-\frac{p}{2}+1}\sum_{t=0}^{n-1} e^{\eta(1-\kappa)tp}\big(\mathbb{E}\big[\,|2\langle\Delta_t,\, Z_{t+1}\Delta_t\rangle|^p\,\big] + \mathbb{E}\big[\,|2\langle\zeta_{t+1},\, \Delta_t\rangle|^p\,\big]\big). \end{aligned}$$

For the moment terms above, we invoke Assumption 3.2, and obtain the following bounds:

$$\mathbb{E}\big[\,|\langle \Delta_t,\, Z_{t+1}\Delta_t\rangle|^p \mid \mathcal{F}_t\big] \leq \|\Delta_t\|_2^p \cdot \mathbb{E}\big[\big(\sum_{j=1}^d \langle e_j,\, Z_{t+1}\Delta_t\rangle^2\big)^{p/2} \mid \mathcal{F}_t\big] \leq \big(p\sigma_L\sqrt{d}\cdot\|\Delta_t\|_2^2\big)^p,$$

$$\mathbb{E}\big[\,|\langle \zeta_{t+1},\, \Delta_t\rangle|^p \mid \mathcal{F}_t\big] \leq \|\Delta_t\|_2^p \cdot \mathbb{E}\big[\big(\sum_{j=1}^d \langle e_j,\, \zeta_{t+1}\rangle^2\big)^{p/2} \mid \mathcal{F}_t\big] \leq \big(p\bar{\sigma}\sqrt{d}\cdot\|\Delta_t\|_2\big)^p.$$

Substituting into the bound above, we find that

$$\begin{aligned}
&\Big(\mathbb{E}\big[\big([M_2]_n\big)^{p/2}\big]\Big)^{1/p}\\
&\leq \tfrac{(\eta(1-\kappa))^{-\frac{1}{p}}\cdot n^{\frac{1}{p}}}{\sqrt{\eta(1-\kappa)}}\Big\{p\sigma_L\sqrt{d}\cdot \max_{0\leq t\leq n}\Big[e^{\eta(1-\kappa)t}\big(\mathbb{E}\|\Delta_t\|_2^{2p}\big)^{1/p}\Big]\\
&\qquad + e^{\frac{\eta(1-\kappa)n}{2}}p\bar{\sigma}\sqrt{d}\max_{0\leq t\leq n}\Big[e^{\eta(1-\kappa)t/2}\big(\mathbb{E}\|\Delta_t\|_2^{p}\big)^{1/p}\Big]\Big\}\\
&\leq \tfrac{1}{\sqrt{\eta(1-\kappa)}}\big(p\sigma_L\sqrt{d}\Phi_n + p\bar{\sigma}\sqrt{e^{\eta(1-\kappa)n}\Phi_n d}\big).
\end{aligned}$$

3.5.4 Proof of Lemma 3.3

Recall the definitions (3.45a) and (3.45b). By Minkowski's inequality, we have the upper bound

$$\Big(\mathbb{E}\big[H_3(t)^p\big]\Big)^{1/p} \leq \big(\mathbb{E}\|N_t\Delta_t\|_2^{2p}\big)^{1/p} + \big(\mathbb{E}\|Z_{t+1}\Delta_t\|_2^{2p}\big)^{1/p} + \big(\mathbb{E}\|\zeta_{t+1}\|_2^{2p}\big)^{1/p} + \big(\mathbb{E}\|\nu_t\|_2^{2p}\big)^{1/p}, \tag{3.65}$$

For the martingale part of the noise, we note that Assumption 3.2 implies that

$$\big(\mathbb{E}\|Z_{t+1}\Delta_t\|_2^{2p} \mid \mathcal{F}_t\big)^{1/p} \leq p^2\sigma_L^2 d\cdot\|\Delta_t\|_2^2, \quad \text{and} \quad \big(\mathbb{E}\|\zeta_{t+1}\|_2^{2p}\big)^{1/p} \leq p^2\bar{\sigma}^2 d.$$

For the additive Markov noise, applying Assumption 3.2 yields $\big(\mathbb{E}\|\nu_t\|_2^{2p}\big)^{1/p} \leq p^2\bar{\sigma}^2 d$.

For the Markov part of the multiplicative noise, we make use of the construction given in Section 3.5.2.1, where we showed that for a given $\tau > 0$, there exists a random variable $\widetilde{s}_t$ such that $\widetilde{s}_t \mid \mathcal{F}_{t-\tau} \sim \xi$, and $\mathbb{E}\big[\rho^p(s_t, \widetilde{s}_t) \mid \mathcal{F}_{t-\tau}\big] \leq c_0 \cdot 2^{1-\frac{\tau}{t_{\mathrm{mix}}}}$. Observe the decomposition

$$N_t\Delta_t = \big(\boldsymbol{L}(s_t) - L(\widetilde{s}_t)\big)\Delta_{t-\tau} + \big(L(\widetilde{s}_t) - \bar{L}\big)\Delta_{t-\tau} + N_t\big(\Delta_t - \Delta_{t-\tau}\big).$$

Using the Lipschitz condition (3.4), we have that

$$\mathbb{E}\big[\|\big(\boldsymbol{L}(s_t) - L(\widetilde{s}_t)\big)\Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau}\big] \leq c_0\cdot 2^{1-\frac{\tau}{t_{\mathrm{mix}}}}\big(\sigma_L d\|\Delta_{t-\tau}\|_2\big)^{2p}.$$

For any $\tau \geq 2pt_{\mathrm{mix}}\log d$, we have the bound

$$\Big(\mathbb{E}\big[\|\big(\boldsymbol{L}(s_t) - L(\widetilde{s}_t)\big)\Delta_{t-\tau}\|_2^{2p}\big]\Big)^{1/p} \leq p^2\sigma_L^2 d\cdot\big(\mathbb{E}\|\Delta_t\|_2^{2p}\big)^{1/p}.$$

By the moment bounds (3.2) on the stationary distribution, we have

$$\mathbb{E}\big[\|\big(\boldsymbol{L}(\widetilde{s}_t) - \bar{L}\big)\Delta_{t-\tau}\|_2^{2p} \mid \mathcal{F}_{t-\tau}\big] \leq \big(2p\sigma_L\sqrt{d}\|\Delta_{t-\tau}\|_2\big)^{2p}.$$

For the last term, we use the Lipschitz condition 3.4 as well as the boundedness condition 3.3 of metric space. In conjunction with the inequality (3.64), for $\tau \geq 2pt_{\mathrm{mix}}\log(c_0 d)$ and stepsize $\eta \leq \frac{1}{48\tau(\sigma_L d+\gamma_{\max})}$, we arrive at the bound

$$\begin{aligned}
&\big(\mathbb{E}\big[\|N_t(\Delta_t - \Delta_{t-\tau})\|_2^{2p}\big]\big)^{1/p} \\
&\leq \sigma_L^2 d^2 \cdot \big(\mathbb{E}\big[\|\Delta_t - \Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} \\
&\leq c\eta^2\sigma_L^2 d^2\tau^2\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} + c\eta^2 p^2\sigma_L^2\bar{\sigma}^2 d^3\tau^2 \\
&\leq c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} + cp^2\bar{\sigma}^2 d,
\end{aligned}$$

for a universal constant $c > 0$.

Collecting the bounds above and substituting into our initial bound (3.65), we find that

$$\big(\mathbb{E}\big[H_3(t)^p\big]\big)^{1/p} \leq c\big(p^2\sigma_L^2 d + \gamma_{\max}^2\big)\big(\mathbb{E}\big[\|\Delta_{t-\tau}\|_2^{2p}\big]\big)^{1/p} + cp^2\bar{\sigma}^2 d,$$

as claimed.

3.6 Proof of Theorem 3.1

From the defining equations (3.3a) and (3.3b), we have the telescoping relation

$$\frac{\theta_n - \theta_{n_0}}{\eta(n-n_0)} = \frac{1}{n-n_0}\sum_{t=n_0}^{n-1}\big(\theta_t - L_{t+1}\theta_t - b_{t+1}\big) = (I-\bar{L})(\widehat{\theta}_n - \bar{\theta}) + \frac{1}{n-n_0}\Psi_{n_0,n} + \frac{1}{n-n_0}\Upsilon_{n_0,n} \tag{3.66}$$

where $\Psi_{n_0,n} = \sum_{t=n_0}^{n-1}\big(L_{t+1}\theta_t + b_{t+1} - \mathbb{E}\big[L_{t+1}\theta_t + b_{t+1}|\mathcal{F}_t\big]\big)$ and $\Upsilon_{n_0,n} := \sum_{t=n_0}^{n-1}\big(\boldsymbol{L}(s_t)\theta_t + \boldsymbol{b}(s_t) - \bar{L}\theta_t - \bar{b}\big)$. Some algebra yields

$$\widehat{\theta}_n - \bar{\theta} = \frac{(I-\bar{L})^{-1}\big(\theta_n - \theta_{n_0}\big)}{\eta(n-n_0)} - \frac{(I-\bar{L})^{-1}\Psi_{n_0,n}}{n-n_0} - \frac{(I-\bar{L})^{-1}\Upsilon_{n_0,n}}{n-n_0} =: I_1 + I_2 + I_3 \tag{3.67}$$

From the triangle inequality, it suffices to bound the norms of I_1, I_2 and I_3.

In the following, we prove a slightly stronger claim, which gives bounds on an arbitrary quadratic loss functional. In particular, given a matrix $Q \succ 0$, we seek bounds on the Q-norm $\|\widehat{\theta}_n - \bar{\theta}\|_Q := \sqrt{(\widehat{\theta}_n - \bar{\theta})^\top Q(\widehat{\theta}_n - \bar{\theta})}$.

3.6.1 Bounding the three terms

We now bound each term in the decomposition (3.67) in turn.

3.6.1.1 Bounding the term I_1

The bound for term I_1 follows directly from Proposition 3.1. In particular, given a sample size $n \geq \frac{8}{\eta(1-\kappa)} \log\big(\|\theta_0 - \bar{\theta}\|_2 d/\eta\big)$ and burn-in period $n_0 = n/2$, we have

$$\mathbb{E}\big[\|\theta_n - \bar{\theta}\|_2^2\big] \leq \frac{c\eta}{1-\kappa}\bar{\sigma}^2\tau d, \quad \text{and} \quad \mathbb{E}\big[\|\theta_{n_0} - \bar{\theta}\|_2^2\big] \leq \frac{c\eta}{1-\kappa}\bar{\sigma}^2\tau d.$$

Noting that $|\!|\!|(I - \bar{L})^{-1}|\!|\!|_{\mathrm{op}} \leq (1-\kappa)^{-1}$, we conclude that

$$\mathbb{E}\big[\|I_1\|_Q^2\big] \leq \lambda_{\max}(Q)\mathbb{E}\big[\|I_1\|_2^2\big] \leq \lambda_{\max}(Q) \cdot \tfrac{c\bar{\sigma}^2\tau d}{\eta(1-\kappa)^3 n^2}. \tag{3.68}$$

3.6.1.2 Bounding the term I_2

For the term I_2, note that the process $(\Psi_t)_{t \geq n_0}$ is a martingale adapted to the natural filtration. Its second moment equals the quadratic variation:

$$\begin{aligned}\mathbb{E}\big[\|I_2\|_Q^2\big] &= \frac{4}{n^2}\mathbb{E}\big[[Q^{1/2}(I-\bar{L})^{-1}\Psi]_{n_0,n}\big] \\ &= \frac{4}{n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|(I-\bar{L})^{-1}\big((L_{t+1} - \boldsymbol{L}(s_t))\theta_t + b_{t+1} - \boldsymbol{b}(s_t)\big)\|_Q^2\big].\end{aligned}$$

By the Cauchy–Schwarz inequality, we have the bound

$$\begin{aligned}\mathbb{E}\big[\|I_2\|_Q^2\big] &\leq \tfrac{8}{n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|(I-\bar{L})^{-1}\zeta_{t+1}\|_Q^2\big] + \tfrac{8}{n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|(I-\bar{L})^{-1}Z_{t+1}\Delta_t\|_Q^2\big] \\ &\leq \tfrac{16}{n}\mathrm{Tr}\big(Q(I-\bar{L})^{-1}{\Sigma^*}_{\mathrm{MG}}(I-\bar{L})^{-\top}\big) + \tfrac{16\sigma_L^2\lambda_{\max}(Q)d}{(1-\kappa)^2 n^2}\sum_{t=n_0}^{n-1}\mathbb{E}\big[\|\Delta_t\|_2^2\big] \\ &\leq \tfrac{16}{n}\mathrm{Tr}\big((I-\bar{L})^{-1}{\Sigma^*}_{\mathrm{MG}}(I-\bar{L})^{-\top}\big) + \lambda_{\max}(Q) \cdot \tfrac{16\sigma_L^2 d}{(1-\kappa)^2 n} \cdot \tfrac{c\eta d\tau}{1-\kappa}\bar{\sigma}^2.\end{aligned} \tag{3.69}$$

3.6.1.3 Bounding the term I_3

Applying the Cauchy-Schwarz inequality yields

$$\mathbb{E}\big[\|(I-\bar{L})^{-1}\Upsilon_{n_0,n}\|_2^2\big] \leq 2\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}\nu_t\|_2^2\big] + 2\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}(I-\bar{L})^{-1}N_t\Delta_t\|_2^2\big]. \tag{3.70}$$

We make use of the two auxiliary lemmas in order to control the terms in the decomposition (3.70).

Lemma 3.7. *Under the setup above, for a sample size n satisfying the bound $\frac{n}{\log n} \geq 2t_{\mathrm{mix}} \log(c_0 d)$, there exists a universal constant $c > 0$ such that*

$$\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} (I - \bar{L})^{-1} \nu_t \|_Q^2\big] \leq (n - n_0) \cdot \mathrm{Tr}\big(Q(I - \bar{L})^{-1} {\Sigma^*}_{\mathrm{Mkv}} (I - \bar{L})^{-\top}\big) \\ + \lambda_{max}(Q) \cdot \frac{c t_{\mathrm{mix}}^2 \bar{\sigma}^2 d}{(1-\kappa)^2} \log^2(c_0 d).$$

See Section 3.6.2.1 for the proof of this claim.

Lemma 3.8. *Under the above conditions, there exists a universal constant $c > 0s$ such that for any scalar $\tau \geq 3t_{\mathrm{mix}} \log^2(c_0 dn)$, stepsize $\eta \in \big(0, \frac{1-\kappa}{c\tau(\sigma_L^2 d + \gamma_{\mathrm{max}}^2)}\big]$ and burn-in time $n_0 \geq \tau + \frac{2}{(1-\kappa)\eta} \log(nd)$, we have $\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} N_t \Delta_t \|_2^2\big] \leq c\eta^2 n^2 \tau^2 d^2 \sigma_L^2 \bar{\sigma}^2$.*

See Section 3.6.2.2 for the proof of this claim.

We now exploit the preceding two lemmas to upper bound the term I_3. We have

$$\mathbb{E}\big[\|I_3\|_Q^2\big] \tag{3.71}$$

$$\leq \tfrac{2}{(n-n_0)^2} \mathbb{E}\big[\| \sum_{t=n_0}^{n-1} (I - \bar{L})^{-1} \nu_t \|_Q^2\big] + \tfrac{2}{(n-n_0)^2} \mathbb{E}\big[\| \sum_{t=n_0}^{n-1} (I - \bar{L})^{-1} N_t \Delta_t \|_Q^2\big] \\ \leq \tfrac{8\mathrm{Tr}\big(Q(I-\bar{L})^{-1} {\Sigma^*}_{\mathrm{Mkv}} (I-\bar{L})^{-\top}\big)}{n} + \lambda_{\mathrm{max}}(Q) \cdot \tfrac{c t_{\mathrm{mix}}^2 \bar{\sigma}^2 d}{(1-\kappa)^2 n^2} \log^2(c_0 d) + \lambda_{\mathrm{max}}(Q) \cdot \tfrac{c\eta^2 \tau^2 d^2 \sigma_L^2 \bar{\sigma}^2}{(1-\kappa)^2}. \tag{3.72}$$

Collecting the bounds (3.68), (3.69), and (3.72), we find that

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_Q^2\big] \leq \tfrac{c}{n} \mathrm{Tr}\big(Q(I - \bar{L})^{-1} ({\Sigma^*}_{\mathrm{MG}} + {\Sigma^*}_{\mathrm{Mkv}})(I - \bar{L})^{-\top}\big) \\ + \lambda_{\mathrm{max}}(Q) \cdot \Big[\frac{c\bar{\sigma}^2 t_{\mathrm{mix}} d}{\eta(1-\kappa)^3 n^2} + \frac{16\sigma_L^2 d}{(1-\kappa)^2 n} \cdot \frac{c\eta d t_{\mathrm{mix}}}{1-\kappa} \bar{\sigma}^2\Big] \\ + \lambda_{\mathrm{max}}(Q) \cdot \Big[\frac{c t_{\mathrm{mix}}^2 \bar{\sigma}^2 d}{(1-\kappa)^2 n^2} \log^2(c_0 dn) + \frac{c\eta^2 t_{\mathrm{mix}}^2 d^2 \sigma_L^2 \bar{\sigma}^2}{(1-\kappa)^2}\Big].$$

For a sample size n lower bounded as $\frac{n}{\log^2 n} \geq \frac{2t_{\mathrm{mix}}(\sigma_L^2 d + \gamma_{\mathrm{max}}^2)}{(1-\kappa)^2} \log(c_0 d)$, we can take the optimal stepsize $\eta = \big[c\big((1-\kappa) n^2 t_{\mathrm{mix}} (\sigma_L^2 d + \gamma_{\mathrm{max}}^2)\big)\big]^{-1/3}$. With this choice, we have

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_Q^2\big] \leq \tfrac{c}{n} \mathrm{Tr}\big(Q(I - \bar{L})^{-1} ({\Sigma^*}_{\mathrm{MG}} + {\Sigma^*}_{\mathrm{Mkv}})(I - \bar{L})^{-\top}\big) \\ + c\lambda_{\mathrm{max}}(Q) \cdot \Big(\frac{\sigma_L^2 d t_{\mathrm{mix}}}{(1-\kappa)^2 n}\Big)^{4/3} \log^2 n. \tag{3.73}$$

Setting $Q := I_d$ completes the proof.

3.6.2 Proof of auxiliary results

In this section, we prove the two auxiliary results used in the proof of Theorem 3.1: namely, Lemma 3.7 and Lemma 3.8.

3.6.2.1 Proof of Lemma 3.7

Given an integer $k \geq 0$, we define the k-step correlation under the stationary Markov chain as

$$\mu_k := \mathbb{E}_{s \sim \xi, s' \sim P^k \delta_s} \big[\langle Q^{1/2}(I - \bar{L})^{-1} \nu(s),\, Q^{1/2}(I - \bar{L})^{-1} \nu(s') \rangle \big].$$

Clearly, we have $\mu_0 \geq 0$, and by Cauchy–Schwarz inequality, for any $k \geq 0$, there is:

$$|\mu_k| \leq \sqrt{\mathbb{E}_{s \sim \xi} \|(I - \bar{L})^{-1} \nu(s)\|_Q^2} \cdot \sqrt{\mathbb{E}_{s' \sim \xi} \|(I - \bar{L})^{-1} \nu(s')\|_Q^2} = \mu_0.$$

The desired quantity can be written as $\mathrm{Tr}\big(Q^{1/2}(I - \bar{L})^{-1} {\Sigma^*}_{\mathrm{Mkv}} (I - \bar{L})^{-\top} Q^{1/2}\big) = \mu_0 + 2 \sum_{k=1}^{+\infty} \mu_k$. Expanding the squared norm yields

$$\begin{aligned}
\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} Q^{1/2}(I - \bar{L})^{-1} \nu_t \|_2^2\big] &= \sum_{n_0 \leq t_1, t_2 \leq n-1} \mathbb{E}\big[\langle Q^{1/2}(I - \bar{L})^{-1} \nu(s_{t_1}),\, Q^{1/2}(I - \bar{L})^{-1} \nu(s_{t_2}) \rangle\big] \\
&= (n - n_0)\mu_0 + 2 \sum_{k=1}^{n-n_0-1} (n - n_0 - k) \mu_k.
\end{aligned}$$

We claim that the cross-correlations μ_k satisfy the bound

$$|\mu_k| \leq c_0 \frac{\bar{\sigma}^2 \|Q\|_{\mathrm{op}} d^2}{(1 - \kappa)^2} \cdot 2^{1 - \frac{k}{2 t_{\mathrm{mix}}}}. \tag{3.74}$$

We return to prove this fact momentarily. Taking it as given, this inequality, in conjunction with the bound $|\mu_k| \leq \mu_0$, can be employed to bound the tail sums needed for the proof. We have

$$\Big| \sum_{k=1}^{n-n_0-1} k \mu_k \Big| \leq \sum_{k=1}^{\tau} \tau |\mu_k| + \sum_{k=\tau+1}^{\infty} k |\mu_k| \leq \tau^2 \mu_0 + 2 c_0 \frac{\bar{\sigma}^2 \|Q\|_{\mathrm{op}} d^2}{(1 - \kappa)^2} \sum_{k=\tau+1}^{\infty} k \cdot 2^{-\frac{k}{2 t_{\mathrm{mix}}}}.$$

With the choice $\tau := 2 t_{\mathrm{mix}} \log(c_0 d)$, simplifying yields

$$\begin{aligned}
\Big| \sum_{k=1}^{n-n_0-1} k \mu_k \Big| &\leq \frac{\tau^2 \bar{\sigma}^2 d \|Q\|_{\mathrm{op}}}{(1 - \kappa)^2} + 2 c_0 \frac{\bar{\sigma}^2 d^2 \|Q\|_{\mathrm{op}}}{(1 - \kappa)^2} \cdot 2 t_{\mathrm{mix}} \big(\tau + 1 + 2 t_{\mathrm{mix}}\big) \cdot 2^{-\frac{\tau+1}{t_{\mathrm{mix}}}} \\
&\leq \frac{2 \tau^2 \bar{\sigma}^2 d}{(1 - \kappa)^2} \|Q\|_{\mathrm{op}},
\end{aligned}$$

and for n satisfying $\frac{n}{\log n} \geq 2\log(c_0 d t_{\mathrm{mix}})$, we have:

$$\sum_{k=n-n_0}^{\infty} |\mu_k| \leq 2c_0 \frac{\bar{\sigma}^2 d^2 \|Q\|_{\mathrm{op}}}{(1-\kappa)^2} \textstyle\sum_{k=\frac{1}{2}n}^{\infty} \cdot 2^{-\frac{k}{2t_{\mathrm{mix}}}} \leq 2c_0 \frac{\bar{\sigma}^2 d^2 \|Q\|_{\mathrm{op}}}{(1-\kappa)^2} \cdot 2^{-\frac{n}{2t_{\mathrm{mix}}}} \leq 2c_0 \frac{\bar{\sigma}^2 d}{(1-\kappa)^2 n^2} \|Q\|_{\mathrm{op}}.$$

Putting together these bounds yields

$$\begin{aligned} \mathbb{E}\big[\| \sum_{t=n_0}^{n-1} (I-\bar{L})^{-1}\nu_t \|_Q^2\big] &= (n-n_0)\big(\mu_0 + 2\sum_{k=1}^{\infty}\mu_k\big) - 2(n-n_0)\sum_{k=n-n_0}^{\infty}\mu_k - 2\sum_{k=1}^{n-n_0-1} k\mu_k \\ &\leq (n-n_0)\cdot \mathrm{Tr}\big((I-\bar{L})^{-1}\Sigma^*{}_{\mathrm{Mkv}}(I-\bar{L})^{-1}\big) + \frac{3\tau^2\bar{\sigma}^2 d}{(1-\kappa)^2}\|Q\|_{\mathrm{op}}, \end{aligned}$$

which completes the proof of the lemma.

Proof of equation (3.74) Let $s_0 \sim \xi$ and $(s_t)_{t\geq 0}$ be a stationary Markov chain starting from s_0. By the construction given in Section 3.5.2.1, there exists a random variable $\widetilde{s}_k$, such that $\widetilde{s}_k$ is independent of s_0, $\widetilde{s}_k \sim \xi$, and such that $\mathbb{E}\big[\rho(s_k, \widetilde{s}_k) \mid s_0\big] \leq c_0 \cdot 2^{1-\frac{k}{t_{\mathrm{mix}}}}$. We then obtain the bound

$$\begin{aligned} |\mu_k| &= \big|\mathbb{E}\big[\langle Q^{1/2}(I-\bar{L})^{-1}\nu(s_0),\, Q^{1/2}(I-\bar{L})^{-1}\nu(s_k)\rangle\big]\big| \\ &\leq \big|\mathbb{E}\big[\langle Q^{1/2}(I-\bar{L})^{-1}\nu(s_0),\, \mathbb{E}\big[Q^{1/2}(I-\bar{L})^{-1}\nu(\widetilde{s}_k) \mid s_0\big]\rangle\big]\big| \\ &\quad + \big|\mathbb{E}\big[Q^{1/2}\langle (I-\bar{L})^{-1}\nu(s_0),\, \mathbb{E}\big[Q^{1/2}(I-\bar{L})^{-1}\big(\nu(s_k)-\nu(\widetilde{s}_k)\big) \mid s_0\big]\rangle\big]\big| \\ &\leq 0 + \sqrt{\mathbb{E}\big[\|Q^{1/2}(I-\bar{L})^{-1}\nu(s_0)\|_2^2\big]} \cdot \sqrt{\mathbb{E}\big[\|Q^{1/2}(I-\bar{L})^{-1}\big(\nu(s_k)-\nu(\widetilde{s}_k)\big)\|_2^2\big]} \\ &\leq \sqrt{\mu_0}\cdot \frac{1}{1-\kappa}\sqrt{\mathbb{E}\big[\rho(s_k,\widetilde{s}_k)^2 \cdot (\sigma_L\|\bar{\theta}\|_2 + \sigma_b)^2 d^2\big]} \\ &\leq c_0 \frac{\bar{\sigma} d}{1-\kappa}\sqrt{\mu_0}\cdot 2^{1-\frac{k}{2t_{\mathrm{mix}}}}. \end{aligned} \tag{3.75}$$

On the other hand, applying the moment condition (3.2) yields $\mu_0 \leq \frac{1}{(1-\kappa)^2} \cdot \mathbb{E}\big[\|\nu(s_0)\|_Q^2\big] \leq \frac{\bar{\sigma}^2 d}{(1-\kappa)^2}\|Q\|_{\mathrm{op}}$. Substituting this bound into our previous inequality (3.75) completes the proof.

3.6.2.2 Proof of Lemma 3.8

The proof of this claim relies on a bootstrap argument: we bound the summation of interest by a more complicated summation that involves product of noise matrices. Recursively applying the result for $m = \log d$ times yields the desired bound.

Lemma 3.9. *Given any integer $m \geq 0$, deterministic sequence $0 = k_0 < k_1 < \cdots < k_m < n_0$, and scalar $\tau \geq 3m t_{\mathrm{mix}} p \log(c_0 dn)$, we have the second moment bound*

$$\begin{aligned}
&\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} \big(\prod_{j=0}^{m} N_{t-k_j} \big) \Delta_{t-k_m} \|_2^2\big] \\
&\leq 2n^2 d^{2m} \sigma_L^{2m+2} \cdot \frac{c\eta}{1-\kappa} d t_{\mathrm{mix}} \bar{\sigma}^2 + 4\eta^2 \tau \sum_{k_{m+1}=k_m+1}^{k_m+\tau} \mathbb{E}\big[\| \sum_{t=n_0}^{n} \big\{ \prod_{j=0}^{m+1} N_{t-k_j} \Delta_{t-k_{m+1}} \big\} \|_2^2\big] \\
&\quad + 4\eta^2 \tau \sum_{k_{m+1}=k_m+1}^{k_m+\tau} \mathbb{E}\big[\| \sum_{t=n_0}^{n} \big\{ \prod_{j=0}^{m} N_{t-k_j} \big(\nu_{t-k_{m+1}} + \zeta_{t-k_{m+1}+1} \big) \big\} \|_2^2\big],
\end{aligned} \tag{3.76a}$$

and in the special case $m = 0$, we have

$$\begin{aligned}
&\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} N_t \Delta_t \|_2^2\big] \\
&\leq c\sigma_L^2 d \cdot \big(n\tau + n^2 \eta^2 \sigma_L^2 d \tau^2 \big) \frac{c\eta}{1-\kappa} d t_{\mathrm{mix}} \bar{\sigma}^2 + 4\eta^2 \tau \sum_{k_1=1}^{\tau} \mathbb{E}\big[\| \sum_{t=n_0}^{n} N_t N_{t-k_1} \Delta_{t-k_1} \|_2^2\big] \\
&\quad + 4\eta^2 \tau \sum_{k_1=1}^{\tau} \mathbb{E}\big[\| \sum_{t=n_0}^{n} N_t \big(\nu_{t-k_1} + \zeta_{t-k_1+1} \big) \|_2^2\big].
\end{aligned} \tag{3.76b}$$

See Appendix B.4.1 for the proof of this lemma.

The following lemma controls the last term of the bound (3.76a):

Lemma 3.10. *Under the setup above, there exists a universal constant $c > 0$, such that for any integer $m > 0$ and deterministic sequence $0 = k_0 < k_1 < \cdots < k_m < n_0$, we have:*

$$\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} \big(\prod_{j=0}^{m-1} N_{t-k_j} \big) \big(\nu_{t-k_m} + \zeta_{t-k_m+1} \big) \|_2^2\big] \leq c\big(n^2 + nd(k_m + t_{\mathrm{mix}} \log(c_0 d)) \big) \sigma_L^{2m} d^{2m} \bar{\sigma}^2.$$

See Appendix B.4.2 for the proof of this lemma.

Taking these lemmas as given, we now proceed with the proof of Lemma 3.8. Given the scalar $\tau := 3 t_{\mathrm{mix}} \log^2(c_0 dn)$, we define

$$\mathfrak{H}_m := \sup_{0=k_0<k_1<\cdots<k_m\leq\tau} \mathbb{E}\big[\| \sum_{t=n_0}^{n-1} \big(\prod_{j=0}^{m} N_{t-k_j} \big) \Delta_{t-k_m} \|_2^2\big]$$

for $m = 0, 1, 2, \cdots, \log d$. By equation (3.76b) and Lemma 3.10, we have the bound

$$\begin{aligned}
\mathfrak{H}_0 &\leq c\sigma_L^2 d \cdot \big(n\tau + n^2 \eta^2 \sigma_L^2 d \tau^2 \big) \frac{c\eta}{1-\kappa} d t_{\mathrm{mix}} \bar{\sigma}^2 + 4\eta^2 \tau^2 \mathfrak{H}_1 \\
&\quad + 4c\eta^2 \tau^2 \big(n^2 + nd(\tau + t_{\mathrm{mix}} \log(c_0 d)) \big) \sigma_L^2 d^2 \bar{\sigma}^2 \\
&\leq 4\eta^2 \tau^2 \mathfrak{H}_1 + c' \eta^2 n^2 \tau^2 d^2 \sigma_L^2 \bar{\sigma}^2.
\end{aligned}$$

In deriving the last inequality, we used the inequalities $\eta \leq \frac{1-\kappa}{\sigma_L^2 d\tau}$ and $n \geq \frac{1}{(1-\kappa)\eta}$.

By equation (3.76a) and Lemma 3.10, we have the recursive relation

$$\begin{aligned} \mathfrak{H}_m &\leq 4\eta^2\tau^2\mathfrak{H}_{m+1} + cn^2 d^{2m+1}\tau\sigma_L^{2m+2} \cdot \frac{\eta \log^3 n}{1-\kappa}\bar{\sigma}^2 + c\eta^2\tau^2 n^2 \sigma_L^{2m+2} d^{2m+2}\bar{\sigma}^2 \\ &\leq 4\eta^2\tau^2\mathfrak{H}_{m+1} + cn^2\sigma_L^{2m} d^{2m}\bar{\sigma}^2 \cdot \log^3 n. \end{aligned}$$

Recursively applying these bounds yields

$$\begin{aligned} \mathfrak{H}_0 &\leq (4\eta^2\tau^2)^m \mathfrak{H}_m + c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2 + c \cdot \sum_{q=1}^{m-1} (4\eta^2\tau^2)^q n^2 \sigma_L^{2q} d^{2q}\bar{\sigma}^2 \\ &\leq (4\eta^2\tau^2)^m \mathfrak{H}_m + 3c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2. \end{aligned}$$

In order to control the term $\mathfrak{H}_m$, we employ the coarse bound

$$\begin{aligned} \mathbb{E}\Big[\Big\| \sum_{t=n_0}^{n-1} \Big(\prod_{j=0}^{m} N_{t-k_j}\Big)\Delta_{t-k_m}\Big\|_2^2\Big] &\leq n \sum_{t=n_0}^{n-1} \mathbb{E}\Big[\Big\|\Big(\prod_{j=0}^{m} N_{t-k_j}\Big)\Delta_{t-k_m}\Big\|_2^2\Big] \\ &\leq n^2(\sigma_L d)^{2m+2} \cdot \frac{c\eta t_{\mathrm{mix}} d\bar{\sigma}^2}{1-\kappa}. \end{aligned}$$

Taking the supremum and noting that $\eta \leq \frac{1-\kappa}{\sigma_L^2 d\tau}$ leads to $\mathfrak{H}_m \leq cn^2\sigma_L^{2m} d^{2m+2}\bar{\sigma}^2$. Consequently, we have established that $\mathfrak{H}_0 \leq 3c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2\big[1 + \big(2\eta\tau\sigma_L d\big)^{\frac{2m+2}{2m}}\big]$. Taking $m = \lceil \log d \rceil$ and $\eta \leq \frac{1}{6\tau\sigma_L d}$, we have $(2\eta\tau\sigma_L d^{\frac{2m+2}{2m}})^{2m} < 1$, and thus $\mathfrak{H}_0 \leq 6c\eta^2 n^2\tau^2 d^2\sigma_L^2\bar{\sigma}^2 \log^3 n$, which completes the proof of this lemma.

3.7 Discussion

In this chapter, we established sharp instance-optimal guarantees for linear stochastic approximation (SA) procedures based on Markovian data. Under ergodicity along with natural tail conditions, we proved non-asymptotic upper bounds on the squared error of both the last iterate of a standard SA scheme, as well as the Polyak–Ruppert averaged sequence. The results highlight two important aspects: an optimal sample complexity of $O(t_{\mathrm{mix}} d)$ for problems in dimension d with mixing time t_{mix}; and an instance-dependent error upper bound for the averaged estimator with carefully chosen stepsize. Complementary to the upper bound, we also showed a non-asymptotic local minimax lower bound over a small neighborhood of a given Markov chain instance, certifying the statistical optimality of the proposed estimators. Our proof of the upper bounds uses a bootstrapping argument of possibly independent interest.

Throughout the chapter, we have introduced novel techniques of analysis and motivated several open questions. In the following, we collect a few interesting future directions:

- **Nonlinear stochastic approximation and controlled dynamics:** This chapter focuses on linear Z-equations where the underlying Markov chain does not involve a control. Though this setting already covers many important examples (as described in Section 3.2.2), its applicability to practical problems is still relatively restricted. To set up a general framework, one could consider a *controlled Markov chain* $(s_t)_{t \geq 0}$ where the transition is given by $s_{t+1} \sim P(\cdot | s_t, \theta_t)$. For any $\theta \in \mathbb{R}^d$, let ξ_θ be the stationary distribution of the Markov chain $P(\cdot|\cdot, \theta)$ induced by the control θ. Given a non-linear operator $H : \mathcal{S} \times \mathbb{R}^d \to \mathbb{R}^d$, suppose that we wish to solve the equation $\mathbb{E}_{s \sim \xi(\theta)} \big[H(\theta; s) \big] = 0$; see [11] for a summary of classical asymptotic theory for such problems. The analysis tools introduced in this chapter provide an avenue by which one could obtain optimal sample complexity bounds (especially in terms of dimension dependency) and instance-dependent guarantees for such problems.

- **Online statistical inference:** By carefully choosing the burn-in period, one can show that the Polyak–Ruppert estimator $\widehat{\theta}_n$ is asymptotically normal and locally minimax optimal. In particular, under suitable conditions, the following limiting result holds true (see [61] for details):

$$\sqrt{n}(\widehat{\theta}_n - \bar{\theta}) \xrightarrow{d} \mathcal{N}\big((I_d - \bar{L})^{-1} ({\Sigma^*}_{\mathrm{MG}} + {\Sigma^*}_{\mathrm{Mkv}}) (I_d - \bar{L})^{-\top} \big). \tag{3.77}$$

  In order to construct confidence intervals for the solution $\bar{\theta}$ with streaming data, it suffices to estimate the asymptotic covariance in equation (3.77). In the i.i.d. setting, online procedures have been developed to estimate such covariances, with non-asymptotic error guarantees [39]. The problem becomes more subtle in the Markovian setting, as the matrix ${\Sigma^*}_{\mathrm{Mkv}}$ involves auto-correlations of the noise process. It is an important open direction to construct online estimators of this matrix to enable inference in a streaming fashion.

- **Model selection and optimal methods for policy evaluation** The policy evaluation problem involves manual choice of two important parameters: the feature vector dimension d and the resolvent parameter λ in TD(λ). In Section 3.4.1.3 and 3.4.2, we provide optimal instance-dependent guarantees on both the approximation factor and the estimation error, for a fixed choice of d and λ. An important direction of future research is to select such parameters adaptively based on data, possibly under a streaming computational model. Ideally, we want the risk of such estimator to attain the infimum of the right hand side of equation (3.39b), over $\lambda \in (0, 1)$ and $d \in \mathbb{N}_+$. A possible candidate approach towards such a model selection problem is the celebrated Lepskii method for adaptive bandwidth selection [118].

.

Part II

Off-policy estimation of linear functionals

Chapter 4

A non-asymptotic theory for semi-parametric efficiency

The problem of estimating a linear functional based on observational data is canonical in both the causal inference and bandit literatures. We analyze a broad class of two-stage procedures that first estimate the treatment effect function, and then use this quantity to estimate the linear functional. We prove non-asymptotic upper bounds on the mean-squared error of such procedures: these bounds reveal that in order to obtain non-asymptotically optimal procedures, the error in estimating the treatment effect should be minimized in a certain weighted L^2-norm. We analyze a two-stage procedure based on constrained regression in this weighted norm, and establish its instance-dependent optimality in finite samples via matching non-asymptotic local minimax lower bounds. These results show that the optimal non-asymptotic risk, in addition to depending on the asymptotically efficient variance, depends on the weighted norm distance between the true outcome function and its approximation by the richest function class supported by the sample size.

4.1 Introduction

A central challenge in both the casual inference and bandit literatures is how to estimate a linear functional associated with the treatment (or reward) function, along with inferential issues associated with such estimators. Of particular interest in causal inference are average treatment effects (ATE) and weighted variants thereof, whereas with bandits and reinforcement learning, one is interested in various linear functionals of the reward function (including elements of the value function for a given policy). In many applications, the statistician has access to only observational data, and lacks the ability to sample the treatment or the actions according to the desired probability distribution. By now, there is a rich body of work on this problem (e.g., [178, 177, 41, 4, 215, 133]), including various types of estimators that are equipped with both asymptotic and non-asymptotic guarantees. We overview this and other past work in the related work section to follow.

In this chapter, we study how to estimate an arbitrary linear functional based on observational data. When formulated in the language of contextual bandits, each such problem involves a state space $\mathbb{X}$, an action space $\mathbb{A}$, and an output space $\mathbb{Y} \subseteq \mathbb{R}$. Given a base measure λ on the action space $\mathbb{A}$—typically, the counting measure for discrete action spaces, or Lebesgue measure for continuous action spaces—we equip each $x \in \mathbb{X}$ with a probability density function $\pi(x, \cdot)$ with respect to λ. This combination defines a probability distribution over $\mathbb{A}$, known either as the *propensity score* (in causal inference) or the *behavioral policy* (in the bandit literature). The conditional mean of any outcome $Y \in \mathbb{Y}$ is specified as $\mathbb{E}[Y \mid x, a] = \mu^*(x, a)$, where the function μ^* is known as the *treatment effect* or the *reward function*, again in the causal inference and bandit literatures, respectively.

Given some probability distribution ξ^* over the state space $\mathbb{X}$, suppose that we observe n i.i.d. triples (X_i, A_i, Y_i) in which $X_i \sim \xi^*$, and

$$A_i \mid X_i \sim \pi(X_i, \cdot), \quad \text{and} \quad \mathbb{E}\big[Y_i \mid X_i, A_i\big] = \mu^*(X_i, A_i), \qquad \text{for } i = 1, 2, \ldots, n. \tag{4.1}$$

We also make use of the conditional variance function

$$\sigma^2(x, a) := \mathbb{E}\Big[\big(Y - \mu^*(X, A)\big)^2 \mid X = x, A = a\Big], \tag{4.2}$$

which is assumed to exist for any $x \in \mathbb{X}$ and $a \in \mathbb{A}$.

For a pre-specified weight function $g : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$, our goal is to estimate the linear functional

$$\tau^* \equiv \tau(\mathcal{I}^*) := \int_{\mathbb{A}} \mathbb{E}_{\xi^*}\Big[g(X, a) \cdot \mu^*(X, a)\Big] d\lambda(a), \tag{4.3}$$

With this set-up, the pair $\mathcal{I}^* := (\xi^*, \mu^*)$ defines a particular *problem instance.* Throughout the chapter, we focus on the case where both the propensity score π and the weight function g are known to the statistician.

Among the interesting instantiations of this general framework are the following:

- **Average treatment effect:** The ATE problem corresponds to estimating the linear functional

  $$\tau^* = \mathbb{E}_{\xi^*}\Big[\mu^*(X, 1) - \mu^*(X, 0)\Big].$$

  It is a special case of equation (4.3), obtained by taking the binary action space $\mathbb{A} = \{0, 1\}$ with λ being the counting measure, along with the weight function $g(x, a) := 2a - 1$.

- **Weighted average treatment effect:** Again with binary actions, suppose that we adopt the weight function $g(x, a) := (2a - 1) \cdot w(x)$, for some given function $w : \mathbb{X} \to \mathbb{R}_+$. With the choice $w(x) := \pi(x, 1)$, this corresponds to average treatment effect on the treated (ATET).

- **Off-policy evaluation for contextual bandits**: For a general finite action space $\mathbb{A}$, a *target policy* is a mapping $x \mapsto \pi^{tar}(x, \cdot)$, corresponding to a probability distribution over the action space. If we take the weight function $g(x, a) := \pi^{tar}(x, a)$ and interpret μ^* as a reward function, then the linear functional (4.3) corresponds to the value of the target policy π^{tar}. Since the observed actions are sampled according to π—which can be different than the target policy π^{tar}—this problem is known as *off-policy* evaluation in the bandit and reinforcement learning literature.

When the propensity score is known, it is a standard fact that one can estimate $\tau(\mathcal{I})$ at a $\sqrt{n}$-rate via an importance-reweighted plug-in estimator. In particular, under mild conditions, the *inverse propensity weighting* (IPW) estimator, given by

$$\widehat{\tau}_n^{IPW} := \frac{1}{n} \sum_{i=1}^{n} \frac{g(X_i, A_i)}{\pi(X_i, A_i)} Y_i, \tag{4.4}$$

is $\sqrt{n}$-consistent, in the sense that $\widehat{\tau}_n^{IPW} - \tau^* = \mathcal{O}_p(1/\sqrt{n})$.

However, the problem is more subtle than might appear at might first: the IPW estimator $\widehat{\tau}_n^{IPW}$ fails to be asymptotically efficient, meaning that its asymptotic variance is larger than the optimal one. This deficiency arises even when the state space $\mathbb{X}$ and action space $\mathbb{A}$ are both binary; for instance, see §3 in Hirano et al. [76]. Estimators that are asymptotically efficient can be obtained by first estimating the treatment effect μ^*, and then using this quantity to form an estimate of $\tau(\mathcal{I})$. Such a combination leads to a semi-parametric method, in which μ^* plays the role of a nuisance function. For example, in application to the ATE problem, Chernozhukov et al. [41] showed that any consistent estimator of μ^* yields an asymptotically efficient estimate of $\tau_g(\mathcal{I})$; see §5.1 in their paper. In the sequel, so as to motivate the procedures analyzed in this chapter, we discuss a broad range of semi-parametric methods that are asymptotically efficient for estimating the linear functional $\tau(\mathcal{I})$.

While such semi-parametric procedures have attractive asymptotic guarantees, they are necessarily applied in finite samples, in which context a number of questions remain open:

- As noted above, we now have a wide array of estimators that are known to be asymptotically efficient, and are thus "equivalent" from the asymptotic perspective. It is not clear, however, which estimator(s) should be used when working with a finite collection of samples, as one always does in practice. Can we develop theory that provides more refined guidance on the choice of estimators in this regime?

- As opposed to a purely parametric estimator (such as the IPW estimate), semi-parametric procedures involve estimating the treatment effect function μ^*. Such non-parametric estimation requires sample sizes that scale non-trivially with the problem dimension, and induce trade-offs between the estimation and approximation error. In what norm should we measure the approximation/estimation trade-offs associated with estimating the treatment effect? Can we relate this trade-off to non-asymptotic and instance-dependent lower bounds on the difficulty of estimating the linear functional τ?

The main goal of this chapter is to give some precise answers to these questions. On the lower bound side, we establish instance-dependent minimax lower bounds on the difficulty of estimating τ. These lower bounds show an interesting elbow effect, in that if the sample size is overly small relative to the complexity of a function class associated with the treatment effect, then there is a penalty in addition to the classical efficient variance. On the upper bound side, we propose a class of weighted constrained least-square estimators that achieve optimal non-asymptotic risk, even in the high-order terms. Both the upper and lower bounds are general, with more concrete consequences for the specific instantiations introduced previously.

Related work: Let us provide a more detailed overview of related work in the areas of semi-parametric estimation and more specifically, the literatures on the treatment effect problem as well as related bandit problems.

In this chapter, we make use of the notion of local minimax lower bounds which, in its asymptotic instantiation, dates back to seminal work of Le Cam [117] and Hájek [72]. These information-based methods were extended to semiparametric settings by Stein [196] and Levit [119, 120], among other authors. Under appropriate regularity assumptions, the optimal efficiency is determined by the worst-case Fisher information of regular parametric sub-models in the tangent space; see the monograph [17] for a comprehensive review.

Early studies of treatment effect estimation were primarily empirical [6]. The unconfoundedness assumption was first formalized by Rosenbaum and Rubin [182], thereby leading to the problem setup described in Section 4.1. A series of seminal papers by Robins and Rotnitzky [178, 177] made connections with the semi-parametric literature; the first semi-parametric efficiency bound, using the tangent-based techniques described in the monograph [17], was formally derived by Hahn [71].

There is now a rich body of work focused on constructing valid inference procedures under various settings, achieving such semiparametric lower bounds. A range of methods have been studied, among them matching procedures [184, 1], inverse propensity weighting [71, 76, 79, 214], outcome regression [38, 78], and doubly robust methods [177, 41, 135, 62]. The two-stage procedure analyzed in the current chapter belongs to the broad category of doubly robust methods.

In their classic paper, Robins and Ritov [176] showed that if no smoothness assumptions are imposed on the outcome model, then the asymptotic variance of the IPW estimator cannot be beaten. This finding can be understood as a worst-case asymptotic statement; in contrast, this chapter takes an instance-dependent perspective, so that any additional structure can be leveraged to obtain superior procedures. Robins et al. [179] derived optimal rates for treatment effect estimation under various smoothness conditions for the outcome function and propensity score function. More recent work has extended this general approach to analyze estimators for other variants of treatment effect (e.g., [96, 4]). There are some connections between our proof techniques and the analysis in this line of work, but our focus is on finite-sample and instance-dependent results, as opposed to global minimax results.

Portions of our work apply to high-dimensional settings, of which sparse linear models are one instantiation. For this class of problems, the recent papers [25, 26, 214] study the relation between sample size, dimension and sparsity level for which $\sqrt{n}$-consistency can be obtained. This body of work applies to the case of unknown propensity scores, which is complementary to our studies with known behavioral policies. To be clear, obtaining $\sqrt{n}$-consistency is always possible under our set-up via the IPW estimator; thus, our focus is on the more refined question of non-asymptotic sample size needed to obtain optimal instance-dependent bounds.

Our work is also related to the notion of second-order efficiency in classical asymptotics. Some past work [43, 43, 34] has studied some canonical semi-parametric problems, including estimating the shift or period of one-dimensional regression functions, and established second-order efficiency asymptotic upper and lower bounds in the exact asymptotics framework. Our instance-dependent lower bounds do not lead to sharp constant factors, but do hold in finite samples. We view it as an important direction for future work to combine exact asymptotic theory with our finite-sample approach so as to obtain second-order efficiency lower bounds with exact first-order asymptotics.

There is also an independent and parallel line of research on the equivalent problem of off-policy evaluation (OPE) in bandits and reinforcement learning. For multi-arm bandits, the paper [123] established the global minimax optimality of certain OPE estimators given a sufficiently large sample size. Wang et al. [215] proposed the "switch" estimator, which switches between importance sampling and regression estimators; this type of procedure, with a particular switching rule, was later shown to be globally minimax optimal for any sample size [133]. Despite desirable properties in a worst-case sense, these estimators are known to be asymptotically inefficient, and the sub-optimality is present even ignoring constant factors (see Section 3 of the paper [76] for some relevant discussion). In the more general setting of reinforcement learning, various efficient off-policy evaluation procedures have been proposed and studied [83, 225, 227, 90]. Other researchers [232, 229, 7, 227] have studied procedures that are applicable to adaptively collected data. It is an interesting open question to see how the perspective of this chapter can be extended to dynamic settings of this type.

Additional notation: Here we collect some notation used throughout the chapter. Given a pair of functions $h_1, h_2 : \mathbb{A} \to \mathbb{R}$ such that $|h_1 h_2| \in \mathbb{L}^1(\lambda)$, we define the inner product

$$\langle h_1,\, h_2 \rangle_\lambda := \int h_1(a) h_2(a) \; d\lambda(a)$$

4.2 Non-asymptotic and instance-dependent upper bounds

We begin with a non-asymptotic analysis of a general class of two-stage estimators of the functional $\tau(\mathcal{I})$. Our upper bounds involve a certain weighted L^2-norm—see equation (4.8a)—which, as shown by our lower bounds in the sequel, plays a fundamental role.

4.2.1 Non-asymptotic risk bounds on two-stage procedures

We first provide some intuition for the class of two-stage estimators that we analyze, before turning to a precise description.

4.2.1.1 Some elementary intuition

We consider two-stage estimators obtained from simple perturbations of the IPW estimator (4.4). Given an auxiliary function $f : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$ and the data set $\{(X_i, A_i, Y_i)\}_{i=1}^n$, consider the estimate

$$\widehat{\tau}_n^f = \frac{1}{n} \sum_{i=1}^n \Big\{ \frac{g(X_i, A_i)}{\pi(X_i, A_i)} Y_i - f(X_i, A_i) + \langle f(X_i, \cdot),\, \pi(X_i, \cdot) \rangle_\lambda \Big\}. \tag{4.5}$$

By construction, for any choice of $f \in \mathbb{L}^2(\xi^* \times \pi)$, the quantity $\widehat{\tau}_n^f$ is an unbiased estimate of τ, so that it is natural to choose f so as to minimize the variance $\mathrm{var}(\widehat{\tau}_n^f)$ of the induced estimator. As shown in Appendix C.1.1, the minimum of this variational problem is achieved by the function

$$f^*(x, a) := \frac{g(x, a) \mu^*(x, a)}{\pi(x, a)} - \langle g(x, \cdot),\, \mu^*(x, \cdot) \rangle_\lambda, \tag{4.6a}$$

where in performing the minimization, we enforced the constraints $\langle f(x, \cdot),\, \pi(x, \cdot) \rangle_\lambda = 0$ for any $x \in \mathbb{X}$. We note that this same function f^* also arises naturally via consideration of Neyman orthogonality.

The key property of the optimizing function f^* is that it induces an estimator $\widehat{\tau}_n^{f^*}$ with asymptotically optimal variance—viz.

$$v_{\text{semi}}^2 := \mathrm{var}\Big(\langle g(X\cdot),\, \mu^*(X,\cdot)\rangle_\lambda\Big) + \int_{\mathbb{A}} \mathbb{E}_{\xi^*}\Big[\frac{g^2(X,a)}{\pi(X,a)}\sigma^2(X,a)\Big] d\lambda(a), \tag{4.6b}$$

where $\sigma^2(x,a) := \mathrm{var}(Y \mid x,a)$ is the conditional variance (4.2) of the outcome. See Appendix C.1.1 for details of this derivation.

4.2.1.2 A class of two-stage procedures

The preceding set-up naturally leads to a broad class of two-stage procedures, which we define and analyze here. Since the treatment effect μ^* is unknown, the optimal function f^* from equation (4.6a) is also unknown to us. A natural approach, then, is the two-stage one: (a) compute an estimate $\widehat{\mu}$ using part of the data; and then (b) substitute this estimate in equation (4.6a) so as to construct an approximation to the ideal estimator $\widehat{\tau}_n^{f^*}$. A standard cross-fitting approach (e.g., [41]) allows one to make full use of data while avoiding the self-correlation bias.

In more detail, we first split the data into two disjoint subsets $\mathcal{B}_1 := (X_i, A_i, Y_i)_{i=1}^{n/2}$ and $\mathcal{B}_2 := (X_i, A_i, Y_i)_{i=n/2+1}^{n}$. We then perform the following two steps:

Step I: For $j \in \{1,2\}$, compute an estimate $\widehat{\mu}_{n/2}^{(j)}$ of μ^* using the data subset $\mathcal{B}_j$, and compute

$$\widehat{f}_{n/2}^{(j)}(x,a) := \frac{g(x,a)\widehat{\mu}_{n/2}^{(j)}(x,a)}{\pi(x,a)} - \langle g(x,\cdot),\, \widehat{\mu}_{n/2}^{(j)}(x,\cdot)\rangle_\lambda. \tag{4.7a}$$

Step II: Use the auxiliary functions $\widehat{f}_{n/2}^{(1)}$ and $\widehat{f}_{n/2}^{(2)}$ to construct the estimate

$$\widehat{\tau}_n := \frac{1}{n}\sum_{i=1}^{n/2}\Big\{\frac{g(X_i,A_i)}{\pi(X_i,A_i)}Y_i - \widehat{f}_{n/2}^{(2)}(X_i,A_i)\Big\} + \frac{1}{n}\sum_{i=n/2+1}^{n}\Big\{\frac{g(X_i,A_i)}{\pi(X_i,A_i)}Y_i - \widehat{f}_{n/2}^{(1)}(X_i,A_i)\Big\}. \tag{4.7b}$$

As described, these two steps should be understood as defining a meta-procedure, since the choice of auxiliary estimator $\widehat{\mu}_{n/2}^{(j)}$ can be arbitrary.

The main result of this section is a non-asymptotic upper bound on the MSE of any such two-stage estimator. It involves the *weighted L^2-norm* $\|\cdot\|_\omega$ given by

$$\|h\|_\omega^2 := \int_{\mathbb{A}} \mathbb{E}_{\xi^*}\Big[\frac{g^2(X,a)}{\pi(X,a)}h^2(X,a)\Big] d\lambda(a), \tag{4.8a}$$

which plays a fundamental role in both upper and lower bounds for the problem. With this notation, we have:

Theorem 4.1. *For any estimator $\widehat{\mu}_{n/2}$ of the treatment effect, the two-stage estimator* (4.7) *has MSE bounded as*

$$\mathbb{E}\Big[|\widehat{\tau}_n - \tau^*|^2 \Big] \leq \frac{1}{n}\Big\{ v_{\text{semi}}^2 + 2\mathbb{E}\big[\|\widehat{\mu}_{n/2} - \mu^*\|_\omega^2\big] \Big\}. \tag{4.8b}$$

See Section 4.4.1 for the proof of this claim.

Note that the upper bound (4.8b) consists of two terms, both of which have natural intepretations. The first term v_{semi}^2 corresponds to the asymptotically efficient variance (4.6b); in terms of the weighted norm (4.8a), it has the equivalent expression

$$v_{\text{semi}}^2 = \text{var}\Big(\langle g(X\cdot),\, \mu^*(X, \cdot)\rangle_\lambda \Big) + \|\sigma\|_\omega^2. \tag{4.8c}$$

The second term corresponds to twice the average estimation error $\mathbb{E}\big[\|\widehat{\mu}_{n/2} - \mu^*\|_\omega^2\big]$, again measured in the weighted squared norm (4.8a). Whenever the treatment effect can be estimated consistently—so that this second term is vanishing in n—we see that the estimator $\widehat{\tau}_n$ is asymptotically efficient, as is known from past work [41]. Of primary interest to us is the guidance provided by the bound (4.8b) in the finite sample regime: in particular, in order to minimize this upper bound, one should construct estimators $\widehat{\mu}$ of the treatment effect that are optimal in the weighted norm (4.8a).

4.2.2 Some non-asymptotic analysis

With this general result in hand, we now propose some explicit two-stage procedures that can be shown to be finite-sample optimal. We begin by introducing the classical idea of an oracle inequality, and making note of its consequences when combined with Theorem 4.1. We then analyze a class of non-parametric weighted least-squares estimators, and prove that they satisfy an oracle inequality of the desired type.

4.2.2.1 Oracle inequalities and finite-sample bounds

At a high level, Theorem 4.1 reduces our problem to an instance of non-parametric regression, albeit one involving the weighted norm $\|\cdot\|_\omega$ from equation (4.8a). In non-parametric regression, there are many methods known to satisfy an attractive "oracle" property (e.g., see the books [206, 213]). In particular, suppose that we construct an estimate $\widehat{\mu}$ that takes values in some function class $\mathcal{F}$. It is said to satisfy an *oracle inequality* for estimating μ^* in the norm $\|\cdot\|_\omega$ if

$$\mathbb{E}\big[\|\widehat{\mu} - \mu^*\|_\omega^2\big] \leq c \inf_{\mu \in \mathcal{F}} \Big\{ \|\mu - \mu^*\|_\omega^2 + \delta_n^2(\mu; \mathcal{F}) \Big\} \tag{4.9}$$

for some universal constant $c \geq 1$. Here the functional $\mu \mapsto \delta_n^2(\mu; \mathcal{F})$ quantifies the $\|\cdot\|_\omega^2$-error associated with estimating some function $\mu \in \mathcal{F}$, whereas the quantity $\|\mu - \mu^*\|_\omega^2$ is the squared approximation error, since the true function μ^* need not belong to the class. We note that the oracle inequality stated here is somewhat more refined than the standard one, since we have allowed the estimation error to be instance-dependent (via its dependence on the choice of μ).

Given an estimator $\widehat{\mu}$ that satisfies such an oracle inequality, an immediate consequence of Theorem 4.1 is that the associated two-stage estimator of $\tau^* \equiv \tau(\mathcal{I})$ has MSE upper bounded as

$$\mathbb{E}\big[\, |\widehat{\tau}_n - \tau^*|^2 \,\big] \leq \frac{1}{n}\Big(v_{\mathrm{semi}}^2 + 2c \inf_{\mu \in \mathcal{F}} \Big\{ \|\mu - \mu^*\|_\omega^2 + \delta_n^2(\mu; \mathcal{F}) \Big\} \Big). \tag{4.10}$$

This upper bound is explicit, and given some assumptions on the approximability of the unknown μ^*, we can use to it choose the "complexity" of the function class $\mathcal{F}$ in a data-dependent manner. See Section 4.2.4 for discussion and illustration of such choices for different function classes.

4.2.2.2 Oracle inequalities for non-parametric weighted least-squares

Based on the preceding discussion, we now turn to the task of proposing a suitable estimator of μ^*, and proving that it satisfies the requisite oracle inequality (4.9). Let $\mathcal{F}$ be a given function class used to approximate the treatment effect μ^*. Given our goal of establishing bounds in the weighted norm (4.8a), it is natural to analyze the *non-parametric weighted least-squares* estimate

$$\widehat{\mu}_m := \arg\min_{\mu \in \mathcal{F}} \Big\{ \frac{1}{m} \sum_{i=1}^m \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \big\{ \mu(X_i, A_i) - Y_i \big\}^2 \Big\}, \tag{4.11}$$

where $\{(X_i, A_i, Y_i)\}_{i=1}^m$ constitute an observed collection of state-action-outcome triples.

Since the pairs (X, A) are drawn from the distribution $\xi^*(x)\pi(x, a)$, our choice of weights ensures that

$$\mathbb{E}\Big[\frac{g^2(X, A)}{\pi^2(X, A)} \big\{ \mu(X, A) - Y \big\}^2 \Big] = \|\mu - \mu^*\|_\omega^2 + \mathbb{E}\Big[\frac{g^2(X, A)}{\pi^2(X, A)} \sigma^2(X, A) \Big].$$

so that (up to a constant offset), we are minimizing an unbiased estimate of $\|\mu - \mu^*\|_\omega^2$. In our analysis, we impose some natural conditions on the function class:

(CC) The function class $\mathcal{F}$ is a *convex and compact* subset of the Hilbert space $\mathbb{L}_\omega^2$.

We also require some tail conditions on functions h that belong to the difference set

$$\partial \mathcal{F} := \{ f_1 - f_2 \mid f_1, f_2 \in \mathcal{F} \}.$$

There are various results in the non-parametric literature that rely on functions being uniformly bounded, or satisfying other sub-Gaussian or sub-exponential tail conditions (e.g., [213]). Here we instead leverage the less restrictive learning-without-concentration framework of Mendelson [138], and require that the following *small probability condition* holds:

(SB) There exists a pair (α_1, α_2) of positive scalars such that

$$\mathbb{P}\Big[\Big|\tfrac{g(X,A)}{\pi(X,A)} h(X,A)\Big| \geq \alpha_1 \|h\|_\omega\Big] \geq \alpha_2 \qquad \text{for all } h \in \partial\mathcal{F}. \tag{4.12}$$

If we introduce the shorthand $\tilde{h} = \frac{g}{\pi} h$, then condition (4.12) can be written equivalently as $\mathbb{P}\big[|\tilde{h}(X,A)| \geq \alpha_1 \|\tilde{h}\|_2\big] \geq \alpha_2$, so that it is a standard small-ball condition on the function $\tilde{h}$; see the papers [106, 138] for more background.

As with existing theory on non-parametric estimation, our risk bounds are determined by the suprema of empirical processes, with "localization" so as to obtain optimal rates. Given a function class $\mathcal{H}$ and a positive integer m, we define the *Rademacher complexities*

$$\mathcal{S}_m^2(\mathcal{H}) := \mathbb{E}\Big[\sup_{f \in \mathcal{H}} \Big\{\frac{1}{m}\sum_{i=1}^m \frac{\varepsilon_i g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \big(Y_i - \mu^*(X_i, A_i)\big) f(X_i, A_i)\Big\}^2\Big], \quad \text{and} \tag{4.13a}$$

$$\mathcal{R}_m(\mathcal{H}) := \mathbb{E}\Big[\sup_{f \in \mathcal{H}} \frac{1}{m}\sum_{i=1}^m \frac{\varepsilon_i g(X_i, A_i)}{\pi(X_i, A_i)} f(X_i, A_i)\Big], \tag{4.13b}$$

where $(\varepsilon_i)_{i=1}^n$ are i.i.d. Rademacher random variables independent of the data.

With this set-up, we are now ready to state some oracle inequalities satisfied by the weighted least-squares estimator (4.11). As in our earlier statement (4.9), these bounds are indexed by some $\mu \in \mathcal{F}$, and our risk bound involves the solutions

$$\tfrac{1}{s}\, \mathcal{S}_m\big((\mathcal{F} - \mu) \cap \mathbb{B}_\omega(s)\big) \leq s\,, \qquad \text{and} \tag{4.14a}$$

$$\tfrac{1}{r}\, \mathcal{R}_m\big((\mathcal{F} - \mu) \cap \mathbb{B}_\omega(r)\big) \leq \tfrac{\alpha_1 \alpha_2}{32}. \tag{4.14b}$$

Let $s_m(\mu)$ and $r_m(\mu)$, respectively, be the smallest non-negative solutions to these inequalities; see Proposition C.2 in Appendix C.1.2 for their guaranteed existence.

Theorem 4.2. *Under the convexity/compactness condition **(CC)** and small-ball condition **(SB)**, the two-stage estimate (4.7) based on the non-parametric least-squares estimate (4.11) satisfies the oracle inequality*

$$\mathbb{E}\big[\big(\widehat{\tau}_n - \tau^*\big)^2\big] \leq \frac{1}{n}\Big\{v_{\text{semi}}^2 + c \inf_{\mu \in \mathcal{F}} \Big(\|\mu - \mu^*\|_\omega^2 + \delta_n^2(\mu; \mathcal{F})\Big)\Big\} \tag{4.15a}$$

where the instance-dependent estimation error is given by

$$\delta_n^2(\mu; \mathcal{F}) := s_{n/2}^2(\mu) + r_{n/2}^2(\mu) + e^{-c'n} \operatorname{diam}_\omega^2(\mathcal{F} \cup \{\mu^*\}), \tag{4.15b}$$

for a pair (c, c') of constants depending only on the small-ball parameters (α_1, α_2).

See Section 4.4.2 for the proof of this theorem.

A few remarks are in order. The bound (4.15a) arises by combining the general bound from Theorem 4.1 with an oracle inequality that we establish for the weighted least-squares estimator (4.11). Compared to the efficient variance v^2_{semi}, this bound includes three additional terms: (i) the critical radii $s_{n/2}(\mu)$ and $r_{n/2}(\mu)$ that solve the fixed point equations; (ii) the approximation error under $\| \cdot \|_\omega$ norm; and (iii) an exponentially decaying term. For any fixed function class $\mathcal{F}$, if we take limits as the sample size n tends to infinity, we see that the asymptotic variance of $\widehat{\tau}_n$ takes the form

$$v^2_{\text{semi}} + c \inf_{\mu \in \mathcal{F}} \|\mu - \mu^*\|^2_\omega.$$

Consequently, the estimator may suffer from an efficiency loss depending on how well the unknown treatment effect μ^* can be approximated (in the weighted norm) by a member of $\mathcal{F}$. When the outcome noise $Y_i - \mu^*(X_i, A_i)$ is of constant order, inspection of equations (4.14a) and (4.14b) reveals that—as n tends to infinity—the critical radius $s_{n/2}(\mu)$ decays at a faster rate than $r_{n/2}(\mu)$. Therefore, the non-asymptotic excess risk—that is, any contribution to the MSE *in addition to* the efficient variance v^2_{semi}—primarily depends on two quantities: (a) the approximation error associated with approximating μ^* using a given function class $\mathcal{F}$, and (b) the (localized) metric entropy of this function class. Interestingly, both of these quantities turn out to be information-theoretically optimal in an instance-dependent sense. More precisely, in Section 4.3, we show that an efficiency loss depending on precisely the same approximation error is unavoidable; we further show that a sample size depending on a local notion of metric entropy is also needed for such a bound to be valid.

4.2.3 A simulation study

We now describe a simulation study that helps to illustrate the elbow effect predicted by our theory, along with the utility of using reweighted estimators of the treatment effect. We can model a missing data problem by using $A \in \mathbb{A} := \{0, 1\}$ as a binary indicator variable for "missingness"—that is, the outcome Y is observed if and only if $A = 1$. Taking ξ as the uniform distribution on the state space $\mathbb{X} := [0, 1]$, we take the weight function $g(x, a) = a$, so that our goal is to estimate the quantity $\mathbb{E}_\xi[\mu^*(X, 1)]$. Within this subsection, we abuse notation slightly by using μ to denote the function $\mu(\cdot, 1)$, and similarly μ^* for $\mu^*(\cdot, 1)$.

We allow the treatment effect to range over the first-order Sobolev smoothness class

$$\mathcal{F} := \Big\{ f : [0,1] \to \mathbb{R} \;\mid\; f(0) = 0, \; \|f\|^2_{\mathbb{H}^1} := \int_0^1 \big(f'(x)\big)^2 dx \leq 1 \Big\},$$

corresponding (roughly) to functions that have a first-order derivative f' with bounded L^2-norm. The function class $\mathcal{F}$ is a particular type of reproducing kernel Hilbert space (cf. Example 12.19 in the book [213]), so it is natural to consider various forms of kernel ridge regression.

Three possible estimators: So as to streamline notation, we let $S_{obs} \subseteq \{1, \ldots, m\}$ denote the subset of indices associated with observed outcomes—that is, $a_i = 1$ if and only if $i \in S_{obs}$. Our first estimator follows the protocol suggested by our theory: more precisely, we estimate the function μ^* using a *reweighted* form of kernel ridge regression (KRR)

$$\widehat{\mu}_{m,\omega} := \arg \min_{\mu \in \mathbb{L}^2([0,1])} \left\{ \sum_{i \in S_{obs}} \frac{g^2(X_i, 1)}{\pi^2(X_i, 1)} \big\{ Y_i - \mu(X_i) \big\}^2 + \lambda_m \|\mu\|_{\mathbb{H}^1}^2 \right\}, \tag{4.16a}$$

where $\lambda_m \geq 0$ is a regularization parameter (to be chosen by cross-validation). Let $\widehat{\tau}_{n,\omega}$ be the output of the two-stage procedure (4.7) when the reweighted KRR estimate is used in the first stage.

So as to isolate the effect of reweighting, we also implement the standard (unweighted) KRR estimate, given by

$$\widehat{\mu}_{m,\mathbb{L}^2} := \arg \min_{\mu \in \mathbb{L}^2([0,1])} \Big\{ \sum_{i \in S_{obs}} \big\{ Y_i - \mu(X_i) \big\}^2 + \lambda_m \|\mu\|_{\mathbb{H}^1}^2 \Big\}. \tag{4.16b}$$

Similarly, we let $\widehat{\tau}_{n,\mathbb{L}^2}$ denote the estimate obtained by using the unweighted KRR estimate as a first-stage quantity.

Finally, so as to provide an (unbeatable) baseline for comparison, we compute the *oracle estimate*

$$\widehat{\tau}_{n,\text{oracle}} := \frac{1}{n} \sum_{i=1}^{n} \Big\{ \frac{(Y_i - \mu^*(X_i))A_i}{\pi(X_i, 1)} + \mu^*(X_i) \Big\}. \tag{4.16c}$$

Here the term “oracle” refers to the fact that it provides an answer given the unrealistic assumption that the true treatment effect μ^* is known. Thus, this estimate cannot be computed based purely on observed quantities, but instead serves as a lower bound for calibrating. For each of these three estimators, we compute its n-rescaled mean-squared error

$$n \cdot \mathbb{E}\big[|\widehat{\tau}_{n,\diamond} - \tau^*|^2 \big] \quad \text{with } \diamond \in \big\{ \omega, \mathbb{L}^2, \text{oracle} \big\}. \tag{4.17}$$

Variance functions: Let us now describe an interesting family of variance functions σ^2 and propensity scores π. We begin by observing that if the standard deviation function σ takes values of the same order as the treatment effect μ^*, then the simple IPW estimator has a variance of the same order as the asymptotic efficient limit v_{semi}^2. Thus, in order to make the problem non-trivial and illustrate the advantage of semiparametric methods, we consider variance functions of the following type: for a given propensity score π and exponent $\gamma \in [0, 1]$, define

$$\sigma^2(x, 1) := \sigma_0^2 \, \big[\pi(x, 1)\big]^{\gamma}, \tag{4.18}$$

where $\sigma_0 > 0$ is a constant pre-factor. Since the optimal asymptotic variance v_{semi} contains the term $\mathbb{E}\big[\frac{\sigma^2(X,1)}{\pi(X,1)}\big]$, this family leads to a term of the form

$$\sigma_0^2 \, \mathbb{E}\left[\frac{1}{\{\pi(X,1)\}^{1-\gamma}}\right],$$

showing that (in rough terms) the exponent γ controls the influence of small values of the propensity score $\pi(X,1)$. At one extreme, for $\gamma = 1$, there is no dependence on these small values, whereas the other extreme $\gamma = 0$, it will be maximally sensitive to small values of the propensity score.

Propensity and treatment effect: We consider the following two choices of propensity scores

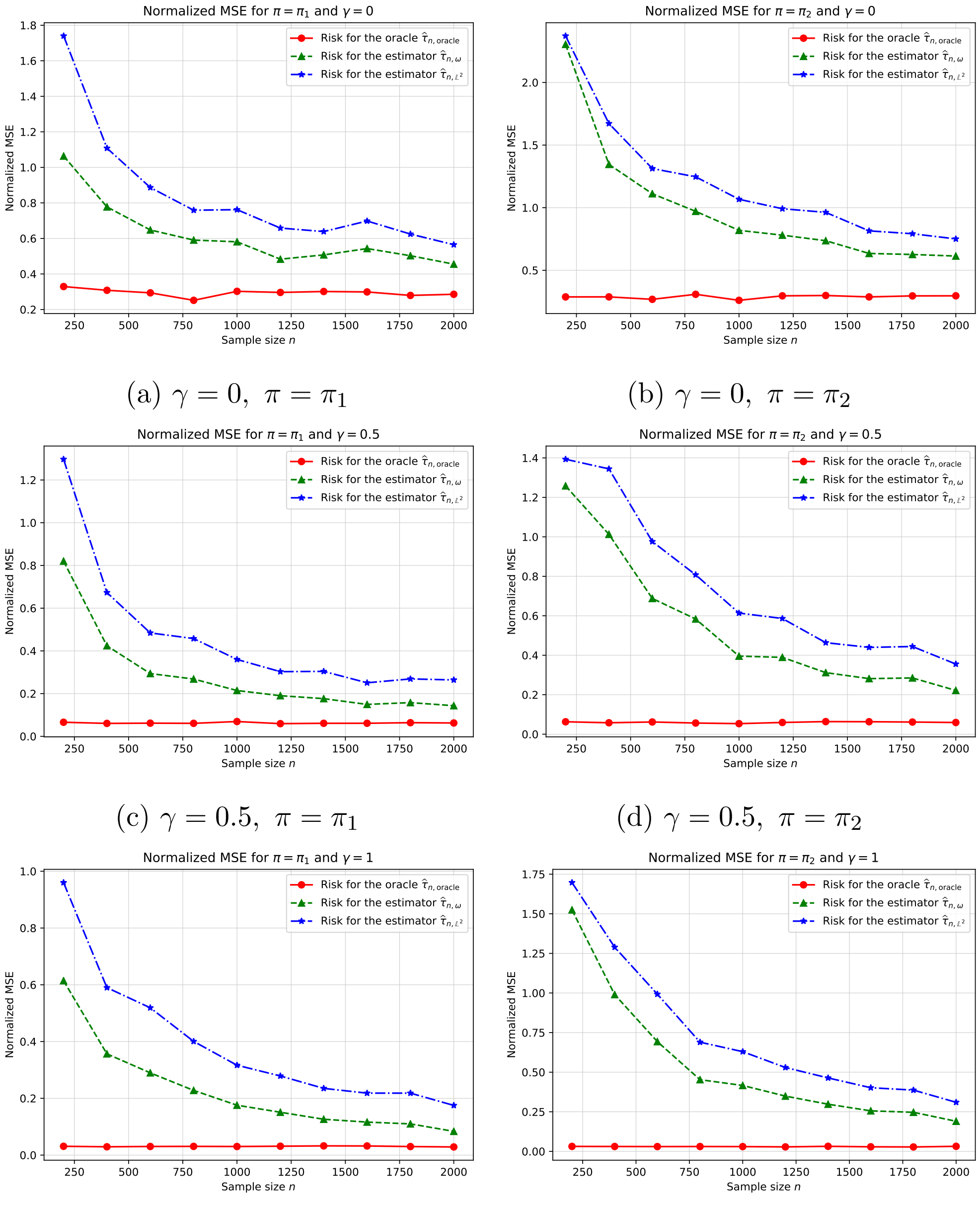
$$\pi_1(x,1) := \tfrac{1}{2} - \big(\tfrac{1}{2} - \pi_{\text{min}}\big)\sin(\pi x), \quad \text{and} \tag{4.19a}$$

$$\pi_2(x,1) := \tfrac{1}{2} - \{\tfrac{1}{2} - \pi_{\text{min}}\big)\sin(\pi x/2)\}, \tag{4.19b}$$

where $\pi_{\text{min}} := 0.005$. At the same time, we take the treatment effect to be the "tent" function

$$\mu^*(x) = \frac{1}{2} - \big|x - \tfrac{1}{2}\big| \qquad \text{for } x \in [0,1]. \tag{4.20}$$

Let us provide the rationale for these choices. Both propensity score functions take values in the interval $[\pi_{\text{min}}, 0.5]$, but achieve the minimal value π_{min} at different points within this interval: $x = 1/2$ for π_1 and at $x = 1$ for π_2. Now observe that for the missing data problem, the risk of the naïve IPW estimator (4.4) contains a term of the form $\mathbb{E}[\frac{\mu^*(X)^2}{\pi(X,1)}]$. Since our chosen treatment effect function (4.20) is maximized at $x = 1/2$, this term is much larger when we set $\pi = \pi_1$, which is minimized at $x = 1/2$. Thus, the propensity score π_1 serves as a "hard" example. On the other hand, the treatment effect is minimized at $x = 1$, where π_2 achieves its minimum, so that this represents an "easy" example.



(a) $\gamma = 0, \ \pi = \pi_1$

(b) $\gamma = 0, \ \pi = \pi_2$

(c) $\gamma = 0.5, \ \pi = \pi_1$

(d) $\gamma = 0.5, \ \pi = \pi_2$

(e) $\gamma = 1, \ \pi = \pi_1$

(f) $\gamma = 1, \ \pi = \pi_2$

Figure 4.1: Plots of the normalized MSE $n \cdot \mathbb{E}[|\widehat{\tau}_{n,\diamond} - \tau^*|]$ for $\diamond \in \{\omega, \mathbb{L}^2, \mathrm{oracle}\}$ versus the sample size. Each marker corresponds to a Monte Carlo estimate based on the empirical average of 1000 independent runs. As indicated in the figure titles, panels (a–f) show the normalized MSE of estimators for combinations of parameters: exponent $\gamma \in \{0, 0.5, 1\}$ in the top, middle and bottom rows respectively, and propensity scores π_1 and π_2 in the left and right columns, respectively. For each run, we used 5-fold cross validation to choose the value of regularization parameter $\lambda_n \in [10^{-1}, 10^2]$.

Simulation set-up and results: For each choice of exponent $\gamma \in \{0, 0.5, 1\}$ and each choice of propensity score $\pi \in \{\pi_1, \pi_2\}$, we implemented the reweighted estimator $\widehat{\tau}_{n,\omega}$, the standard estimator $\widehat{\tau}_{n,\mathbb{L}^2}$ and the oracle estimator $\widehat{\tau}_{n,\mathrm{oracle}}$. For each simulation, we varied the sample size over the range $n \in \{2000, 4000, \ldots, 18000, 20000\}$. For each run, we use 5-fold cross validation to choose the value of regularization parameter $\lambda_n \in [10^{-1}, 10^2]$. For each estimator and choice of simulation parameters, we performed a total of 1000 independent runs, and used them to form a Monte Carlo estimate of the true MSE.

Figure 4.1 provides plots of the n-rescaled mean-squared error (4.17) versus the sample size n for each of the three estimators in each of the six set-ups (three choices of γ, crossed with two choices of propensity score). In order to interpret the results, first note that consistent with the classical theory, the n-rescaled MSE of the oracle estimator stays at a constant level for different sample sizes. (There are small fluctuations, to be expected, since the quantity $\widehat{\tau}_{n,\mathrm{oracle}}$ itself is an empirical average over n samples.) Due to the design of our problem instances, the naïve IPW estimator (4.4) has much larger mean-squared error; in fact, it is so large that we do not include it in the plot, since doing so would change the scaling of the vertical axis. On the other hand, both the reweighted KRR two-stage estimate $\widehat{\tau}_{n,\omega}$ and the standard KRR two-stage estimate $\widehat{\tau}_{n,\mathbb{L}^2}$ exhibit the elbow effect suggested by our theory: when the sample size is relatively small, the high-order terms in the risk dominate, yielding a large normalized MSE. However, as the sample size increases, these high-order terms decay at a faster rate, so that the renormalized MSE eventually converges to the asymptotically optimal limit (i.e., the risk of the oracle estimator $\widehat{\tau}_{n,\mathrm{oracle}}$). In all our simulation instances, the weighted estimator $\widehat{\tau}_{n,\omega}$, which uses a *reweighted* non-parametric least-squares estimate in the first stage, outperforms the standard two-stage estimator $\widehat{\tau}_{n,\mathbb{L}^2}$ that does not reweight the objective. Again, this behavior is to be expected from our theory: in our bounds, the excess MSE due to errors in estimating the treatment effect is measured using the weighted norm.

4.2.4 Implications for particular models

We now return to our theoretical thread, and illustrate the consequences of our general theory for some concrete classes of outcome models.

4.2.4.1 Standard linear functions

We begin with the simplest case, namely that of linear outcome functions. For each $j = 1, \ldots, d$, let $\phi_j : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$ be a basis function, and consider functions that are linear in this representation—viz. $f_\theta(x, a) = \sum_{j=1}^{d} \theta_j \phi_j(x, a)$ for some parameter vector $\theta \in \mathbb{R}^d$. For a radius[1] $R_2 > 0$, we define the function class

$$\mathcal{F} := \Big\{ f_\theta \mid \; \|\theta\|_2 \leq R_2 \Big\}.$$

[1] We introduce this radius only to ensure compactness; in our final bound, the dependence on R_2 is exponentially decaying, so that it is of little consequence.

Our result assumes the existence of the following moment matrices:

$$\Sigma := \mathbb{E}\left[\frac{g^2(X,A)}{\pi^2(X,A)}\phi(X,A)\phi(X,A)^\top\right], \quad \text{and}$$
$$\Gamma_\sigma := \mathbb{E}\left[\frac{g^4(X,A)}{\pi^4(X,A)}\sigma^2(X,A)\phi(X,A)\phi(X,A)^\top\right].$$

With this set-up, we have:

Corollary 4.1. *Under the small-ball condition **(SB)**, given a sample size satisfying the lower bound $n \geq c_0\big\{d + \log(R_2\lambda_{max}(\Sigma))\big\}$, the estimate $\widehat{\tau}_n$ satisfies the bound*

$$\mathbb{E}\Big[|\widehat{\tau}_n - \tau^*|^2\Big] \leq \frac{1}{n}\left\{v_{\text{semi}}^2 + c\inf_{\mu\in\mathcal{F}}\|\mu-\mu^*\|_\omega^2\right\} + \frac{c}{n^2}\operatorname{trace}\Big(\Sigma^{-1}\Gamma_\sigma\Big), \tag{4.21}$$

where the constants (c_0, c) depend only on the small-ball parameters (α_1, α_2).

See Appendix C.2.1 for the proof of this corollary.

A few remarks are in order. First, Corollary 4.1 is valid in the regime $n \gtrsim d$, and the higher order term scales as $\mathcal{O}\left(d/n^2\right)$ in the worst case. Consequently, the optimal efficiency $v_{\text{semi}}^2 + \inf_{\mu\in\mathcal{F}}\|\mu-\mu^*\|_\omega^2$ is achieved when the sample size n exceeds the dimension d for linear models.[2]

It is worth noting, however, that in the well-specified case with $\mu^* \in \mathcal{F}$, the high-order term $\frac{c}{n^2}\operatorname{trace}(\Sigma^{-1}\Gamma_\sigma)$ in equation (4.21) does not necessarily correspond to the optimal risk for estimating the function μ^* under the weighted norm $\|\cdot\|_\omega$. Indeed, in order to estimate the function μ^* with the optimal semi-parametric efficiency under a linear model, an estimator that reweights samples with the function $\frac{1}{\sigma^2(X_i,A_i)}$ is the optimal choice, leading to a higher order term of the form $\frac{c}{n^2}\operatorname{trace}(\bar{\Sigma}^{-1}\Sigma)$, where $\bar{\Sigma} := \mathbb{E}\left[\frac{1}{\sigma^2(X,A)}\phi(X,A)\phi(X,A)^\top\right]$.[3] In general, the question of achieving optimality with respect to both the approximation error and high-order terms (under the $\|\cdot\|_\omega$-norm) is currently open.

4.2.4.2 Sparse linear models

Now we turn to sparse linear models for the outcome function. Recall the basis function set-up from Section 4.2.4.1, and the linear functions $f_\theta = \sum_{j=1}^d \theta_j\phi_j(x,a)$. Given a radius $R_1 > 0$, consider the class of linear functions induced by parameters with bounded ℓ_1-norm—viz.

$$\mathcal{F} := \Big\{f_\theta \;\mid\; \|\theta\|_1 \leq R_1\Big\}.$$

[2] We note in passing that the constant pre-factor c in front of the term $n^{-1}\inf_{\mu\in\mathcal{F}}\|\mu-\mu^*\|_\omega^2$ can be reduced to 1 using the arguments in Corollary 4.5.

[3] Note that $\begin{bmatrix}\bar{\Sigma} & \Sigma\\ \Sigma & \Gamma_\sigma\end{bmatrix} = \operatorname{cov}\left(\begin{bmatrix}\sigma(X,A)^{-1}\phi(X,A)\\ \frac{g(X,A)}{\pi(X,A)}\sigma(X,A)\phi(X,A)\end{bmatrix}\right) \succeq 0$. Taking the Schur complement we obtain that $\Gamma_\sigma \succeq \Sigma\bar{\Sigma}^{-1}\Sigma$, which implies that $\operatorname{trace}(\Sigma^{-1}\Gamma_\sigma) \geq \operatorname{trace}(\bar{\Sigma}^{-1}\Sigma)$.

Sparse linear models of this type arise in many applications, and have been the subject of intensive study (e.g., see the books [75, 213] and references therein).

We assume that the basis functions and outcome noise $Y - \mu^*(X, A)$ satisfy the moment bounds

$$\mathbb{E}\Big[\Big|\frac{g(X,A)}{\pi(X,A)}\big(Y - \mu^*(X,A)\big)\Big|^{\ell}\Big] \leq (\bar{\sigma}\sqrt{\ell})^{\ell}, \quad \text{for any } \ell = 1, 2, \ldots, \text{ and} \tag{4.22a}$$

$$\max_{j=1,\ldots,d} \mathbb{E}\Big[\Big|\frac{g(X,A)}{\pi(X,A)}\phi_j(X,A)\Big|^{\ell}\Big] \leq (\sigma\sqrt{\ell})^{\ell}, \quad \text{for any } \ell = 1, 2, \ldots. \tag{4.22b}$$

Under these conditions, we have the following guarantee:

Corollary 4.2. *Under the small-ball condition **(SB)** and the moment bounds (4.22), for any sparsity level $k = 1, \ldots, d$ and sample size n such that $n \geq c_0\Big\{\frac{\sigma^2 k \log(d)}{\lambda_{min}(\Sigma)} + \log^2(d) + \log(R_1 \cdot \lambda_{max}(\Sigma))\Big\}$, we have*

$$\mathbb{E}\Big[|\widehat{\tau}_n - \tau^*|^2\Big] \leq \frac{v^2_{\text{semi}}}{n} + \frac{c}{n} \inf_{\substack{\|\bar{\theta}\|_1 = R_1 \\ \|\bar{\theta}\|_0 \leq k}} \Big\{\|\mu^* - \langle\bar{\theta}, \phi(\cdot, \cdot)\rangle\|^2_\omega + \frac{\bar{\sigma}^2 \|\bar{\theta}\|_0 \log(d)}{n} \cdot \frac{\sigma^2}{\lambda_{min}(\Sigma)}\Big\},$$

where the constants (c_0, c) depend only on the small ball parameters (α_1, α_2).

See Appendix C.2.2 for the proof of this corollary.

A few remarks are in order. First, the additional risk term compared to the semiparametric efficient limit v^2_{semi}/n is similar to existing oracle inequalities for sparse linear regression (e.g., §7.3 in the book [213]). Notably, it adapts to the sparsity level of the approximating vector $\bar{\theta}$. The complexity of the auxiliary estimation task is characterized by the sparsity level $\|\bar{\theta}\|_0$ of the target function, which appears in both the high-order term of the risk bound and the sample size requirement. On the other hand, note that the $\|\cdot\|_\omega$-norm projection of the function μ^* to the set $\mathcal{F}$ may not be sparse. Instead of depending on the (potentially large) local complexity of such projection, the bound in Corollary 4.2 is adaptive to the trade-off between the sparsity level $\|\bar{\theta}\|_0$ and the approximation error $\|\mu^* - \langle\bar{\theta}, \phi(\cdot, \cdot)\rangle\|_\omega$.

4.2.4.3 Hölder smoothness classes

Let us now consider a non-parametric class of outcome functions. With state space $\mathbb{X} = [0,1]^{d_x}$ and action space $\mathbb{A} = [0,1]^{d_a}$, define the total dimension $p := d_x + d_a$. Given an integer order of smoothness $k > 0$, consider the class

$$\mathcal{F}_k := \Big\{\mu : [0,1]^p \to \mathbb{R} \ \mid \sup_{(x,a) \in [0,1]^p} |\partial^\alpha \mu(x,a)| \leq 1 \quad \text{for any } \alpha\mathbb{N}^p \text{ satisfying } \|\alpha\|_1 \leq k\Big\}.$$

Here for a multi-index $\alpha \in \mathbb{N}^p$, the quantity $\partial^\alpha f$ denotes the mixed partial derivative

$$\partial^\alpha f(x,a) := \Big(\prod_{j=1}^p \frac{\partial^{\alpha_j}}{\partial x_j^{\alpha_j}}\Big) f(x,a).$$

We impose the following assumptions on the likelihood ratio and random noise

$$\mathbb{E}\Big[\Big| \frac{g(X,A)}{\pi(X,A)} \big(Y - \mu^*(X,A)\big) \Big|^\ell \Big] \leq (\bar{\sigma}\sqrt{\ell})^\ell \quad \text{and} \tag{4.23a}$$

$$\mathbb{E}\Big[\Big| \frac{g(X,A)}{\pi(X,A)} \Big|^\ell \Big] \leq (\sigma\sqrt{\ell})^\ell, \quad \text{for any } \ell \in \mathbb{N}_+. \tag{4.23b}$$

Additionally, we impose the $L_2 - L_4$ hypercontractivity condition

$$\sqrt{\mathbb{E}\Big[\big(\tfrac{g(X,A)}{\pi(X,A)} f(X,A) \big)^4 \Big]} \leq M_{2\to 4}\, \mathbb{E}\Big[\big(\tfrac{g(X,A)}{\pi(X,A)} f(X,A) \big)^2 \Big] \quad \text{for any } f \in \mathcal{F}_k, \tag{4.23c}$$

which is slightly stronger than the small-ball condition **(SB)**.

Our result involves the sequences $\bar{r}_n := c_{\sigma,p/k}\, n^{-k/p} \log n$, and

$$\bar{s}_n := c_{\sigma,p/k} \bar{\sigma} \cdot \begin{cases} n^{-\frac{k}{2k+p}} & \text{if } p < 2k \\ n^{-1/4}\sqrt{\log n} & \text{if } p = 2k, \\ n^{-\frac{k}{2p}} & \text{if } p > 2k, \end{cases}$$

where the constant $c_{\sigma,p/k}$ depends on the tuple $(\sigma, p/k, M_{2\to 4})$.

With this notation, when the outcome function is approximated by the class $\mathcal{F}_k$, we have the following guarantee for treatment effect estimation:

Corollary 4.3. *Under the small-ball condition* ***(SB)*** *and the moment bounds* (4.23)*(a)–(c), we have*

$$\mathbb{E}\Big[|\widehat{\tau}_n - \tau^*|^2 \Big] \leq \frac{1}{n} \Big\{ v_{\text{semi}}^2 + c \inf_{\mu \in \mathcal{F}_k} \|\mu^* - \mu\|_\omega^2 \Big\} + \frac{c}{n} \Big\{ \bar{s}_n^2 + \bar{r}_n^2 \Big\}, \tag{4.24}$$

where the constant c depends only on the small-ball parameters (α_1, α_2).

See Appendix C.2.3 for the proof of this corollary.

It is worth making a few comments about this result. First, in the high-noise regime where $\bar{\sigma} \gtrsim 1$, the term $\bar{s}_n$ is dominant. This particular rate is optimal in the Donsker regime ($p < 2k$), but is sub-optimal when $p > 2k$. However, this sub-optimality only appears in high-order terms, and is of lower order for a sample size[4] n such that $\log n \gg (p/k)$. Indeed, even if the least-square estimators is sub-optimal for nonparametric estimation in non-Donsker classes, the reweighted least-square estimator (4.11) may still be desirable, as it is able to approximate the projection of the function μ^* onto the class $\mathcal{F}_k$, under the weighted norm $\|\cdot\|_\omega$.

[4] In fact, this lower bound cannot be avoided, as shown by our analysis in Section 4.3.2.1.

4.2.4.4 Monotone functions

We now consider a nonparametric problem that involves shape constraints—namely, that of monotonic functions. Let $\phi : \mathbb{X} \times \mathbb{A} \to [0, 1]$ be a one-dimensional feature mapping. We consider the class of outcome functions that are monotonic with respect to this feature—namely, the function class

$$\mathcal{F} := \Big\{ (x, a) \to f\big(\phi(x, a)\big) \;\mid\; f : [0, 1] \to [0, 1] \text{ is non-decreasing} \Big\}.$$

We assume the outcome and likelihood ratio are uniformly bounded—specifically, that

$$|Y_i| \leq 1 \quad \text{and} \quad \left| \frac{g(X_i, A_i)}{\pi(X_i, A_i)} \right| \leq b \quad \text{almost surely for } i = 1, 2, \ldots, n. \tag{4.25}$$

Under these conditions, we have the following result:

Corollary 4.4. *Under the small-ball condition* ***(SB)*** *and boundedness condition* (4.25), *we have*

$$\mathbb{E}\Big[|\widehat{\tau}_n - \tau^*|^2 \Big] \leq \frac{1}{n} \Big\{ v_{\text{semi}}^2 + c \inf_{\mu \in \mathcal{F}} \|\mu - \mu^*\|_\omega^2 \Big\} + \frac{c}{n} \Big(\frac{b^2}{n} \Big)^{2/3}, \tag{4.26}$$

where the constants (c_0, c) *depend only on the small-ball parameters* (α_1, α_2).

See Appendix C.2.4 for the proof of this corollary.

Note that compared to Corollaries 4.1– 4.3, Corollary 4.4 requires a stronger uniform bound on the likelihood ratio g/π: it is referred to as the *strict overlap condition* in the causal inference literature. In our analysis, this condition is required to make use of existing bracketing-based localized entropy control. Corollary 4.4 holds for any sample size $n \geq 1$, and we establish a matching lower bound as a consequence of Proposition 4.1 to be stated in the sequel. It should be noted that the likelihood ratio bound b might be large, in which case the high-order term in Corollary 4.4 could be dominant (at least for small sample sizes). As with previous examples, optimal estimation of the scalar τ^* requires optimal estimation of the function μ^* under $\|\cdot\|_\omega$-norm. How to do so optimally for isotonic classes appears to be an open question.

4.2.5 Non-asymptotic normal approximation

Note that the oracle inequality in Theorem 4.2 involves an approximation factor depending on the small-ball condition in Assumption **(SB)**, as well as other universal constants. Even with sample size n tending to infinity, the result of Theorem 4.2 does not ensure that the auxiliary estimator $\widehat{\mu}_{n/2}$ converges to a limiting point. This issue, while less relevant for the mean-squared error bound in Theorem 4.2, assumes importance in the inferential setting. In this case, we do need the auxiliary estimator to converge so as to be able to characterize the approximation error.

In order to address this issue, we first define the orthogonal projection within the class

$$\bar{\mu} := \arg\min_{\mu \in \mathcal{F}} \|\mu - \mu^*\|_\omega. \tag{4.27}$$

Our analysis also involves an additional squared Rademacher complexity, one which involves the *difference* $\mu^* - \bar{\mu}$. It is given by

$$\mathcal{D}_m^2(\mathcal{H}) := \mathbb{E}\Big[\sup_{f \in \mathcal{H}} \Big\{\frac{1}{m}\sum_{i=1}^m \varepsilon_i \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \big[\mu^*(X_i, A_i) - \bar{\mu}(X_i, A_i)\big] \, f(X_i, A_i)\Big\}^2\Big]. \tag{4.28a}$$

We let $d_m > 0$ be the unique solution to the fixed point equation

$$\tfrac{1}{d}\,\mathcal{D}_m\big((\mathcal{F} - \bar{\mu}) \cap \mathbb{B}_\omega(d)\big) = d. \tag{4.28b}$$

The existence and uniqueness is guaranteed by an argument analogous to that used in the proof of Proposition C.2.

In order to derive a non-asymptotic CLT, we need a finite fourth moment

$$M_4 := \mathbb{E}\Big[\Big\{\frac{g(X, A)}{\pi(X, A)}\big(Y - \bar{\mu}(X, A)\big) + \langle g(X, \cdot), \, \bar{\mu}(X, \cdot)\rangle_\lambda\Big\}^4\Big].$$

The statement also involves the excess variance

$$v^2(\bar{\mu}) := \mathbb{E}\Big[\operatorname{var}\Big(\frac{g(X, A)}{\pi(X, A)} \cdot \{\mu^*(X, A) - \bar{\mu}(X, A)\} \mid X\Big)\Big].$$

With these definitions, we have the following guarantee:

Corollary 4.5. *Under Assumptions **(CC)** and **(SB)**, the two-stage estimator* (4.7) *satisfies the Wasserstein distance bound*

$$\mathcal{W}_1\big(\sqrt{n}\widehat{\tau}_n, Z\big) \leq \tfrac{4\sqrt{M_4}}{[v_{\text{semi}} + v(\bar{\mu})]} \frac{1}{\sqrt{n}} + c\big\{r_{n/2} + s_{n/2} + d_n\big\} + \operatorname{diam}_\omega(\mathcal{F} \cup \{\mu^*\}) \cdot e^{-c'n}, \tag{4.29}$$

where $Z \sim \mathcal{N}\big(0, v_{\text{semi}}^2 + v^2(\bar{\mu})\big)$*, and the pair* (c, c') *of constants depend only on the small-ball parameters* (α_1, α_2)*.*

See Section 4.4.3 for the proof of this corollary.

A few remarks are in order. First, in the limit $n \to +\infty$, Corollary 4.5 guarantees asymptotic normality of the estimate $\widehat{\tau}_n$, with asymptotic variance $v_{\text{semi}}^2 + v^2(\bar{\mu})$. In contrast, the non-asymptotic result given here makes valid inference possible at a finite-sample level, by taking into account the estimation error for auxiliary functions.

Compared to the risk bound in Theorem 4.2, the right-hand-side of equation (4.29) contains two terms: the first term $\frac{4M_4^2}{(v_{\mathrm{semi}}+v(\bar{\mu}))\sqrt{n}}$ is the Berry–Esseen error, and an additional critical radius $d_{n/2}$ depending on the localized multiplier Rademacher complexity. When the approximation error $\mu^* - \bar{\mu}$ is of order $o(1)$, the multiplier Rademacher complexity $\mathcal{D}_{n/2}$ becomes (asymptotically) smaller than the Rademacher complexity $\mathcal{S}_{n/2}$, resulting in a critical radius $d_{n/2}(\bar{\mu})$ smaller than $s_{n/2}(\bar{\mu})$. On the other hand, the efficiency loss in Corollary 4.5 is the exact variance $v^2(\bar{\mu})$ with unity pre-factor, which exhibits a smaller efficiency loss compared to Theorem 4.2.

Excess variance compared to approximation error: It should be noted that the excess variance term $v^2(\bar{\mu})$ in Corollary 4.5 is smaller than the best approximation error $\inf_{\mu^* \in \mathcal{F}} \|\bar{\mu} - \mu^*\|_\omega^2$. Indeed, the difference $\Delta := \|\bar{\mu} - \mu^*\|_\omega^2 - v^2(\bar{\mu})$ can be written as

$$\begin{aligned}\Delta &= \mathbb{E}\Big[\Big(\frac{g(X,A)}{\pi(X,A)} \cdot \big(\mu^* - \bar{\mu}\big)(X,A)\Big)^2\Big] - \mathbb{E}\Big[\operatorname{var}\Big(\frac{g(X,A)}{\pi(X,A)} \cdot \big(\mu^* - \bar{\mu}\big)(X,A) \mid X\Big)\Big] \\ &= \mathbb{E}_{\xi^*}\big[\langle g(X,\cdot),\, (\mu^* - \bar{\mu})(X,\cdot)\rangle_\lambda^2\big]. \qquad (4.30)\end{aligned}$$

When considering minimax risk over a local neighborhood around the function μ^*, the difference term computed above is dominated by the supremum of the asymptotic efficient variance v_{semi}^2 evaluated within this neighborhood. Consequently, the upper bound induced by Corollary 4.5 does not contradict the local minimax lower bound in Theorem 4.3; and since the difference $\|\bar{\mu} - \mu^*\|_\omega^2 - v^2(\bar{\mu})$ does not involve the importance weight ratio g/π, this term is usually much smaller than the weighted norm term $\|\bar{\mu} - \mu^*\|_\omega^2$.

On the other hand, Corollary 4.5 and equation (4.30) provide guidance on the way of achieving the optimal pointwise exact asymptotic variance. In particular, when we choose a function class $\mathcal{F}$ such that $\langle h(x,\cdot),\, g(x,\cdot)\rangle_\lambda = 0$ for any $h \in \mathcal{F}$ and $x \in \mathbb{X}$, the expression (4.30) becomes a constant independent of the choice of $\bar{\mu}$. For such a function class, a function $\bar{\mu}$ that minimizes the approximation error $\|\mu - \mu^*\|_\omega^2$ will also minimize the variance $v^2(\mu)$. Such a class can be easily constructed from any function class $\mathcal{H}$ by taking a function $h \in \mathcal{H}$ and replacing it with $f(x,a) := h(x,a) - \langle h(x,\cdot),\, g(x,\cdot)\rangle_\lambda$. And the optimal variance can still be written in the form of approximation error:

$$v(\bar{\mu}) = \|\bar{\mu} - \widetilde{\mu}^*\|_\omega, \quad \text{where } \widetilde{\mu}^*(x,a) := \mu^*(x,a) - \langle \mu^*(x,\cdot),\, g(x,\cdot)\rangle_\lambda.$$

Indeed, the functional v can be seen as the induced norm of $\|\cdot\|_\omega$ in the quotient space generated by $\mathbb{L}_\omega^2$ modulo the subspace $\mathbb{L}^2(\xi^*)$ that contains functions depending only on the state but not action.

4.3 Minimax lower bounds

Thus far, we have derived upper bounds for particular estimators of the linear functional $\tau(\mathcal{I})$, ones that involve the weighted norm (4.8a). In this section, we turn to the complementary question of deriving local minimax lower bounds for the problem. Recall that any given problem instance is characterized by a quadruple of the form (ξ^*, π, μ^*, g). In this section, we state some lower bounds that hold uniformly over all estimators that are permitted to know both the policy π and the weight function g. With (π, g) known, the instance is parameterized by the pair (ξ^*, μ^*), and we derive two types of lower bounds:

- In Theorem 4.3, we study local minimax bounds in which the unknown probability distribution ξ^* and potential outcome function are allowed to range over suitably defined neighborhoods of a given target pair (ξ^*, μ^*), respectively, but without structural conditions on the function classes.
- In Proposition 4.1, we impose structural conditions on the function class $\mathcal{F}$ used to model μ^*, and prove a lower bound that involves the complexity of $\mathcal{F}$—in particular, via its fat shattering dimension. This lower bound shows that if the sample size is smaller than the function complexity, then any estimator has a mean-squared error larger than the efficient variance.

4.3.1 Instance-dependent bounds under mis-specification

Given a problem instance $\mathcal{I}^* = (\xi^*, \mu^*)$ and an error function $\delta : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$, we consider the local neighborhoods

$$\mathfrak{N}^{\text{val}}_{\delta}(\mu^*) := \Big\{ \mu \;\mid\; |\mu(x, a) - \mu^*(x, a)| \leq \delta(x, a) \quad \text{for } (x, a) \in \mathbb{X} \times \mathbb{A} \Big\}, \tag{4.31a}$$

$$\mathfrak{N}^{\text{prob}}(\xi^*) := \Big\{ \xi \;\mid\; D_{\mathrm{KL}}\left(\xi \,\|\, \xi^*\right) \leq \tfrac{1}{n} \Big\}. \tag{4.31b}$$

Our goal is to lower bound the *local minimax risk*

$$\mathscr{M}_n\big(\mathbb{C}_\delta(\mathcal{I}^*)\big) := \inf_{\widehat{\tau}_n} \sup_{\mathcal{I} \in \mathbb{C}_\delta(\mathcal{I}^*)} \mathbb{E}\left|\tau - \widehat{\tau}_n\right|^2 \quad \text{where } \mathbb{C}_\delta(\mathcal{I}^*) := \big\{(\xi, \mu) \in \mathfrak{N}^{\text{prob}}(\xi^*)\big\} \times \mathfrak{N}^{\text{val}}_{\delta}(\mu^*). \tag{4.32}$$

Let us now specify the assumptions that underlie our lower bounds.

Assumptions for lower bound: First, we require some tail control on certain random variables, stated in terms of the $(2, 4)$-*moment-ratio* $\|Y\|_{2\to 4} := \frac{\sqrt{\mathbb{E}[Y^4]}}{\mathbb{E}[Y^2]}$.

(MR) The random variables

$$Z(X, A) := \frac{\delta(X, A) g(X, A)}{\pi(X, A)}, \quad \text{and} \quad Z'(X, A) := \langle \mu^*(X, \cdot),\, g(X, \cdot) \rangle_\lambda - \tau(\mathcal{I}^*) \tag{4.33}$$

have finite $(2, 4)$-moment ratios $M_{2\to 4} := \|Z\|_{2\to 4}$ and $M'_{2\to 4} := \|Z'\|_{2\to 4}$.

Second, we require the existence of a constant $c_{\max} > 0$ such that the distribution ξ^* satisfies the following *compatibility condition*.

(COM) For a finite state space $\mathbb{X}$, we require $\xi^*(x) \leq c_{\max}/|\mathbb{X}|$ for all $x \in \mathbb{X}$. If $\mathbb{X}$ is infinite, we require that ξ^* is non-atomic (i.e., $\xi^*(\{x\}) = 0$ for all $x \in \mathbb{X}$), and set $c_{\max} = 1$ for concreteness.

Finally, we impose a lower bound on the *local neighborhood size*:

(LN) The neighborhood function δ satisfies the lower bound

$$\sqrt{n}\, \delta(x, a) \geq \frac{g(x, a)\sigma^2(x, a)}{\pi(x, a)\|\sigma\|_\omega} \quad \text{for any } (x, a) \in \mathbb{X} \times \mathbb{A}. \tag{4.34}$$

In the following statement, we use c and c' to denote universal constants.

Theorem 4.3. *Under Assumptions* ***(MR)***, ***(COM)*** *and* ***(LN)***, *given a sample size lower bounded as* $n \geq c' \max\{(M'_{2\to4})^2, M^2_{2\to4}\}$, *the local minimax risk over the class* $\mathbb{C}_\delta(\mathcal{I}^*)$ *is lower bounded as*

$$\mathscr{M}_n\big(\mathbb{C}_\delta(\mathcal{I}^*)\big) \geq \frac{c}{n}\begin{cases} v^2_{\text{semi}} & \text{if } n \geq \frac{|\mathbb{X}|}{c_{\max}} \\ v^2_{\text{semi}} + \|\delta\|^2_\omega & \text{otherwise.} \end{cases} \tag{4.35}$$

We prove this claim in Section 4.5.1.

It is worth understanding the reasons for each of the assumptions required for this lower bound to hold. The compatibility condition **(COM)** is needed to ensure that no single state can take a significant proportion of probability mass under ξ^*. If this condition is violated, then it could be possible to construct a low MSE estimate of the outcome function via an empirical average, which would then break our lower bound. The neighborhood condition **(LN)** ensures that the set of problems considered by the adversary is large enough to be able to capture the term $\|\sigma\|^2_\omega$ in the optimal variance v^2_{semi}. Without this assumption, the "local-neighborhood" restriction on certain states-action pairs could be more informative than the data itself.

Now let us understand some consequences of Theorem 4.3. First, it establishes the information-theoretic optimality of Theorem 4.2 and Corollary 4.5 in an instance-dependent sense. Consider a function class $\mathcal{F}$ that approximately contains the true outcome function μ^*; more formally, consider the δ-approximate version of $\mathcal{F}$ given by

$$\mathcal{F}_\delta := \Big\{\widetilde{\mu} \in \mathbb{L}^2_\omega \mid \exists \mu \in \mathcal{F} \text{ such that } |\mu(x, a) - \widetilde{\mu}(x, a)| \leq \delta(x, a) \quad \text{for all } (x, a) \in \mathbb{X} \times \mathbb{A}\Big\},$$

and let us suppose that $\mu^* \in \mathcal{F}_\delta$. With this notation, Theorem 4.3 implies a lower bound of the form

$$\inf_{\widehat{\tau}_n} \sup_{\substack{\mu \in \mathcal{F}_\delta \\ \xi \in \mathfrak{N}^{prob}(\xi^*)}} \mathbb{E}\big[|\tau - \widehat{\tau}_n|^2\big] \geq \frac{c}{n}\Big\{\sup_{\mu \in \mathcal{F}} \text{var}\left(\langle g(X, \cdot),\, \mu(X, \cdot)\rangle_\lambda + \|\sigma\|^2_\omega + \|\delta\|^2_\omega\right\}. \tag{4.36}$$

Thus, we see that the efficiency loss due to errors in estimating the outcome function is unavoidable; moreover, this loss is measured in the weighted norm $\| \cdot \|_\omega$ that also appeared centrally in our upper bounds.

It is also worth noting that for a finite cardinality state space $\mathbb{X}$, Theorem 4.3 exhibits a "phase transition" in the following sense: for a sample size $n \gg |\mathbb{X}|$, the lower bound is simply a non-asymptotic version of the semi-parametric efficiency lower bound (up to the pre-factor[5] $c > 1$). On the other hand, when $n < |\mathbb{X}|$, then the term $\|\delta\|_\omega^2/n$ starts to play a significant role. For an infinite state space $\mathbb{X}$ without atoms, the lower bound (4.35) holds for any sample size n.

By taking $\mu^* = 0$ and $\delta(x,a) = 1$ for all (x,a), equation (4.35) implies the global minimax lower bound

$$\inf_{\widehat{\tau}_n} \sup_{\|\mu\|_\infty \leq 1, \; \xi = \xi^*} \mathbb{E}\big[|\tau - \widehat{\tau}_n|^2 \big] \geq \frac{c}{n} \int_{\mathbb{A}} \mathbb{E}_{\xi^*} \Big[\frac{g^2(X,a)}{\pi(X,a)} \Big] d\lambda(a), \tag{4.37}$$

valid whenever $n \leq |\mathbb{X}|$.

The χ^2-type term on the right-hand side of this bound is related to—but distinct from—results from past work on off-policy evaluation in bandits [215, 133]. In this past work, a term of this type arose due to noisiness of the observations. In contrast, our lower bound (4.37) is valid even if the observed outcome is noiseless, and the additional risk depending on the weighted norm arises instead from the impossibility of estimating μ^* itself.

4.3.2 Lower bounds for structured function classes

As we have remarked, in the special case of a finite state space ($|\mathbb{X}| < \infty$), Theorem 4.3 exhibits an interesting transition at the boundary $n \asymp |\mathbb{X}|$. On the other hand, for an infinite state space, the stronger lower bound in Theorem 4.3—namely, that involving $\|\delta\|_\omega^2$—is always in force. It should be noted, however, that this strong lower bound depends critically on the fact that Theorem 4.3 imposes *no conditions* on the function class $\mathcal{F}$ of possible treatment effects, so that the error necessarily involves the local perturbation δ.

In this section, we undertake a more refined investigation of this issue. In particular, when some complexity control is imposed upon $\mathcal{F}$, then the lower bounds again exhibit a transition: any procedure pays a price only when the sample size is sufficiently small relative to the complexity of $\mathcal{F}$. In doing so, we assess the complexity of $\mathcal{F}$ using the *fat-shattering dimension*, a scale-sensitive version of the VC dimension [95, 3].

(FS) A collection of data points $(s_i)_{i=1}^N$ is *shattered at scale* δ by a function class $\mathcal{H} : \mathbb{X} \to \mathbb{R}$ means that for any subset $S \subseteq \{1, \ldots, N\}$, there exists a function $f \in \mathcal{H}$ and a vector $t \in \mathbb{R}^N$ such that

$$f(s_i) \geq t_i + \delta \quad \text{for all } i \in S, \text{ and} \quad f(s_i) \leq t_i - \delta \quad \text{for all } i \notin S. \tag{4.38}$$

[5] Using slightly more involved argument, this pre-factor can actually be made arbitrarily close to unity.

The fat-shattering dimension $\mathrm{fat}_\delta(\mathcal{H})$ is the largest integer N for which there exists some sequence $(s_i)_{i=1}^N$ shattered by $\mathcal{H}$ at scale δ.

In order to illustrate a transition depending on the fat shattering dimension, we consider the minimax risk

$$\mathscr{M}_n(\mathcal{F}) := \inf_{\widehat{\tau}_n} \sup_{\substack{\mu \in \mathcal{F} \\ \xi \in \mathcal{P}(\mathbb{X})}} \mathbb{E}\big[|\widehat{\tau}_n - \tau(\xi, \mu)|^2 \big],$$

specializing to the case of a finite action space $\mathbb{A}$ equipped with the counting measure λ. We further assume that the class $\mathcal{F}$ is a product of classes associated to each action, i.e., $\mathcal{F} = \bigotimes_{a \in \mathbb{A}} \mathcal{F}_a$, with $\mathcal{F}_a$ being a *convex subset* of real-valued functions on the state space $\mathbb{X}$. We also assume the existence[6] of a sequence $\{s_j\}_{j=1}^D$ that, for each action $a \in \mathbb{A}$, is shattered by $\mathcal{F}_a$ at scale δ_a. Analogous to the moment ratio assumption **(MR)**, we need an additional assumption that

$$M_{2\to4} := \|\frac{g(X, A)}{\pi(X, A)} \delta_A \|_{2\to4} < +\infty, \quad \text{for } X \sim \mathcal{U}(\{x_j\}_{j=1}^D) \text{ and } A \sim \pi(X, \cdot). \tag{4.39}$$

Proposition 4.1. *With the set-up given above, there are universal constants (c, c') such that for any sample size satisfying $n \geq M_{2\to4}^2$ and $n \leq c'D$, we have the lower bound*

$$\mathscr{M}_n(\mathcal{F}) \geq \frac{c}{n} \Big\{ \frac{1}{D} \sum_{j=1}^D \sum_{a \in \mathbb{A}} \frac{g^2(s_j, a)}{\pi(s_j, a)} \delta_a^2 \Big\}. \tag{4.40}$$

See Section 4.5.2 for the proof of this claim.

A few remarks are in order. First, if we take ξ to be the uniform distribution over the sequence $\{x_j\}_{j=1}^D$, the right-hand-side of the bound (4.40) is equal to $\frac{c}{n}\|\delta\|_\omega^2$. Thus, Proposition 4.1 is the analogue of our earlier lower bound (4.32) under the additional restriction that the treatment effect function μ^* belong the given function class $\mathcal{F}$. This lower bound holds as long as $n \leq c'D$, so that the fat shattering dimension D as opposed to the state space cardinality $|\mathbb{X}|$ (for a discrete state space) demarcates the transition between different regimes.

An important take-away of Proposition 4.1 is that the sample size must exceed the "complexity" of the function class $\mathcal{F}$ in order for the asymptotically efficient variance v_{semi}^2 to be dominant. More precisely, suppose that—for some scale $\delta > 0$—the sample size is smaller than the fat-shattering dimension $\mathrm{fat}_\delta(\mathcal{F})$. In this regime, the naïve IPW estimator (4.4) is actually instance-optimal, even when there is no noise. Observe that its risk contains a term of the form $\sum_{a \in \mathbb{A}} \mathbb{E}\big[\frac{g^2(X,a)}{\pi(X,a)}\big]$, which is *not present* in the asymptotically efficient variance v_{semi}^2.

By contrast, suppose instead that the sample size exceeds the fat-shattering dimension. In this regime, it is possible to obtain non-trivial estimates of the treatment effect, so

[6] Thus, per force, we have $D \leq \mathrm{fat}_{\delta_a}(\mathcal{F}_a)$ for each $a \in \mathbb{A}$.

that superior estimates of τ^* are possible. From the point of view of our theory, one can use the fat shattering dimension $\mathrm{fat}_\delta(\mathcal{F})$ to control the δ-covering number [140], and hence the Rademacher complexities that arise in our theory. Doing so leads to non-trivial radii $(s_{n/2}, r_{n/2})$ in Theorem 4.2, and consequently, the asymptotically efficient variance will become the dominant term. We illustrate this line of reasoning via various examples in Section 4.2.4.

It should be noted that a sample size scaling with the fat-shattering dimension is also known to be necessary and sufficient to learn the function μ^* with $o(1)$ error [95, 9, 3]. These classical results, in combination with our Proposition 4.1 and Theorem 4.2, exhibit that necessary conditions on the sample size for consistent estimation of the function μ^* are equivalent to those requiring for achieving the asymptotically efficient variance in estimating the scalar τ^*.

Worst-case interpretation: It is worthwhile interpreting the bound (4.40) in a worst-case setting. Consider a problem with binary action space $\mathbb{A} = \{0, 1\}$ and $g(x, a) = 2a - 1$. Suppose that we use a given function class $\mathcal{H}$ (consisting of functions from the state space $\mathbb{X}$ to the interval $[0, 1]$) as a model[7] of both of the functions $\mu^*(\cdot, 0)$ and $\mu^*(\cdot, 1)$. Given a scalar $\pi_{\min} \in (0, 1/2)$, let $\Pi(\pi_{\min})$ be the set of propensity score functions such that $\pi(x, 1) \in [\pi_{\min}, 1 - \pi_{\min}]$ for any $x \in \mathbb{X}$. By taking the worst-case over this class, we find that there are universal constants $c, c' > 0$ such that

$$\sup_{\pi \in \Pi(\pi_{\min})} \inf_{\widehat{\tau}_n} \sup_{\mu^* \in \mathcal{H}} \mathbb{E}\big[\,|\widehat{\tau}_n - \tau|^2\,\big] \geq c \begin{cases} \dfrac{1}{n} + \dfrac{\bar{\delta}^2}{n\pi_{\min}} & \text{for } n \leq c' \mathrm{fat}_{\bar{\delta}}(\mathcal{F}), \\ \dfrac{1}{n} & \text{otherwise}, \end{cases} \tag{4.41}$$

for any $\bar{\delta} \in (0, 1)$. The validity of this lower bound does not depend on noise in the outcome observations (and therefore applies to noiseless settings). Since $\pi_{\min} \in (0, 1)$, any scalar $\bar{\delta} \gg \sqrt{\pi_{\min}}$ yields a non-trivial risk lower bound for sample sizes n below the threshold $\mathrm{fat}_{\bar{\delta}}(\mathcal{F})$.

Relaxing the convexity requirement: Proposition 4.1 is based on the assumption each function class $\mathcal{F}_a$ is convex. This requirement can be relaxed if we require instead that the sequence $\{x_i\}_{i=1}^D$ be shattered with the inequalities (4.38) all holding with equality—that is, for any subset S, there exists a function $f \in \mathcal{H}$ and a vector $t \in \mathbb{R}^D$ such that

$$f(s_i) = t_i + \delta \quad \text{for all } i \in S, \text{ and} \quad f(s_i) = t_i - \delta \quad \text{for all } i \notin S. \tag{4.42}$$

For example, any class of functions mapping $\mathbb{X}$ to the binary set $\{0, 1\}$ satisfies this condition with $D = \mathrm{VC}(\mathcal{F})$ and $\delta = 1/2$. In the following, we provide additional examples of non-convex function classes that satisfy equation (4.42).

[7] We write $\mu^* \in \mathcal{H}$ as a shorthand for this set-up.

4.3.2.1 Examples of fat-shattering lower bounds

We discuss examples of the fat-shattering lower bound (4.40) in this section. We first describe some implications for convex classes. We then treat some non-convex classes using the strengthened shattering condition (4.42).

Example 4.1 (Smoothness class in high dimensions)**.** We begin with a standard Hölder class on the domain $\mathbb{X} = [-1, 1]^p$. For some index $k = 1, 2, \ldots$, we consider functions that are k-order smooth in the following sense

$$\mathcal{F}_k^{(\mathrm{Lip})} := \Big\{ f : [-1, 1]^p \to \mathbb{R} \;\mid\; \sup_{x \in \mathbb{X}} \max_{\alpha \in \mathbb{N}^p,\ \|\alpha\|_1 \leq k} |\partial^\alpha f(x)| \leq 1 \Big\}. \tag{4.43}$$

By inspection, the class $\mathcal{F}$ is convex. We can lower bound its fat shattering dimension by combining classical results on L^2-covering number of smooth functions [103] with the relation between fat shattering dimension and covering number [140], we conclude that

$$\mathrm{fat}_t\big(\mathcal{F}_k^{(\mathrm{Lip})}\big) \geq 2^{p/k}, \qquad \text{for a sufficiently small scale } t > 0. \tag{4.44}$$

Consequently, for a function class with a constant order of smoothness (i.e., not scaling with the dimension p), the sample size required to approach the asymptotototically optimal efficiency scales exponentially in p. ♣

Example 4.2 (Single index models)**.** Next we consider a class of single index models with domain $\mathbb{X} = [-1, 1]^p$. Since our main goal is to understand scaling issues, we may assume that p is an integer power of 2 without loss of generality. Given a differentiable function $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi(0) = 0$ and $\varphi'(x) \geq \ell_\varphi > 0$ for all $x \in \mathbb{R}$, we consider ridge functions of the form $g_\beta(x) := \varphi\big(\langle \beta, \, x \rangle\big)$. For a radius $R > 0$, we define the class

$$\mathcal{F}_R^{GLM} := \Big\{ g_\beta \;\mid\; \|\beta\|_2 \leq R \Big\}. \tag{4.45}$$

Let us verify the strengthened shattering condition (4.42). Suppose that the vectors $\{x_j\}_{j=1}^p$ define the Hadamard basis in p dimensions, and so are orthonormal. Taking $t_j = 0$ for $j = 1, \ldots, p$, given any binary vector $\zeta \in \{-1, 1\}^p$, we define the p-dimensional vector

$$\beta(\zeta) = \frac{1}{p} \sum_{j=1}^p \varphi^{-1}\big(\zeta_j a R\big) \, x_j,$$

Given the orthonormality of the vectors $\{x_j\}_{j=1}^p$, we have

$$\langle \beta(\zeta), \, x_\ell \rangle = \varphi^{-1}\big(\zeta_\ell a R\big) \qquad \text{for each } \ell = 1, \ldots, p,$$

and thus $g_{\beta(\zeta)}(s_\ell) = \zeta_\ell a R$ for each $\ell = 1, 2, \ldots, p$. Consequently, the function class $\mathcal{F}_R^{GLM}$ satisfies the strengthened shattering condition (4.42) with fat shattering dimension $D = p$ and scale $\delta = aR$. So when the outcome follows a generalized linear model, a sample size must be at least of the order p in order to match the optimal asymptotic efficiency. ♣

Example 4.3 (Sparse linear models)**.** Once again take the domain $[-1, 1]^p$, and consider linear functions of the form $f_\beta(x) = \langle \beta, \, x \rangle$ for some parameter vector $\beta \in \mathbb{R}^p$. Given a positive integer $s \in \{1, \ldots, p\}$, known as the *sparsity index*, we consider the set of s-sparse linear functions

$$\mathcal{F}_s^{sparse} := \Big\{ f_\beta \;\mid\; |\text{supp}(\beta)| \leq s, \;\; \text{and} \;\; \|\beta\|_\infty \leq 1 \Big\}. \tag{4.46}$$

As noted previously, sparse linear models of this type have a wide range of applications (e.g., see the book [75]).

In Appendix C.2.5, we prove that the strong shattering condition (4.42) holds with fat shattering dimension $D \asymp s \log\big(\frac{ep}{s}\big)$. Consequently, if the outcome functions μ^* follow a sparse linear model, at least $\Omega\Big(s \log\big(\frac{ep}{s}\big)\Big)$ samples are needed to make use of this fact. ♣

4.4 Proofs of upper bounds

In this section, we prove the upper bounds on the estimation error (Theorem 4.1 and Theorem 4.2), along with corollaries for specific models.

4.4.1 Proof of Theorem 4.1

The error can be decomposed into three terms as $\widehat{\tau}_n - \tau^* = T_* - T_1 - T_2$, where

$$\begin{aligned}
T_* &:= \frac{1}{n} \sum_{i=1}^{n} \Big\{ \tfrac{g(X_i, A_i)}{\pi(X_i, A_i)} Y_i - \tau^* - f^*(X_i, A_i) \Big\}, \\
T_1 &:= \frac{1}{n} \sum_{i=1}^{n/2} \big(\widehat{f}^{(2)}_{n/2}(X_i, A_i) - f^*(X_i, A_i) \big), \\
T_2 &:= \frac{1}{n} \sum_{i=n/2+1}^{n} \big(\widehat{f}^{(1)}_{n/2}(X_i, A_i) - f^*(X_i, A_i) \big).
\end{aligned}$$

Since the terms in the summand defining T_* are i.i.d., a straightforward computation yields

$$\mathbb{E}[T_*^2] = \frac{1}{n} \mathbb{E}\Big[\Big(\tfrac{g(X_i, A_i)}{\pi(X_i, A_i)} Y_i - \tau^* - f^*(X_i, A_i) \Big)^2 \Big] = \frac{v_{\text{semi}}^2}{n},$$

corresponding to the optimal asymptotic variance. For the cross term $\mathbb{E}[T_1 T_2]$, applying the Cauchy-Schwarz inequality yields

$$|\mathbb{E}[T_1 T_2]| \leq \sqrt{\mathbb{E}[T_1^2]} \cdot \sqrt{\mathbb{E}[T_2^2]} \leq \tfrac{1}{2n} \mathbb{E}\big[\|\widehat{\mu}_{n/2} - \mu^*\|_\omega^2 \big].$$

Consequently, in order to complete the proof, it suffices to show that

$$\mathbb{E}[T_1^2] \;=\; \mathbb{E}[T_2^2] = \tfrac{1}{2n} \mathbb{E}\big[\|\widehat{\mu}_{n/2} - \mu^*\|_\omega^2 \big], \quad \text{and} \tag{4.47a}$$

$$\mathbb{E}[T_1 T_*] = \mathbb{E}[T_2 T_*] = 0. \tag{4.47b}$$

Proof of equation (4.47a)**:** We begin by observing that $\mathbb{E}\big[T_1^2 \mid \mathcal{B}_2\big] = \frac{1}{2n}\|\widehat{f}^{(2)}_{n/2} - f^*\|^2_{\xi\times\pi}$. Now recall equations (4.6a) and (4.7a) that define f^* and $\widehat{f}^{(2)}_{n/2}$ respectively. From these definitions, we have

$$\begin{aligned}\|\widehat{f}^{(2)}_{n/2} - f^*\|^2_{\xi\times\pi} &= \mathbb{E}_{X\sim\xi}\Big[\operatorname{var}_{A\sim\pi(X,\cdot)}\Big(\frac{g(X,A)}{\pi(X,A)}\big(\widehat{\mu}^{(2)}_{n/2}(X,A) - \mu^*(X,A)\big) \mid X\Big) \mid \mathcal{B}_2\Big]\\ &\leq \mathbb{E}_{(X,A)\sim\xi\times\pi}\Big[\frac{g^2(X,A)}{\pi^2(X,A)}\big(\widehat{\mu}^{(2)}_{n/2}(X,A) - \mu^*(X,A)\big)^2 \mid \mathcal{B}_2\Big] = \|\widehat{\mu}^{(2)}_{n/2} - \mu^*\|^2_{\omega}.\end{aligned}$$

Putting together the pieces yields $\mathbb{E}[T_1^2] \leq \frac{1}{2n}\mathbb{E}[\|\widehat{\mu}^{(2)}_{n/2} - \mu^*\|^2_{\omega}]$ as claimed. A similar argument yields the same bound for $\mathbb{E}[T_2^2]$.

Proof of equation (4.47b)**:** We first decompose the term T_* into two parts:

$$T_{*,j} := \frac{1}{n}\sum_{i=n(j-1)/2+1}^{nj/2}\Big\{\frac{g(X_i,A_i)}{\pi(X_i,A_i)}Y_i - \tau^* - f^*(X_i,A_i)\Big\}, \quad \text{for } j \in \{1,2\}.$$

Since for any $x \in \mathbb{X}$, the functions $f^*(x,\cdot)$ and $\widehat{f}^{(2)}_{n/2}(x,\cdot)$ are both zero-mean under $\pi(x,\cdot)$, we have the following identity.

$$\mathbb{E}\big[T_{*,2}T_1 \mid \mathcal{B}_2\big] = \frac{1}{n}\sum_{i=1}^{n/2}\mathbb{E}\Big[T_{*,2}\cdot\mathbb{E}\big[\widehat{f}^{(2)}_{n/2}(X_i,A_i) - f^*(X_i,A_i) \mid X_i\big] \mid \mathcal{B}_2\Big] = 0.$$

Similarly, we have $\mathbb{E}\big[T_{*,1}T_2\big] = 0$. It remains to study the terms $\mathbb{E}\big[T_{*,j}T_j\big]$ for $j \in \{1,2\}$. We start with the following expansion:

$$\begin{aligned}T_{*,1}\cdot T_1 = \frac{1}{n^2}\sum_{i=1}^{n/2}&\Big\{\frac{g(X_i,A_i)}{\pi(X_i,A_i)}Y_i - \tau^* - f^*(X_i,A_i)\Big\}\cdot\big(\widehat{f}^{(2)}_{n/2}(X_i,A_i) - f^*(X_i,A_i)\big)\\ &+ \frac{1}{n^2}\sum_{1\leq i\neq\ell\leq n/2}\big\{\tfrac{g(X_i,A_i)}{\pi(X_i,A_i)}Y_i - \tau^* - f^*(X_i,A_i)\big\}\cdot\big(\widehat{f}^{(2)}_{n/2}(X_\ell,A_\ell) - f^*(X_\ell,A_\ell)\big).\end{aligned}$$

For $i \neq \ell$, by the unbiasedness of T_*, we note that:

$$\mathbb{E}\Big[\Big\{\frac{g(X_i,A_i)}{\pi(X_i,A_i)}Y_i - \tau^* - f^*(X_i,A_i)\Big\}\cdot\big(\widehat{f}^{(2)}_{n/2}(X_\ell,A_\ell) - f^*(X_\ell,A_\ell)\big) \mid \mathcal{B}_2, X_\ell\Big] = 0.$$

So we have that:

$$\begin{aligned}\mathbb{E}\big[T_{*,1}T_1\big] &= \tfrac{1}{2n}\mathbb{E}\Big[\Big\{\frac{g(X,A)}{\pi(X,A)}\mu^*(X,A) - \tau^* - f^*(X,A)\Big\}\cdot\big(\widehat{f}^{(2)}_{n/2}(X,A) - f^*(X,A)\big)\Big]\\ &= \tfrac{1}{2n}\mathbb{E}\Big[\Big(\langle g(X,\cdot),\, \mu^*(X,\cdot)\rangle - \tau^*\Big)\cdot\big(\widehat{f}^{(2)}_{n/2}(X,A) - f^*(X,A)\big)\Big]\\ &= \tfrac{1}{2n}\mathbb{E}\Big[\Big(\langle g(X,\cdot),\, \mu^*(X,\cdot)\rangle - \tau^*\Big)\cdot\mathbb{E}\big[\widehat{f}^{(2)}_{n/2}(X,A) - f^*(X,A) \mid X, \mathcal{B}_2\big]\Big] = 0.\end{aligned}$$

4.4.2 Proof of Theorem 4.2

Based on Theorem 4.1 and the discussion thereafter, it suffices to prove an oracle inequality on the squared error $\mathbb{E}\big[\|\widehat{\mu}_n - \mu^*\|_\omega^2\big]$. So as to ease the notation, for any pair of functions $f, g : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$, we define the empirical inner product

$$\langle f,\, g\rangle_m := \frac{1}{m}\sum_{i=1}^{m} \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} f(X_i, A_i) g(X_i, A_i),$$

and the induced norm $\|f\|_m := \sqrt{\langle f,\, f\rangle_m}$.

With this notation, observe that our weighted least-squares estimator is based on minimizing the objective $\|Y - \mu\|_m^2 = \frac{1}{m}\sum_{i=1}^m \frac{g^2(X_i,A_i)}{\pi^2(X_i,A_i)}\big(Y_i - \mu(X_i, A_i)\big)^2$, where we have slightly overloaded our notation on Y— viewing it as a function such that $Y(X_i, A_i) = Y_i$ for each i.

By the convexity of Ω and the optimality condition that defines $\widehat{\mu}_m$, for any function $\mu \in \mathcal{F}$ and scalar $\beta \in (0, 1)$, we have $\|Y - \mu\|_m^2 \leq \|Y_i - \big(t\mu + (1-t)\widehat{\mu}_m\big)\|_m^2$. Taking the limit $t \to 0^+$ yields the basic inequality

$$\|\widehat{\Delta}_m\|_m^2 \leq \langle \mu^* - Y,\, \widehat{\Delta}_m\rangle_m + \langle \widehat{\Delta}_m,\, \widetilde{\Delta}\rangle_m, \tag{4.48}$$

where define the estimation error $\widehat{\Delta}_m := \widehat{\mu}_m - \mu$, and the approximation error $\widetilde{\Delta} := \mu^* - \mu$. By applying the Cauchy–Schwarz inequality to the last term in equation (4.48), we find that

$$\langle \widehat{\Delta}_m,\, \widetilde{\Delta}\rangle_m \leq \|\widehat{\Delta}_m\|_m \cdot \|\widetilde{\Delta}\|_m \leq \frac{1}{2}\|\widehat{\Delta}_m\|_m^2 + \frac{1}{2}\|\widetilde{\Delta}\|_m^2.$$

Combining with inequality (4.48) yields the bound

$$\|\widehat{\Delta}_m\|_m^2 \leq \frac{2}{m}\sum_{i=1}^{m} W_i \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \widehat{\Delta}_m(X_i, A_i) + \|\widetilde{\Delta}\|_m^2, \tag{4.49}$$

where $W_i := \mu^*(X_i, A_i) - Y_i$ is the *outcome noise* associated with observation i.

The remainder of our analysis involves controlling different terms in the bound (4.49). There are two key ingredients in the argument:

- First, we need to relate the empirical $\mathbb{L}^2$-norm $\|\cdot\|_m$ with its population counterpart $\|\cdot\|_\omega$. Lemma 4.1 stated below provides this control.

- Second, using the Rademacher complexity $\mathcal{S}_m$ from equation (4.13a), we upper bound the weighted empirical average term associated with the outcome noise $W_i = \mu^*(X_i, A_i) - Y_i$ on the right-hand-side of equation (4.49). This bound is given in Lemma 4.2.

Define the event

$$\mathscr{E}_\omega := \Big\{ \|f\|_m^2 \geq \frac{\alpha_2 \alpha_1^2}{16} \|f\|_\omega^2 \quad \text{for all } f \in \mathcal{F}^* \setminus \mathbb{B}_\omega(r_m) \Big\}. \tag{4.50}$$

The following result provides tail control on the complement of this event.

Lemma 4.1. *There exists a universal constant $c' > 0$ such that*

$$\mathbb{P}(\mathscr{E}_\omega^c) \leq \exp\big(-\tfrac{\alpha_2^2}{c'} m\big). \tag{4.51}$$

See Section 4.4.2.1 for the proof.

For any (non-random) scalar $r > 0$, we also define the event

$$\mathscr{E}(r) := \big\{ \|\widehat{\Delta}_m\|_\omega \geq r \big\}.$$

On the event $\mathscr{E}_\omega \cap \mathscr{E}(r_m)$, our original bound (4.49) implies that

$$\|\widehat{\Delta}_m\|_\omega^2 \leq \frac{32}{\alpha_2 \alpha_1^2 m} \sum_{i=1}^m W_i \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \widehat{\Delta}_m(X_i, A_i) + \frac{16}{\alpha_2 \alpha_1^2} \|\widetilde{\Delta}\|_m^2. \tag{4.52}$$

In order to bound the right-hand-side of equation (4.52), we need a second lemma that controls the empirical process in terms of the critical radius s_m defined by the fixed point relation (4.14a).

Lemma 4.2. *We have*

$$\mathbb{E}\Big[\mathbf{1}_{\mathscr{E}(s_m)} \cdot \frac{2}{m} \sum_{i=1}^m W_i \tfrac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \widehat{\Delta}_m(X_i, A_i) \Big] \leq s_m \sqrt{\mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2\big]}. \tag{4.53}$$

See Section 4.4.2.2 for the proof.

With these two auxiliary lemmas in hand, we can now complete the proof of the theorem itself. In order to exploit the basic inequality (4.52), we begin by decomposing the MSE as $\mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2\big] \leq \sum_{j=1}^3 T_j$, where

$$\begin{aligned} &T_1 := \mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2 \mathbf{1}_{\mathscr{E}_\omega \cap \mathscr{E}(r_m) \cap \mathscr{E}(s_m)}\big], \\ &T_2 := \mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2 \mathbf{1}_{[\mathscr{E}(r_m) \cap \mathscr{E}(s_m)]^c}\big], \quad \text{and} \quad T_3 := \mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2 \mathbf{1}_{\mathscr{E}_\omega^c}\big]. \end{aligned}$$

We analyze each of these terms in turn.

Analysis of T_1: Combining the bound (4.52) with Lemma 4.2 yields

$$\begin{aligned} T_1 &\leq \tfrac{32}{\alpha_2 \alpha_1^2 m} \mathbb{E}\Big[\mathbf{1}_{\mathscr{E}(r_m)} \cdot \sum_{i=1}^m W_i \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \widehat{\Delta}_m(X_i, A_i) \Big] + \tfrac{16}{\alpha_2 \alpha_1^2} \mathbb{E}\big[\|\widetilde{\Delta}\|_m^2 \big] \\ &\leq \tfrac{32}{\alpha_2 \alpha_1^2} s_m \sqrt{\mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2 \big]} + \tfrac{16}{\alpha_2 \alpha_1^2} \mathbb{E}\big[\|\widetilde{\Delta}\|_m^2 \big] \\ &= \tfrac{32}{\alpha_2 \alpha_1^2} s_m \sqrt{\mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2 \big]} + \tfrac{16}{\alpha_2 \alpha_1^2} \|\widetilde{\Delta}\|_\omega^2, \end{aligned} \tag{4.54a}$$

where the final equality follows since $\mathbb{E}\big[\|\widetilde{\Delta}\|_m^2\big] = \|\widetilde{\Delta}\|_\omega^2$, using the definition of the empirical $\mathbb{L}^2$-norm, and the fact that the approximation error $\widetilde{\Delta}$ is a deterministic function.

Bounding T_2: On the event $[\mathscr{E}(r_m) \cap \mathscr{E}(s_m)]^c = \mathscr{E}^c(r_m) \cup \mathscr{E}^c(s_m)$, we are guaranteed to have $\|\widehat{\Delta}_m\|_\omega^2 \leq s_m^2 + r_m^2$, and hence

$$T_2 \leq s_m^2 + r_m^2. \tag{4.54b}$$

Analysis of T_3: Since the function class $\mathcal{F}$ is bounded, we have

$$T_3 \leq \mathrm{diam}_\omega^2(\mathcal{F} \cup \{\mu^*\}) \cdot \mathbb{P}\big(\mathscr{E}_\omega^c\big) \leq \mathrm{diam}_\omega^2(\mathcal{F} \cup \{\mu^*\}) \cdot e^{-c\alpha_2^2 m} \tag{4.54c}$$

for a universal constant $c > 0$.

Finally, substituting the bounds (4.54a), (4.54b) and (4.54c) into our previous inequality $\mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2\big] \leq \sum_{j=1}^3 T_j$ yields

$$\mathbb{E}\big[\|\widehat{\Delta}_m^2\|_\omega\big] \leq \tfrac{32 s_m}{\alpha_2 \alpha_1^2} \sqrt{\mathbb{E}[\|\widehat{\Delta}_m\|_\omega^2]} + \tfrac{16}{\alpha_2 \alpha_1^2} \|\widetilde{\Delta}\|_\omega^2 + (s_m^2 + r_m^2) + \mathrm{diam}_\omega^2(\mathcal{F} \cup \{\mu^*\}) \cdot e^{-c\alpha_2^2 m}.$$

Note that this is a self-bounding relation for the quantity $\mathbb{E}\big[\|\widehat{\Delta}_m^2\|_\omega\big]$. With the choice $m = n/2$, it implies the the MSE bound

$$\begin{aligned} \mathbb{E}\big[\|\widehat{\mu}_{n/2} - \mu^*\|_\omega^2\big] &\leq 2\mathbb{E}\big[\|\widehat{\mu}_{n/2} - \mu\|_\omega^2\big] + 2\mathbb{E}\big[\|\widehat{\Delta}_{n/2}\|_\omega^2\big] \\ &\leq \big(2 + \tfrac{2c'}{\alpha_1 \alpha_2^2}\big) \|\widetilde{\Delta}\|_\omega^2 + \tfrac{c'}{\alpha_1^2 \alpha_2^4} s_{n/2}^2 + \tfrac{c'}{\alpha_1 \alpha_2^2} r_{n/2}^2 + \mathrm{diam}_\omega^2(\mathcal{F} \cup \{\mu^*\}) \cdot e^{-c\alpha_2^2 n/2}, \end{aligned}$$

for a pair (c, c') of positive universal constants. Combining with Theorem 4.1 and taking the infimum over $\mu \in \mathcal{F}$ completes the proof.

4.4.2.1 Proof of Lemma 4.1

The following lemma provides a lower bound on the empirical norm, valid uniformly over a given function class $\mathcal{H} \subseteq \big\{ h / \|h\|_\omega \;\mid\; h \in \mathcal{F}^* \backslash \{0\} \big\}$.

Lemma 4.3. *For a failure probability* $\varepsilon \in (0,1)$, *we have*

$$\inf_{h \in \mathcal{H}} \|h\|_m^2 \geq \frac{\alpha_2 \alpha_1^2}{4} - 4\alpha_1 \mathcal{R}_m(\mathcal{H}) - c\alpha_1^2 \cdot \Big\{ \sqrt{\tfrac{\log(1/\varepsilon)}{m}} + \tfrac{\log(1/\varepsilon)}{m} \Big\} \tag{4.55}$$

with probability at least $1-\varepsilon$.

See Appendix C.1.3 for the proof of this lemma.

Taking it as given for now, we proceed with the proof of Lemma 4.1. For any deterministic radius $r > 0$, we define the set

$$\mathcal{H}_r := \Big\{ h / \|h\|_\omega \;\mid\; h \in \mathcal{F}^*, \text{ and } \|h\|_\omega \geq r \Big\}.$$

By construction, the sequence $\{\mathcal{H}_r\}_{r>0}$ consists of nested sets—that is, $\mathcal{H}_r \subseteq \mathcal{H}_s$ for $r > s$—and all are contained within the set $\big\{ h/\|h\|_\omega \mid h \in \mathcal{F} \backslash \{0\} \big\}$. By convexity of the class $\mathcal{F}$, for any $h \in \mathcal{F}$ such that $\|h\|_\omega \geq r$, we have $r\, h/\|h\|_\omega \in \mathcal{F} \cap \mathbb{B}(r)$. Consequently, we can bound the Rademacher complexity as

$$\begin{aligned}\mathcal{R}_m(\mathcal{H}_r) = \mathbb{E}\Big[\sup_{h \in \mathcal{H}_r} \sum_{i=1}^m \varepsilon_i \frac{g(X_i, A_i) h(X_i, A_i)}{\pi(X_i, A_i)} \Big] &\leq \frac{1}{r} \mathbb{E}\Big[\sup_{h \in \mathcal{F}^* \cap \mathbb{B}_\omega(r)} \sum_{i=1}^m \varepsilon_i \frac{g(X_i, A_i) h(X_i, A_i)}{\pi(X_i, A_i)} \Big] \\ &= \frac{1}{r} \mathcal{R}_m(\mathcal{F}^* \cap \mathbb{B}_\omega(r)).\end{aligned}$$

By combining this inequality with Lemma 4.3, we find that

$$\inf_{f \in \mathcal{F} \backslash \mathbb{B}_\omega(r)} \frac{\|f\|_m^2}{\|f\|_\omega^2} \geq \frac{\alpha_2 \alpha_1^2}{4} - \frac{4\alpha_1}{r} \mathcal{R}_m\big(\mathcal{F}^* \cap \mathbb{B}_\omega(r)\big) - c\alpha_1^2 \cdot \Big\{ \sqrt{\frac{\log(1/\varepsilon)}{m}} + \frac{\log(1/\varepsilon)}{m} \Big\} \tag{4.56}$$

with probability at least $1 - \varepsilon$. This inequality is valid for any deterministic radius $r > 0$.

By the definition (4.14b) of the critical radius r_m, inequality (4.14b) holds for any $r > r_m$. We now set $r = r_m$ in equation (4.56). Doing so allows us to conclude that given a sample size satisfying $m \geq \frac{1024 c^2}{\alpha_2^2} \log(1/\varepsilon)$, we have

$$\frac{4\alpha_1}{r_m} \mathcal{R}_m\big(\mathcal{F}^* \cap \mathbb{B}_\omega(r_m)\big) \leq \frac{\alpha_2 \alpha_1^2}{16}, \quad \text{and} \quad c\alpha_1^2 \cdot \Big\{ \sqrt{\frac{\log(1/\varepsilon)}{m}} + \frac{\log(1/\varepsilon)}{m} \Big\} \leq \frac{\alpha_2 \alpha_1^2}{16}.$$

Combining with equation (4.56) completes the proof of Lemma 4.1.

4.4.2.2 Proof of Lemma 4.2

Recall our notation $W_i := \mu^*(X_i, A_i) - Y_i$ for the outcome noise. Since the set Ω is convex, on the event $\mathscr{E}(s_m)$, we have

$$\tfrac{1}{\|\widehat{\Delta}_m\|_\omega} \sum_{i=1}^m W_i \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \widehat{\Delta}_m(X_i, A_i) \leq \tfrac{1}{s_m} \sup_{h \in \mathcal{F}^* \cap \mathbb{B}_\omega(s_m)} \sum_{i=1}^m W_i \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} h(X_i, A_i). \tag{4.57}$$

Define the empirical process supremum

$$Z_m(s_m) := \sup_{h \in \mathcal{F}^* \cap \mathbb{B}_\omega(s_m)} \frac{1}{m} \sum_{i=1}^m W_i \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} h(X_i, A_i).$$

Since the all-zeros function 0 is an element of $\mathcal{F}^* \cap \mathbb{B}_\omega(s_m)$, we have $Z_m(s_m) \geq 0$. Equation (4.57) implies that

$$\mathbb{E}\Big[\mathbf{1}_{\mathscr{E}(s_m)} \cdot \tfrac{2}{m} \sum_{i=1}^m \tfrac{g^2(X_i,A_i)}{\pi^2(X_i,A_i)} \big(\mu^*(X_i, A_i) - Y_i\big) \widehat{\Delta}_m(X_i, A_i)\Big]$$
$$\leq \mathbb{E}\Big[\frac{\|\widehat{\Delta}_m\|_\omega}{s_m} Z_m(s_m)\Big] \leq \sqrt{\mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2\big]} \cdot \sqrt{s_m^{-2}\mathbb{E}\big[Z_m^2(s_m)\big]}, \quad (4.58)$$

where the last step follows by applying the Cauchy–Schwarz inequality.

Define the symmetrized random variable

$$Z'_m(s_m) := \sup_{h \in \mathcal{F}^* \cap \mathbb{B}_\omega(s_m)} \frac{1}{m} \sum_{i=1}^m \varepsilon_i \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \big(\mu^*(X_i, A_i) - Y_i\big) h(X_i, A_i),$$

where $\{\varepsilon_i\}_{i=1}^m$ is an i.i.d. sequence of Rademacher variables, independent of the data. By a standard symmetrization argument (e.g., §2.4.1 in the book [213]), there are universal constants (c, c') such that

$$\mathbb{P}\big[Z_m(s_m) > t\big] \leq c'\mathbb{P}\big[Z'_m(s_m) > ct\big], \quad \text{for any } t > 0.$$

Integrating over t yields the bound

$$\mathbb{E}[Z_m^2(s_m)] \leq c^2 c' \mathbb{E}[Z'^2_m(s_m)] = c^2 c' \mathcal{S}_m^2(s_m) \overset{(i)}{=} c^2\, c' s_m^2,$$

where equality (i) follows from the definition of s_m. Substituting this bound back into equation (4.58) completes the proof of Lemma 4.2.

4.4.3 Proof of Corollary 4.5

Define the function $\bar{f}(x, a) := \frac{g(x,a)}{\pi(x,a)} \bar{\mu}(x, a) - \langle g(x, \cdot),\, \bar{\mu}(x, \cdot)\rangle_\lambda$, which would be optimal if $\bar{\mu}$ were the true treatment function. It induces the estimate

$$\widehat{\tau}_{n,\bar{f}} = \frac{1}{n} \sum_{i=1}^n \Big\{ \frac{g(X_i, A_i)}{\pi(X_i, A_i)} \big(Y_i - \bar{\mu}(X_i, A_i)\big) + \langle g(X_i, \cdot),\, \bar{\mu}(X_i, \cdot)\rangle_\lambda \Big\},$$

which has (n-rescaled) variance $n \cdot \mathbb{E}\Big[\big|\widehat{\tau}_{n,\bar{f}} - \tau^*\big|^2\Big] = v_{\text{semi}}^2 + v^2(\bar{\mu})$, where v_{semi}^2 is the efficient variance, and $v^2(\bar{\mu}) := \text{var}\Big(\frac{g(X,A)}{\pi(X,A)} \cdot \big(\mu^* - \bar{\mu}\big)(X, A)\Big)$. Let us now compare two-stage estimator $\widehat{\tau}_n$ with this idealized estimator. We have

$$\begin{aligned}\mathbb{E}[|\widehat{\tau}_{n,f} - \widehat{\tau}_n|^2] &\leq \frac{2}{n^2}\mathbb{E}\Big[\big|\sum_{i=1}^{n/2}(\bar{f} - \widehat{f}^{(2)}_{n/2})(X_i, A_i)\big|^2\Big] + \frac{2}{n}\mathbb{E}\Big[\big|\sum_{i=n/2+1}^{n}(\bar{f} - \widehat{f}^{(1)}_{n/2})(X_i, A_i)\big|^2\Big]\\ &\leq \frac{4}{n}\mathbb{E}\big[\|\widehat{\mu}_{n/2} - \bar{\mu}\|_\omega^2\big].\end{aligned}$$

Thus, we are guaranteed the Wasserstein bound

$$\mathcal{W}_1\big(\sqrt{n}\widehat{\tau}_n, \sqrt{n}\widehat{\tau}_{n,\bar{f}}\big) \leq \mathcal{W}_2\big(\sqrt{n}\widehat{\tau}_n, \sqrt{n}\widehat{\tau}_{n,\bar{f}}\big) \leq 2\sqrt{\mathbb{E}\big[\|\widehat{\mu}_{n/2} - \bar{\mu}\|_\omega^2\big]}.$$

Consequently, by the triangle inequality for the Wasserstein distance, it suffices to establish a normal approximation guarantee for the idealized estimator $\widehat{\tau}_{n,\bar{f}}$, along with control on the error induced by approximating the function $\bar{\mu}$ using an empirical estimator.

Normal approximation for $\widehat{\tau}_{n,\bar{f}}$: We make use of the following non-asymptotic central limit theorem:

Proposition 4.2 ([183], Theorem 3.2 (restated))**.** *Given* i.i.d. *zero-mean random variables $\{X_i\}_{i=1}^n$ with finite fourth moment, the rescaled sum $W_n := \sum_{i=1}^n X_i/\sqrt{n}$ satisfies the Wasserstein bound*

$$\mathcal{W}_1\big(W_n, Z\big) \leq \tfrac{1}{\sqrt{n}}\Big\{\tfrac{\mathbb{E}[|X_1|^3]}{\mathbb{E}[X_1^2]} + \sqrt{\tfrac{2\mathbb{E}[X_1^4]}{\pi\mathbb{E}[X_1^2]}}\Big\} \qquad \textit{where } Z \sim \mathcal{N}(0, \mathbb{E}[X_1^2]).$$

Since we have $\mathbb{E}[|X_1|^3] \leq \sqrt{\mathbb{E}[X_1^2]\cdot\mathbb{E}[X_1^4]}$, this bound implies that $\mathcal{W}_1\big(W_n, Z\big) \leq \frac{2}{\sqrt{n}}\cdot\sqrt{\mathbb{E}[X_1^4]/\mathbb{E}[X_1^2]}$. Applying this bound to the empirical average $\widehat{\tau}_{n,\bar{f}}$ yields

$$\mathcal{W}_1\big(\sqrt{n}\widehat{\tau}_{n,\bar{f}}, \mathcal{Z}\big) \leq \tfrac{2}{\sqrt{n}}\cdot\sqrt{\tfrac{M_4}{v^2_{\mathrm{semi}} + v^2(\bar{\mu})}},$$

as claimed.

Bounds on the estimation error $\|\widehat{\mu}_{n/2} - \bar{\mu}\|_\omega$: From the proof of Theorem 4.2, recall the basic inequality (4.48)—viz.

$$\|\widehat{\Delta}_m\|_m^2 \leq \frac{1}{m}\sum_{i=1}^m W_i\tfrac{g^2(X_i,A_i)}{\pi^2(X_i,A_i)}\widehat{\Delta}_m(X_i, A_i) + \langle\widehat{\Delta}_m, \widetilde{\Delta}\rangle_m, \tag{4.59}$$

where $W_i = \mu^*(X_i, A_i) - Y_i$ is the outcome noise.

As before, we define the approximation error $\widetilde{\Delta} := \mu^* - \bar{\mu}$. Since $\bar{\mu} = \arg\min_{h\in\mathcal{F}}\|h - \mu^*\|_\omega$ is the projection of μ^* onto $\mathcal{F}$, and $\widehat{\mu}_m \in \mathcal{F}$ is feasible for this optimization problem, the first-order optimality condition implies that $\langle \widehat{\Delta}_m,\, \widetilde{\Delta} \rangle_\omega \leq 0$. By adding this inequality to our earlier bound (4.59) and re-arranging terms, we find that

$$\|\widehat{\Delta}_m\|_m^2 \leq \frac{1}{m} \sum_{i=1}^m \tfrac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \big(\mu^*(X_i, A_i) - Y_i \big) \widehat{\Delta}_m(X_i, A_i) + \big(\langle \widehat{\Delta}_m,\, \widetilde{\Delta} \rangle_m - \langle \widehat{\Delta}_m,\, \widetilde{\Delta} \rangle_\omega \big). \tag{4.60}$$

Now define the empirical process suprema

$$Z_m(r) := \sup_{h \in \mathcal{F}^* \cap \mathbb{B}_\omega(r)} \frac{1}{m} \sum_{i=1}^m \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} \big(\mu^*(X_i, A_i) - Y_i \big) h(X_i, A_i), \quad \text{and}$$

$$Z_m'(s) := \sup_{h \in \mathcal{F}^* \cap \mathbb{B}_\omega(s)} \frac{1}{m} \sum_{i=1}^m \Big(\frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} h(X_i, A_i) \widetilde{\Delta}(X_i, A_i) - \langle \widetilde{\Delta},\, h \rangle_\omega \Big).$$

From the proof of Theorem 4.2, recall the events

$$\mathscr{E}_\omega := \Big\{ \|f\|_m^2 \geq \tfrac{\alpha_2 \alpha_1^2}{16} \|f\|_\omega^2, \quad \text{for any } f \in \mathcal{F}^* \setminus \mathbb{B}_\omega(r_m) \Big\}, \quad \text{and} \quad \mathscr{E}(r) := \big\{ \|\widehat{\Delta}_m\|_\omega \geq r \big\}.$$

Introduce the shorthand $u_m = \max\{r_m, s_m, d_m\}$. On the event $\mathscr{E}_\omega \cap \mathscr{E}(u_m)$, the basic inequality (4.60) implies that

$$\begin{aligned}\frac{\alpha_2 \alpha_1^2}{16} \|\widehat{\Delta}_m\|_\omega^2 \leq \|\widehat{\Delta}_m\|_m^2 &\overset{(i)}{\leq} Z_m(\|\widehat{\Delta}_m\|_\omega) + Z_m'(\|\widehat{\Delta}_m\|_\omega) \\ &\overset{(ii)}{\leq} \frac{\|\widehat{\Delta}_m\|_\omega}{r_m} Z_m(r_m) + \frac{\|\widehat{\Delta}_m\|_\omega}{s_m} Z_m'(s_m),\end{aligned}$$

where step (ii) follows from the non-increasing property of the functions $r \mapsto r^{-1} Z_m(r)$ and $s \mapsto s^{-1} Z_m'(s)$.

So there exists a universal constant $c > 0$ such that

$$\mathbb{E}\Big[\|\widehat{\Delta}_m\|_\omega^2 \mathbf{1}_{\mathscr{E}_\omega \cap \mathscr{E}(u_m)} \Big] \leq \frac{c}{\alpha_2^2 \alpha_1^4} \Big\{ \frac{1}{s_m^2} \mathbb{E}\big[Z_m^2(s_m) \big] + \frac{1}{d_m^2} \mathbb{E}\big[\big\{ Z_m'(d_m) \big\}^2 \big] \Big\}.$$

Via the same symmetrization argument as used in the proof of Theorem 4.2, there exists a universal constant $c > 0$ such that

$$\mathbb{E}[Z_m^2(s_m)] \leq c \mathcal{S}_m^2 \big(\mathcal{F}^* \cap \mathbb{B}_\omega(s_m) \big), \quad \text{and} \quad \mathbb{E}\Big[\big(Z_m'(d_m) \big)^2 \Big] \leq c \mathcal{D}_m^2 \big(\mathcal{F}^* \cap \mathbb{B}_\omega(d_m) \big).$$

By the definition of the critical radius s_m, we have

$$\frac{1}{s_m} \mathcal{S}_m \big(\mathcal{F}^* \cap \mathbb{B}_\omega(s_m) \big) = s_m, \quad \text{and} \quad \frac{1}{d_m} \mathcal{D}_m \big(\mathcal{F}^* \cap \mathbb{B}_\omega(d_m) \big) = d_m.$$

Combining with the moment bound above, we arrive at the conclusion:

$$\begin{aligned}\mathbb{E}[\|\widehat{\Delta}_m\|_\omega^2] &\leq \mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2 \mathbf{1}_{\mathscr{E}_\omega \cap \mathscr{E}(u_m)} \big] + \mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2 \mathbf{1}_{\mathscr{E}(u_m)^c} \big] + \mathbb{E}\big[\|\widehat{\Delta}_m\|_\omega^2 \mathbf{1}_{\mathscr{E}_\omega^c} \big] \\ &\leq \Big(1 + \frac{c}{\alpha_2^4 \alpha_1^2} \Big) \cdot \big(r_m^2 + s_m^2 + d_m^2 \big) + \mathrm{diam}_\omega^2(\mathcal{F} \cup \{\mu^*\}) \cdot \mathbb{P}(\mathscr{E}_\omega^c) \\ &\leq \Big(1 + \frac{c}{\alpha_2^4 \alpha_1^2} \Big) \cdot \big(r_m^2 + s_m^2 + d_m^2 \big) + \mathrm{diam}_\omega^2(\mathcal{F}) \cdot e^{-c \alpha_2^2 m}.\end{aligned}$$

Substituting into the Wasserstein distance bound completes the proof of Corollary 4.5.

4.5 Proofs of minimax lower bounds

In this section, we prove the two minimax lower bounds—namely, Theorem 4.3 and Proposition 4.1.

4.5.1 Proof of Theorem 4.3

It suffices to show that the minimax risk $\mathscr{M}_n \equiv \mathscr{M}_n\big(\mathbb{C}_\delta(\mathcal{I}^*)\big)$ satisfies the following three lower bounds:

$$\mathscr{M}_n \geq \frac{c}{n} \operatorname{var}_{\xi^*}\left(\langle g(X, \cdot),\, \mu^*(X, \cdot)\rangle_\lambda\right) \quad \text{for } n \geq 4(M'_{2\to 4})^2, \tag{4.61a}$$

$$\mathscr{M}_n \geq \frac{c}{n}\|\sigma\|_\omega^2 \quad \text{for } n \geq 16, \tag{4.61b}$$

$$\mathscr{M}_n \geq \frac{c}{n}\|\delta\|_\omega^2 \quad \text{for } n \in \big[M_{2\to 4}^2, c'|\mathbb{X}|/c_{\max}\big]. \tag{4.61c}$$

Given these three inequalities, the minimax risk $\mathscr{M}_n$ can be lower bounded by the average of the right-hand side quantities, assuming that n is sufficiently large. Since c is a universal constant, these bounds lead to the conclusion of Theorem 4.3.

Throughout the proof, we use $\mathbb{P}_{\mu^*,\xi}$ to denote the law of a sample (X, A, Y) under the problem instance defined by outcome function μ^* and data distribution ξ. We further use $\mathbb{P}^{\otimes n}_{\mu^*,\xi}$ to denote its n-fold product, as is appropriate given our i.i.d. data $(X_i, A_i, Y_i)_{i=1}^n$.

4.5.1.1 Proof of the lower bound (4.61a)

The proof is based on Le Cam's two-point method: we construct a family of probability distributions $\{\xi_s \mid s > 0\}$, each contained in the local neighborhood $\mathfrak{N}^{\mathrm{prob}}(\xi^*)$. We choose the parameter s small enough to ensure that the probability distributions $\mathbb{P}^{\otimes n}_{\xi_s,\mu^*}$ and $\mathbb{P}^{\otimes n}_{\xi^*,\mu^*}$ are "indistinguishable", but large enough to ensure that the functional values $\tau(\xi_s, \mu^*)$ and $\tau(\xi^*, \mu^*)$ are well-separated. See §15.2.1–15.2.2 in the book [213] for more background.

More precisely, Le Cam's two-point lemma guarantees that for any distribution $\xi_s \in \mathfrak{N}^{\mathrm{prob}}(\xi^*)$, the minimax risk is lower bounded as

$$\mathscr{M}_n \geq \frac{1}{4}\Big\{1 - d_{\mathrm{TV}}\Big(\mathbb{P}^{\otimes n}_{\mu^*,\xi_s}, \mathbb{P}^{\otimes n}_{\mu^*,\xi^*}\Big)\Big\} \cdot \big\{\tau(\xi_s, \mu^*) - \tau(\xi^*, \mu^*)\big\}^2, \tag{4.62}$$

Recall that throughout this section, we work with the sample size lower bound

$$n \geq 4(M'_{2\to 4})^2. \tag{4.63}$$

Now suppose that under the condition (4.63), we can exhibit a choice of s within the family $\{\xi_s \mid s > 0\}$ such that the functional gap satisfies the lower bound

$$\tau(\xi_s, \mu^*) - \tau(\xi^*, \mu^*) \geq \frac{1}{16\sqrt{n}}\sqrt{\operatorname{var}\left(\langle g(X, \cdot),\, \mu^*(X, \cdot)\rangle_\lambda\right)}, \tag{4.64a}$$

whereas the TV distance satisfies the upper bound

$$d_{\mathrm{TV}}\Big(\mathbb{P}^{\otimes n}_{\mu^*,\xi_s}, \mathbb{P}^{\otimes n}_{\mu^*,\xi^*}\Big) \leq \frac{1}{3}. \tag{4.64b}$$

These two inequalities, in conjunction with Le Cam's two point bound (4.62), imply the claimed lower bound (4.61a).

With this overview in place, it remains to define the family $\{\xi_s \mid s > 0\}$, and prove the bounds (4.64a) and (4.64b).

Family of perturbations: Define the real-valued function

$$h(x) := \langle \mu^*(x, \cdot),\, g(x, \cdot)\rangle_\lambda - \mathbb{E}_{\xi^*}\big[\langle \mu^*(X, \cdot),\, g(X, \cdot)\rangle_\lambda\big],$$

along with its truncated version

$$h_{tr}(x) := \begin{cases} h(x) & \text{if } |h(x)| \leq 2M'_{2\to 4} \cdot \sqrt{\mathbb{E}_{\xi^*}[h^2(X)]}, \text{ and} \\ \operatorname{sgn}(h(x)) \cdot \sqrt{\mathbb{E}_{\xi^*}[h^2(X)]} & \text{otherwise.} \end{cases}$$

For each $s > 0$, we define the *tilted probability measure*

$$\xi_s(x) := Z_s^{-1} \xi^*(x) \exp\big(s h_{tr}(x)\big), \quad \text{where } Z_s = \textstyle\sum_{x \in \mathbb{X}} \xi^*(x) \exp\big(s h_{tr}(x)\big).$$

It can be seen that the tilted measure satisfies the bounds

$$\exp\big(-s\|h_{tr}\|_\infty\big) \leq \frac{\xi_s(x)}{\xi^*(x)} \leq \exp\big(s\|h_{tr}\|_\infty\big) \qquad \text{for any } x \in \mathbb{X},$$

whereas the normalization constant is sandwiched as

$$\exp\big(-s\|h_{tr}\|_\infty\big) \leq Z_s \leq \exp\big(s\|h_{tr}\|_\infty\big).$$

Throughout this section, we choose

$$s := (4\|h_{tr}\|_{\mathbb{L}^2(\xi^*)}\sqrt{n})^{-1}, \tag{4.65a}$$

which ensures that

$$s\|h_{tr}\|_\infty = \frac{1}{4\sqrt{n}} \cdot \frac{\|h_{tr}\|_\infty}{\|h_{tr}\|_{\mathbb{L}^2(\xi^*)}} \overset{(i)}{\leq} \frac{1}{\sqrt{8n}} \frac{2M'_{2\to 4}\|h\|_{\mathbb{L}^2(\xi^*)}}{\|h\|_{\mathbb{L}^2(\xi^*)}} \overset{(ii)}{\leq} \frac{1}{8}, \tag{4.65b}$$

where step (i) follows from the definition of the truncated function h_{tr}, and step (ii) follows from the sample size condition (4.63).

Proof of the lower bound (4.64a)**:** First we lower bound the gap in the functional. We have

$$\begin{aligned}\tau(\xi_s, \mu^*) - \tau(\xi^*, \mu^*) &= \mathbb{E}_{\xi_s}[\langle \mu^*(X, \cdot),\, g(X, \cdot)\rangle_\lambda] - \mathbb{E}_{\xi^*}[\langle \mu^*(X, \cdot),\, g(X, \cdot)\rangle_\lambda] \\ &= \mathbb{E}_{\xi^*}\Big[h(X) e^{s h_{tr}(X)}\Big] \;/\; \mathbb{E}_{\xi^*}\Big[e^{s h_{tr}(X)}\Big]. \end{aligned} \tag{4.66}$$

Note that $|s h_{tr}(X)| \leq 1/8$ almost surely by construction. Using the elementary inequality $|e^z - 1 - z| \leq z^2$, valid for all $z \in [-1/4, 1/4]$, we obtain the lower bound

$$\mathbb{E}_{\xi^*}\Big[h(X) e^{s h_{tr}(X)}\Big] \geq \mathbb{E}_{\xi^*}\big[h(X)\big] + s\mathbb{E}_{\xi^*}\big[h(X) h_{tr}(X)\big] - s^2 \mathbb{E}_{\xi^*}\big[|h(X)| \cdot |h_{tr}(X)|^2\big] \tag{4.67}$$

Now we study the three terms on the right-hand-side of equation (4.67). By definition, we have $\mathbb{E}_{\xi^*}[h(X)] = 0$. Since the quantities $h(X)$ and $h_{tr}(X)$ have the same sign almost surely, the second term admits a lower bound

$$\mathbb{E}_{\xi^*}\big[h(X) h_{tr}(X)\big] \geq \mathbb{E}\big[h_{tr}^2(X)\big] \geq \frac{1}{2}\mathbb{E}\big[h^2(X)\big],$$

where the last step follows from Lemma C.2.

Focusing on the third term in the decomposition (4.67), we note that Cauchy-Schwarz inequality yields

$$\mathbb{E}_{\xi^*}\big[|h(X)| \cdot |h_{tr}(X)|^2\big] \leq \sqrt{\mathbb{E}\big[h^2(X)\big]} \cdot \sqrt{\mathbb{E}\big[h^4(X)\big]} \leq \sqrt{M'_{2\to4}} \cdot \Big\{\mathbb{E}\big[h^2(X)\big]\Big\}^{3/2},$$

where the last step follows from the definition of the constant $M'_{2\to4}$.

Combining these bounds with equation (4.67) and substituting the choice (4.65a) of the parameter s, we obtain the following lower bound on the functional gap

$$\begin{aligned}\mathbb{E}_{\xi^*}\Big[h(X) e^{s h_{tr}(X)}\Big] &\geq s\|h\|^2_{\mathbb{L}^2(\xi^*)} - s^2\sqrt{M'_{2\to4}}\|h\|^3_{\mathbb{L}^2(\xi^*)} \\ &\geq \frac{1}{8\sqrt{n}}\|h\|_{\mathbb{L}^2(\xi^*)} - \frac{\sqrt{M'_{2\to4}}}{16n}\|h\|_{\mathbb{L}^2(\xi^*)} \\ &\geq \frac{3}{32\sqrt{n}}\|h\|_{\mathbb{L}^2(\xi^*)},\end{aligned}$$

where the last step follows because $n \geq 4(M'_{2\to4})^2$.

On the other hand, since $|s h_{tr}(X)| \leq 1/8$ almost surely, we have $\mathbb{E}_{\xi^*}\big[e^{s h_{tr}(X)}\big] \leq 3/2$. Combining with the bound above and substituting into the expression (4.66), we find that we find that

$$\tau(\xi_s, \mu^*) - \tau(\xi^*, \mu^*) \geq \frac{3}{32\sqrt{n}}\|h\|_{\mathbb{L}^2(\xi^*)} \;/\; \mathbb{E}_{\xi^*}\big[e^{s h_{tr}(X)}\big] \geq \frac{1}{16\sqrt{n}}\|h\|_{\mathbb{L}^2(\xi^*)},$$

which is equivalent to the claim (4.64a).

Proof of the upper bound (4.64b)**:** Pinsker's inequality ensures that

$$d_{\mathrm{TV}}\Big(\mathbb{P}^{\otimes n}_{\mu^*,\xi_s}, \mathbb{P}^{\otimes n}_{\mu^*,\xi^*}\Big) \leq \sqrt{\frac{1}{2}\chi^2\left(\mathbb{P}^{\otimes n}_{\mu^*,\xi_s} \mid\mid \mathbb{P}^{\otimes n}_{\mu^*,\xi^*}\right)}, \tag{4.68}$$

so that it suffices to bound the χ^2-divergence. Beginning with the divergence between ξ_s and ξ^* (i.e., without the tensorization over n), we have

$$\begin{aligned}\chi^2\left(\xi_s \mid\mid \xi^*\right) = \mathrm{var}_{\xi^*}\Big(\xi_s(X)/\xi^*(X)\Big) &= \frac{1}{Z_s^2}\ \mathrm{var}_{\xi^*}\left(e^{sh_{tr}(X)} - 1\right)\\ &\leq \exp\big(2s\|h_{tr}\|_\infty\big)\cdot \mathbb{E}_{\xi^*}\big[|e^{sh_{tr}(X)} - 1|^2\big]\\ &\leq \exp\big(4s\|h_{tr}\|_\infty\big)\cdot s^2\mathbb{E}_{\xi^*}\big[h^2_{tr}(X)\big]. \end{aligned} \tag{4.69}$$

where the last step follows from the elementary inequality $|e^x - 1| \leq e^{|x|}\cdot|x|$, valid for any $x \in \mathbb{R}$. Given the choice of tweaking parameter s, we have $\exp\big(4s\|h_{tr}\|_\infty\big) \leq 2$.

The definition of the truncated function h_{tr} implies that $\mathbb{E}_{\xi^*}\big[h^2_{tr}(X)\big] \leq \mathbb{E}_{\xi^*}[h^2(X)]$. Combining this bound with our earlier inequality (4.69) yields

$$\chi^2\left(\xi_s \mid\mid \xi^*\right) \leq 2s^2\mathbb{E}_{\xi^*}\big[h^2_{tr}(X)\big] \leq \frac{1}{8n},$$

which certifies that $\xi_s \in \mathfrak{N}^{prob}(\xi^*)$, as required for the validity of our construction.

Finally, by the tensorization property of the χ^2-divergence, we have

$$\chi^2\left(\mathbb{P}^{\otimes n}_{\mu^*,\xi_s} \mid\mid \mathbb{P}^{\otimes n}_{\mu^*,\xi^*}\right) \leq \Big(1 + \frac{1}{8n}\Big)^n - 1 \ \leq\ \tfrac{3}{20}.$$

Combining with our earlier statement (4.68) of Pinsker's inequality completes the proof of the upper bound (4.64b).

4.5.1.2 Proof of equation (4.61b)

The proof is also based on Le Cam's two-point method. Complementary to equation (4.61a), we take the source distribution ξ^* to be fixed, and perturb the outcome function μ^*. Given a pair $\mu_{(s)}, \mu_{(-s)}$ of outcome functions in the local neighborhood $\mathfrak{N}^{val}_\delta$, Le Cam's two-point lemma implies

$$\mathscr{M}_n \geq \frac{1}{4}\Big\{1 - d_{\mathrm{TV}}\Big(\mathbb{P}^{\otimes n}_{\mu_{(s)},\xi^*}, \mathbb{P}^{\otimes n}_{\mu_{(-s)},\xi^*}\Big)\Big\}\cdot\big\{\tau(\xi^*, \mu_{(s)}) - \tau(\xi^*, \mu_{(-s)})\big\}^2, \tag{4.70}$$

With this set-up, our proof is based on constructing a pair $\big(\mu_{(s)}, \mu_{(-s)}\big)$ of outcome functions within the neighborhood $\mathfrak{N}^{val}_\delta(\mu^*)$ such that

$$\tau(\xi^*, \mu_{(s)}) - \tau(\xi^*, \mu_{(-s)}) \geq \frac{1}{2\sqrt{n}}\|\sigma\|_\omega, \quad \text{and} \tag{4.71a}$$

$$d_{\mathrm{TV}}\Big(\mathbb{P}^{\otimes n}_{\mu_{(s)},\xi^*}, \mathbb{P}^{\otimes n}_{\mu_{(-s)},\xi^*}\Big) \leq \frac{1}{3}. \tag{4.71b}$$

Construction of problem instances: Consider the noisy Gaussian observation model

$$Y_i \mid X_i, A_i \sim N\Big(\mu^*(X_i, A_i), \sigma^2(X_i, A_i)\Big) \qquad \text{for } i = 1, 2, \ldots, n. \tag{4.72}$$

We construct a pair of problem instances as follows: for any $s > 0$, define the functions

$$\mu^*_{(s)}(x, a) = \mu^*(x, a) + s\tfrac{g(x,a)}{\pi(x,a)}\sigma^2(x, a), \quad \text{and} \quad \mu^*_{(-s)}(x, a) = \mu^*(x, a) - s\tfrac{g(x,a)}{\pi(x,a)}\sigma^2(x, a)$$

for any $(x, a) \in \mathbb{X} \times \mathbb{A}$.

Throughout this section, we make the choice $s := \frac{1}{4\|\sigma\|_\omega \sqrt{n}}$. Under such choice, the compatibility condition (4.34) ensures that

$$|\mu_{(zs)}(x, a) - \mu^*(x, a)| = s\frac{g(x, a)}{\pi(x, a)}\sigma^2(x, a) \leq \delta(x, a),$$

for any $(x, a) \in \mathbb{X} \times \mathbb{A}$ and $z \in \{-1, 1\}$

This ensures that both $\mu_{(s)}$ and $\mu_{(-s)}$ belong to the neighborhood $\mathfrak{N}^{val}_\delta(\mu^*)$. It remains to prove the two bounds required for Le Cam's two-point arguments.

Proof of equation (4.71a): For the target linear functional under our construction, we note that

$$\tau(\xi^*, \mu_{(s)}) - \tau(\xi^*, \mu_{(-s)}) = 2s\mathbb{E}_{\xi^*}\Big[\langle \frac{g(X, \cdot)}{\pi(X, \cdot)}\sigma^2(X, \cdot),\, g(X, \cdot)\rangle_\lambda\Big] = 2s\|\sigma\|^2_\omega = \frac{1}{2\sqrt{n}}\|\sigma\|_\omega,$$

which establishes the bound (4.71a).

Proof of the bound (4.71b): It order to bound the total variation distance, we study the KL divergence between the product distributions $\mathbb{P}^{\otimes n}_{\mu_{(zs)}, \xi^*}$ for $z \in \{-1, 1\}$. Indeed, we have

$$\begin{aligned} D_{\mathrm{KL}}\left(\mathbb{P}^{\otimes n}_{\mu_{(s)}, \xi^*} \parallel \mathbb{P}^{\otimes n}_{\mu_{(-s)}, \xi^*}\right) &\overset{(i)}{=} nD_{\mathrm{KL}}\left(\mathbb{P}_{\mu_{(s)}, \xi^*} \parallel \mathbb{P}_{\mu_{(-s)}, \xi^*}\right) \\ &\overset{(ii)}{\leq} n\mathbb{E}\Big[D_{\mathrm{KL}}\left(\mathcal{L}(Y \mid X, A)|_{\mu_{(s)}} \parallel \mathcal{L}(Y \mid X, A)|_{\mu_{(-s)}}\right)\Big], \end{aligned} \tag{4.73}$$

where in step (i), we use the tensorization property of KL divergence, and in step (ii), we use convexity of KL divergence. The expectation is taken with respect to $X \sim \xi^*$ and $A \sim \pi(X, \cdot)$.

Noting that the conditional law $\mathcal{L}(Y \mid X, A)|_{\mu_{(zs)}}$ is Gaussian under both problem instances, we have

$$D_{\mathrm{KL}}\left(\mathcal{L}(Y \mid x, a)|_{\mu_{(s)}} \parallel \mathcal{L}(Y \mid x, a)|_{\mu_{(-s)}}\right) = \frac{4s^2 g^2(x, a)}{\pi^2(x, a)}\sigma^2(x, a).$$

Substituting into equation (4.73), we find that $D_{\mathrm{KL}}\left(\mathbb{P}^{\otimes n}_{\mu_{(s)},\xi^*} \parallel \mathbb{P}^{\otimes n}_{\mu_{(-s)},\xi^*}\right) \leq 4ns^2\|\sigma\|_\omega^2$. For a sample size $n \geq 16$, with the choice of the perturbation parameter $s = \frac{1}{4\sqrt{n}\|\sigma\|_\omega}$, an application of Pinsker's inequality leads to the bound

$$d_{\mathrm{TV}}\left(\mathbb{P}^{\otimes n}_{\mu_{(s)},\xi^*}, \mathbb{P}^{\otimes n}_{\mu_{(-s)},\xi^*}\right) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}\left(\mathbb{P}^{\otimes n}_{\mu^s,\xi^*} \parallel \mathbb{P}^{\otimes n}_{\mu^{(-s)},\xi^*}\right)} \leq \frac{1}{2\sqrt{2}}, \tag{4.74}$$

which completes the proof of equation (4.71b).

4.5.1.3 Proof of equation (4.61c)

The proof is based on Le Cam's mixture-vs-mixture method (cf. Lemma 15.9, [213]). We construct a pair $(\mathbb{Q}_1, \mathbb{Q}_{-1})$ of probability distributions supported on the neighborhood $\mathfrak{N}^{val}_\delta(\mu^*)$; these are used to define two mixture distributions with the following properties:

- The mixture distributions have TV distance bounded as
$$d_{\mathrm{TV}}\left(\int \mathbb{P}^{\otimes n}_{\mu,\xi^*} d\mathbb{Q}^*_1(\mu), \int \mathbb{P}^{\otimes n}_{\mu,\xi^*} d\mathbb{Q}^*_{-1}(\mu)\right) \leq \frac{1}{4}. \tag{4.75}$$
See Lemma 4.4 for details.

- There is a large gap in the target linear functional when evaluated at functions in the support of $\mathbb{Q}^*_1$ and $\mathbb{Q}^*_{-1}$. See Lemma 4.5 for details.

For any binary function $\zeta : \mathbb{X} \times \mathbb{A} \rightarrow \{-1, 1\}$, we define the perturbed outcome function

$$\mu_\zeta(x, a) := \mu^*(x, a) + \zeta(x, a) \cdot \delta(x, a) \quad \text{for all } (x, a) \in \mathbb{X} \times \mathbb{A}.$$

By construction, we have $\mu_\zeta \in \mathfrak{N}^{val}_\delta(\mu^*)$ for any binary function ζ. Now consider the function

$$\rho(x, a) := \begin{cases} \dfrac{g(x,a)\delta(x,a)}{\|\delta\|_\omega \pi(x,a)} & \dfrac{|g(x,a)|\delta(x,a)}{\pi(x,a)} \leq 2M_{2\rightarrow 4}\|\delta\|_\omega \\ \mathrm{sgn}\big(g(x,a)\big), & \text{otherwise.} \end{cases}$$

It can be seen that $\mathbb{E}[\rho^2(X, A)] \leq 1$ where the expectation is taken over a pair $X \sim \xi^*$ and $A \sim \pi(X, \cdot)$.

For a scalar $s \in \big(0, \frac{1}{2M_{2\rightarrow 4}}\big]$ and a sign variable $z \in \{-1, 1\}$, we define the probability distribution

$$\mathbb{Q}^s_z := \mathcal{L}(\mu_\zeta), \quad \text{where} \quad \zeta \sim \prod_{x \in \mathbb{X}, a \in \mathbb{A}} \mathrm{Ber}\Big(\frac{1 + zs\rho(x,a)}{2}\Big). \tag{4.76}$$

Having constructed the mixture distributions, we are ready to prove the lower bound (4.61c). The proof relies on the following two lemmas on the properties of the mixture distributions:

Lemma 4.4. *The total variation distance between mixture-of-product distributions is upper bounded as*

$$d_{\mathrm{TV}}\left(\int \mathbb{P}_{\mu,\xi^*}^{\otimes n} d\mathbb{Q}_1^s(\mu), \int \mathbb{P}_{\mu,\xi^*}^{\otimes n} d\mathbb{Q}_{-1}^s(\mu) \right) \leq 2s\sqrt{n} + 4 \cdot e^{-n/4}. \tag{4.77}$$

Lemma 4.5. *Given a state space with cardinality lower bounded as* $|\mathbb{X}| \geq 128 c_{\max}/s^2$*, we have*

$$\mathbb{P}_{\mu \sim \mathbb{Q}_1^s}\Big\{ \tau(\xi^*, \mu) \geq \tau(\xi^*, \mu^*) + \frac{s}{8}\|\delta\|_\omega \Big\} \geq 1 - 2 \cdot e^{-4}, \quad \textit{and} \tag{4.78a}$$

$$\mathbb{P}_{\mu \sim \mathbb{Q}_{-1}^s}\Big\{ \tau(\xi^*, \mu) \leq \tau(\xi^*, \mu^*) - \frac{s}{8}\|\delta\|_\omega \Big\} \geq 1 - 2 \cdot e^{-4}. \tag{4.78b}$$

We prove these lemmas at the end of this section.

Taking these two lemmas as given, we now proceed with the proof of equation (4.61c). Based on Lemma 4.5, we define two sets of functions as follows:

$$\mathscr{E}_1 := \Big\{ \mu_\zeta \mid \zeta \in \{-1, 1\}^{\mathbb{X} \times \mathbb{A}},\ \tau(\xi^*, \mu_\zeta,) \geq \tau(\xi^*, \mu^*) + \frac{s}{8}\|\delta\|_\omega \Big\}, \quad \text{and}$$
$$\mathscr{E}_{-1} := \Big\{ \mu_\zeta \mid \zeta \in \{-1, 1\}^{\mathbb{X} \times \mathbb{A}},\ \tau(\xi^*, \mu_\zeta) \leq \tau(\xi^*, \mu^*) - \frac{s}{8}\|\delta\|_\omega \Big\}.$$

When the sample size requirement in equation (4.61c) is satisfied, Lemma 4.5 implies that $\mathbb{Q}_z^s(\mathscr{E}_z) \geq 1 - e^{-4}$ for $z \in \{-1, 1\}$. We set $s = \frac{1}{16\sqrt{n}}$, and define

$$\mathbb{Q}_z^* := \mathbb{Q}_1^s \big| \mathscr{E}_z, \quad \text{for } z \in \{-1, 1\}. \tag{4.79}$$

By construction, the probability distributions $\mathbb{Q}_1^*$ and $\mathbb{Q}_{-1}^*$ have disjoint support, and for any pair $\mu \in \mathrm{supp}(\mathbb{Q}_1^*) \subseteq \mathscr{E}_1$ and $\mu' \in \mathrm{supp}(\mathbb{Q}_{-1}^*) \subseteq \mathscr{E}_{-1}$, we have:

$$\tau\big(\xi^*, \mu\big) \geq \tau\big(\xi^*, \mu^*\big) + \frac{\|\delta\|_\omega}{128\sqrt{n}}, \quad \text{and} \quad \tau\big(\xi^*, \mu'\big) \leq \tau\big(\xi^*, \mu^*\big) - \frac{\|\delta\|_\omega}{128\sqrt{n}}. \tag{4.80}$$

Furthermore, combining the conclusions in Lemma 4.4 and Lemma 4.5 using Lemma C.1, we obtain the total variation distance upper bound:

$$\begin{aligned}
&d_{\mathrm{TV}}\left(\int \mathbb{P}_{\mu,\xi^*}^{\otimes n} d\mathbb{Q}_1^*(\mu), \int \mathbb{P}_{\mu,\xi^*}^{\otimes n} d\mathbb{Q}_{-1}^*(\mu) \right) \\
&\leq \frac{1}{1 - 2 \cdot e^{-4}} d_{\mathrm{TV}}\left(\int \mathbb{P}_{\mu,\xi^*}^{\otimes n} d\mathbb{Q}_1^s(\mu), \int \mathbb{P}_{\mu,\xi^*}^{\otimes n} \mathbb{Q}_{-1}^s(\mu) \right) + 4 \cdot e^{-4} \\
&\leq \frac{1/8 + 4 \cdot e^{-n/4}}{1 - 2 \cdot e^{-4}} + 4 \cdot e^{-4} \leq \frac{1}{4},
\end{aligned}$$

which completes the proof of equation (4.75).

Combining equation (4.80) and (4.75), we can invoke Le Cam's mixture-vs-mixture lemma, and conclude that

$$\mathscr{M}_n \geq \tfrac{1}{4}\Big\{1 - d_{\mathrm{TV}}\Big(\int \mathbb{P}_{\mu,\xi^*}^{\otimes n} d\mathbb{Q}_1^*(\mu), \int \mathbb{P}_{\mu,\xi^*}^{\otimes n} d\mathbb{Q}_{-1}^*(\mu)\Big)\Big\} \cdot \inf_{\substack{\mu \in \mathrm{supp}(\mathbb{Q}_1) \\ \mu' \in \mathrm{supp}(\mathbb{Q}_{-1})}} \big[\tau(\xi^*, \mu) - \tau(\xi^*, \mu')\big]_+^2$$

$$\geq \frac{c\|\delta\|_\omega^2}{n},$$

for a universal constant $c > 0$. This completes the proof of equation (4.61c).

Proof of Lemma 4.4: Our proof exploits a Poissonization device, which makes the number of observations random, and thereby simplifies calculations. For $z \in \{-1, 1\}$, denote the mixture-of-product distribution:

$$\mathbb{Q}_z^{(s,\otimes n)} := \int \mathbb{P}_{\mu,\xi^*}^{\otimes n} d\mathbb{Q}_z^s(\mu).$$

We construct a pair $\big(\mathbb{Q}_1^{(s,\mathrm{Poi})}, \mathbb{Q}_{-1}^{(s,\mathrm{Poi})}\big)$ of mixture distributions as follows: randomly draw the sample size $\nu \sim \mathrm{Poi}(2n)$ independent of ζ and random sampling of data. For each $z \in \{-1, 1\}$, we let $\mathbb{Q}_z^{(s,\mathrm{Poi})}$ be the mixture distribution:

$$\mathbb{Q}_z^{(s,\mathrm{Poi})} := \sum_{k=0}^{+\infty} \mathbb{Q}_z^{(s,\otimes k)} \cdot \mathbb{P}(\nu = k).$$

By a known lower tail bound for a Poisson random variable (c.f. [22], §2.2), we have

$$\mathbb{P}\big[\underbrace{\nu \geq n}_{=:\widetilde{\mathscr{E}}_n}\big] \geq 1 - e^{-n/4}, \tag{4.81}$$

We note that on the event $\widetilde{\mathscr{E}}_n$, the probability law $\mathbb{Q}_z^{(s,\otimes n)}$ is actually the projection of the law $\mathbb{Q}_z^{(s,\mathrm{Poi})}\big|\widetilde{\mathscr{E}}_n$ on the first n observations. Consequently, we can use Lemma C.1 to bound the total variation distance between the original mixture distributions using that of the Poissonized models:

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathbb{Q}_1^{(s,\otimes n)}, \mathbb{Q}_{-1}^{(s,\otimes n)}\big) &\leq d_{\mathrm{TV}}\Big(\mathbb{Q}_1^{(s,\mathrm{Poi})}\big|\widetilde{\mathscr{E}}_n, \mathbb{Q}_{-1}^{(s,\mathrm{Poi})}\big|\widetilde{\mathscr{E}}_n\Big) + 4\mathbb{P}\big(\widetilde{\mathscr{E}}_n^c\big) \\ &\leq \frac{1}{\mathbb{P}\big(\widetilde{\mathscr{E}}_n\big)} d_{\mathrm{TV}}\Big(\mathbb{Q}_1^{(s,\mathrm{Poi})}, \mathbb{Q}_{-1}^{(s,\mathrm{Poi})}\Big) + 4 \cdot e^{-n/4} \\ &\leq 2 d_{\mathrm{TV}}\Big(\mathbb{Q}_1^{(s,\mathrm{Poi})}, \mathbb{Q}_{-1}^{(s,\mathrm{Poi})}\Big) + 4 \cdot e^{-n/4}, \end{aligned} \tag{4.82}$$

valid for any $n \geq 4$.

It remains to bound the total variation distance between the Poissonized mixture distributions. We start by considering the empirical count function

$$M(x, a) := \sum_{i=1}^{\nu} \mathbf{1}\Big\{X_i = x,\ A_i = a\Big\} \qquad \text{for all } (x, a) \in \mathcal{S} \times \mathbb{A},$$

Note that conditionally on the value of ν, the vector $(M(x,a))_{x\in\mathbb{X},a\in\mathbb{A}}$ follows a multinomial distribution. Since $\nu \sim \mathrm{Poi}(2n)$, we have

$$\forall x \in \mathbb{X},\; a \in \mathbb{A} \quad M(x,a) \sim \mathrm{Poi}\big(2n\xi^*(x)\pi(x,a)\big), \quad \text{independent of each other.}$$

For each $(x,a) \in \mathbb{X} \times \mathbb{A}$ and $z \in \{-1,1\}$, we consider a probability distribution $\mathbb{Q}'_z(x,a)$ defined by the following sampling procedure:

(a) Sample $M(x,a) \sim \mathrm{Poi}\big(2n\xi^*(x)\pi(x,a)\big)$.

(b) Sample $\zeta(x,a) \sim \mathrm{Ber}\big(\frac{1+sz\rho(x,a)}{2}\big)$.

(c) Generate a (possibly empty) set of $M(x,a)$ independent observations from the conditional law of Y given $X = x$ and $A = a$.

By independence, for any $z \in \{-1,1\}$, it is straightforward to see that:

$$\mathbb{Q}_z^{(s,\mathrm{Poi})} = \prod_{(x,a)\in\mathcal{S}\times\mathbb{A}} \mathbb{Q}'_z(x,a),$$

Pinsker's inequality, combined with the tensorization of the KL divergence, guarantees that

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathbb{Q}_1^{(s,\mathrm{Poi})}, \mathbb{Q}_{-1}^{(s,\mathrm{Poi})}\big) &\leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}\left(\mathbb{Q}_1^{(s,\mathrm{Poi})} \parallel \mathbb{Q}_{-1}^{(s,\mathrm{Poi})}\right)} \\ &= \sqrt{\frac{1}{2} \sum_{x\in\mathbb{X},a\in\mathbb{A}} D_{\mathrm{KL}}\left(\mathbb{Q}'_1(x,a) \parallel \mathbb{Q}'_{-1}(x,a)\right)}. \end{aligned} \tag{4.83}$$

Note that the difference between the probability distributions $\mathbb{Q}'(x,a)$ and $\mathbb{Q}'_{-1}(x,a)$ lies only in the parameter of the Bernoulli random variable $\zeta(x,a)$, which is observed if and only if $M(x,a) > 0$. By convexity of KL divergence, we have:

$$\begin{aligned} &D_{\mathrm{KL}}\left(\mathbb{Q}'_1(x,a) \parallel \mathbb{Q}'_{-1}(x,a)\right) \\ &\leq \mathbb{P}\Big(M(x,a) > 0\Big) \cdot D_{\mathrm{KL}}\left(\mathrm{Ber}\big(\frac{1+s\rho(x,a)}{2}\big) \parallel \mathrm{Ber}\big(\frac{1-s\rho(x,a)}{2}\big)\right) \\ &\leq 4\big(1 - e^{-2n\xi^*(x)\pi(x,a)}\big) \cdot s^2\rho^2(x,a) \\ &\leq 8n\xi^*(x)\pi(x,a)s^2\rho^2(x) \\ &\leq 8ns^2\xi^*(x)\frac{g(x,a)\delta^2(x,a)}{\pi(x,a)\|\delta\|_\omega^2} \end{aligned}$$

Substituting back to the decomposition result (4.83), we conclude that

$$d_{\mathrm{TV}}\big(\mathbb{Q}_1^{(s,\mathrm{Poi})}, \mathbb{Q}_{-1}^{(s,\mathrm{Poi})}\big) \leq \sqrt{\frac{1}{2} \sum_{x\in\mathbb{X},a\in\mathbb{A}} 8ns^2\xi^*(x)\frac{g^2(x,a)\delta^2(x,a)}{\pi(x,a)\|\delta\|_\omega^2}} \leq 2s\sqrt{n}.$$

Finally, combining with equation (4.82) completes the proof.

Proof of Lemma 4.5: Under our construction, we can compute the expectation of the target linear functional $\tau(\mathcal{I})$ under both distributions. In particular, for $z = 1$, we have

$$\begin{aligned}
&\mathbb{E}_{\mu \sim \mathbb{Q}_1^s}\big[\tau(\xi^*, \mu)\big] \\
&= \tau(\xi^*, \mu^*) + \frac{s}{2} \cdot \mathbb{E}_{\xi^*}\Big[\int_{\mathbb{A}} \delta(X, a) g(X, a) \rho(X, a) d\lambda(a)\Big] \\
&\geq \tau(\xi^*, \mu^*) + \frac{s}{2\|\delta\|_\omega} \cdot \mathbb{E}\Big[\frac{\delta^2(X, A) g(X, A)^2}{\pi^2(X, A)} \mathbf{1}\Big\{\frac{|g(X, A)| |\delta(X, A)}{\pi(X, A)} \leq 2M_{2\to 4}\|\delta\|_\omega\Big\}\Big],
\end{aligned}$$

where the last expectation is taken with respect to $X \sim \xi^*$ and $A \sim \pi(X, \cdot)$.

Applying Lemma C.2 to the random variable $g(X, A)\delta(X, A)/\pi(X, A)$ yields

$$\mathbb{E}\Big[\frac{\delta^2(X, A) g^2(X, A)}{\pi^2(X, A)} \mathbf{1}\Big\{\frac{|g(X, A)| |\delta(X, A)}{\pi(X, A)} \leq 2M_{2\to 4}\|\delta\|_\omega\Big\}\Big] \geq \frac{1}{2}\mathbb{E}\Big[\frac{\delta^2(X, A)\, g^2(X, A)}{\pi^2(X, A)}\Big].$$

Consequently, we have the lower bound on the expected value under $\mathbb{Q}_1^s$

$$\mathbb{E}_{\mu \sim \mathbb{Q}_1^s}\big[\tau(\xi^*, \mu)\big] \geq \tau(\xi^*, \mu^*) + \frac{s}{4}\|\delta\|_\omega. \tag{4.84a}$$

Similarly, under the distribution $\mathbb{Q}_{-1}^s$, we note that:

$$\mathbb{E}_{\mu \sim \mathbb{Q}_{-1}^s}\big[\tau(\xi^*, \mu)\big] \leq \tau(\xi^*, \mu^*) - \frac{s}{4}\|\delta\|_\omega. \tag{4.84b}$$

We now consider the concentration behavior of random function $\mu \sim \mathbb{Q}_z^s$ for each choice of $z \in \{-1, 1\}$. Since the random signs are independent at each state-action pair $(x, a) \in \mathbb{X} \times \mathbb{A}$, we can apply Hoeffding's inequality: more precisely, with with probability $1 - 2e^{-2t}$, we have

$$\begin{aligned}
\big|\tau(\xi^*, \mu) - \mathbb{E}_{\mu \sim \mathbb{Q}_z^s}\big[\tau(\xi^*, \mu)\big]\big| &\overset{(i)}{\leq} \sqrt{t \cdot \sum_{x \in \mathbb{X}, a \in \mathbb{A}} \xi^{*2}(x) g^2(x, a) \delta^2(x, a)} \\
&\leq \sqrt{\frac{tc_{\max}}{|\mathbb{X}|} \cdot \sum_{x \in \mathbb{X}, a \in \mathbb{A}} \xi^*(x) g^2(x, a) \delta^2(x, a)} \leq \sqrt{\frac{t \cdot c_{\max}}{|\mathbb{X}|}}\|\delta\|_\omega,
\end{aligned}$$

where in step (i), we use the compatibility condition $\xi^*(x) \leq \frac{c_{\max}}{|\mathbb{X}|}$ for any $x \in \mathbb{X}$.

Given a state space with cardinality lower bounded as $|\mathbb{X}| \geq 128 c_{\max}/s^2$, we can combine the concentration bound with the expectation bounds (4.84) so as to obtain

$$\begin{aligned}
\mathbb{P}_{\mu \sim \mathbb{Q}_1^s}\Big\{\tau(\xi^*, \mu) \geq \tau(\xi^*, \mu^*) + s\sqrt{\tfrac{t}{128}}\|\delta\|_\omega\Big\} &\geq 1 - 2 \cdot e^{-2t}, \quad \text{and} \\
\mathbb{P}_{\mu \sim \mathbb{Q}_{-1}^s}\Big\{\tau(\xi^*, \mu) \leq \tau(\xi^*, \mu^*) - s\sqrt{\tfrac{t}{128}}\|\delta\|_\omega\Big\} &\geq 1 - 2 \cdot e^{-2t}.
\end{aligned}$$

Setting $t = 2$ completes the proof of Lemma 4.5.

4.5.2 Proof of Proposition 4.1

Let the input distribution ξ^* be the uniform distribution over the sequence $\{x_j\}_{j=1}^D$. It suffices to show that

$$\inf_{\widehat{\tau}_n} \sup_{\mu \in \mathcal{F}} \mathbb{E}\big[\,|\widehat{\tau}_n - \tau(\xi^*, \mu)|^2\,\big] \geq \frac{c}{n}\left\{\frac{1}{D}\sum_{j=1}^D \sum_{a \in \mathbb{A}} \frac{g^2(s_j, a)}{\pi(s_j, a)}\delta_a^2\right\}. \tag{4.85}$$

Recall that we are given a sequence $\{x_j\}_{j=1}^D$ such that for each $a \in \mathbb{A}$, the function class $\mathcal{F}_a$ shatters it at scale δ_a. Let $\{t_{j,a}\}_{j=1}^D$ be the sequence of function values in the fat-shattering definition (4.38). Note that since the class $\mathcal{F}$ is convex, we have

$$\bigotimes_{j=1}^D \bigotimes_{a \in \mathbb{A}} [t_{j,a} - \delta_a, t_{j,a} + \delta_a] \subseteq \bigotimes_{a \in \mathbb{A}} \Big\{(f_a(x_j))_{j \in [D]} \;\mid\; f_a \in \mathcal{F}_a\Big\}.$$

Note that this distribution satisfies the compatibility condition with $c_{\max} = 1$ and the hyper-contractivity condition with a constant $M_{2\to4} = \|g(X, A)\delta_A / \pi(X, A)\|_{2\to4}$. Invoking equation (4.61c) over the local neighborhood $\mathfrak{N}_\delta^{val}(t)$ yields the claimed bound (4.85).

4.6 Discussion

We have studied the problem of evaluating linear functionals of the outcome function (or reward function) based on observational data. In the bandit literature, this problem corresponds to off-policy evaluation for contextual bandits. As we have discussed, the classical notion of semi-parametric efficiency characterizes the optimal asymptotic distribution, and the finite-sample analysis undertaken in this chapter enriches this perspective. First, our analysis uncovered the importance of a particular weighted $\mathbb{L}^2$-norm for estimating the outcome function μ^*. More precisely, optimal estimation of the scalar τ^* is equivalent to optimal estimation of the outcome function μ^* under such norm, in the sense of minimax risk over a local neighborhood. Furthermore, when the outcome function is known to lie within some function class $\mathcal{F}$, we showed that a sample size scaling with the complexity of $\mathcal{F}$ is necessary and sufficient to achieve such bounds non-asymptotically.

Our result lies at the intersection of decision-making problems and the classical semi-parametric theories, which motivates several promising directions of future research on both threads:

- Our analysis reduces the problem of obtaining finite-sample optimal estimates for linear functionals to the nonparametric problem of estimating the outcome function under a weighted norm. Although the re-weighted least-square estimator (4.11) converges to the best approximation of the treatment effect function in the class, it is not clear whether it always achieves the optimal trade-off between the approximation and estimation errors. How to optimally estimate the nonparametric component under weighted norm (so as to optimally estimate the scalar τ^* in finite sample) for a variety of function classes is an important direction of future research, especially with weight functions.

- The analysis of the current chapter was limited to i.i.d. data, but similar issues arise with richer models of data collection. There are recent lines of research on how to estimate linear functionals with adaptively collected data (e.g. when the data are generated from an exploratory bandit algorithm [228, 189, 97]), or with an underlying Markov chain structure (e.g. in off-policy evaluation problems for reinforcement learning [83, 225, 90, 227]). Many results in this literature build upon the asymptotic theory of semi-parametric efficiency, so that it is natural to understand whether extensions of our techniques could be used to obtain finite-sample optimal procedures in these settings.

- The finite-sample lens used in this chapter reveals phenomena in semi-parametric estimation that are washed away in the asymptotic limit. This chapter has focused on a specific class of semi-parametric problems, but more broadly, we view it as interesting to see whether such phenomena exist for other models in semi-parametric estimation. In particular, if a high-complexity object—such as a regression or density function—needs to be estimated in order to optimally estimate a low-complexity object—such as a scalar—it is important to characterize the minimal sample size requirements, and the choice of nonparametric procedures for the nuisance component that are finite-sample optimal.

Chapter 5

Beyond semi-parametric efficiency: a kernel-based analysis

Continuing the discussion in Chapter 4, we study optimal procedures for estimating a linear functional based on observational data. In many problems of this kind, a widely used assumption is strict overlap, i.e., uniform boundedness of the importance ratio, which measures how well the observational data covers the directions of interest. When it is violated, the classical semi-parametric efficiency bound and the bounds in Chapter 4 can easily become infinite, so that the instance-optimal risk depends on the function class used to model the regression function. For any convex and symmetric function class $\mathcal{F}$, we derive a non-asymptotic local minimax bound on the mean-squared error in estimating a broad class of linear functionals. This lower bound refines the classical semi-parametric one, and makes connections to moduli of continuity in functional estimation. When $\mathcal{F}$ is a reproducing kernel Hilbert space, we prove that this lower bound can be achieved up to a constant factor by analyzing a computationally simple regression estimator. We apply our general results to various families of examples, thereby uncovering a spectrum of rates that interpolate between the classical theories of semi-parametric efficiency (with $\sqrt{n}$-consistency) and the slower minimax rates associated with non-parametric function estimation.

5.1 Introduction

Estimation and inference problems based on observational data arise in various applications, and are studied in the fields of causal inference, econometrics, and reinforcement learning. An interesting subclass of such problems are semi-parametric in nature: they involve estimating the value of a linear functional in the presence of one or more unknown non-parametric "nuisance" functions.

More concretely, suppose that we observe n i.i.d. triples of the form (X_i, A_i, Y_i), where each triple is drawn according to the following procedure

- the state variable X_i is drawn from some distribution ξ^* over the state space $\mathbb{X}$.

- the action variable A_i is drawn with conditional distribution $A_i \mid X_i \sim \pi(\cdot \mid X_i)$, where π is a *behavioral policy*, also known as the propensity score in the causal inference literature.
- the response or outcome variable Y_i has conditional expectation $\mathbb{E}\big[Y_i \mid X_i, A_i\big] = \mu^*(X_i, A_i)$, where μ^* is the *regression function*, also known as the treatment effect.

The distribution ξ^* and regression function μ^* are both unknown, and we would like to estimate a known functional that depends on both of them. More precisely, given a known family of functions $\{\omega(\cdot \mid x) \mid x \in \mathbb{X}\}$, where each $\omega(\cdot \mid x)$ is a signed Radon measure over the action space $\mathbb{A}$, consider the functional

$$(\mu, \xi) \mapsto \mathscr{L}_\omega(\mu, \xi) := \mathbb{E}_{X \sim \xi}\left[\int_{\mathbb{A}} \mu^*(X, a) d\omega(a \mid X)\right] \;=\; \int_{\mathbb{X}} \int_{\mathbb{A}} \mu^*(x, a)\, d\omega(a \mid x)\, d\xi(x). \tag{5.1}$$

Our goal is to estimate $\tau^* := \mathscr{L}_\omega(\mu^*, \xi^*)$—the value of this functional at the unknown pair (μ^*, ξ^*). The behaviorial policy π is also unknown, and it plays the role of another non-parametric nuisance, since it affects the joint distribution of the samples (X_i, A_i) that we observe. Special cases of this set-up include estimating the average treatment effect (ATE), and off-policy evaluation for contextual bandit problems. We also consider a variant in which, instead of taking the expectation over $X \sim \xi$, we evaluate at a fixed state s_0. This latter set-up is appropriate for the conditional average treatment effect (CATE).

There are a variety of settings—involving particular assumptions on the regression function and behavioral policy—under which estimates of τ^* based on n samples are consistent at the classical $\sqrt{n}$-rate. Moreover, via the classical notion of semi-parametric efficiency [120], we have a refined understanding of the optimal instance-dependent constants that should accompany this $\sqrt{n}$-rate [71]. However, there are also various settings—of interest in practice—in which the efficiency bound is infinite, and the classical $\sqrt{n}$-convergence no longer holds. This issue is not only theoretical in nature: when applied to problems of this type, many standard estimators for τ^*—being motivated by classical considerations—no longer perform well.

At a high level, there are at least two types of phenomena that can invalidate classical $\sqrt{n}$-consistency. First, if both the regression function and behavioral policy need to be estimated from classes with high complexity, the difficulty of doing so—as opposed to only the fluctuations intrinsic to the target functional—can become dominant. For instance, the paper [179] studies a variety of such cases involving Hölder classes; see also the paper [96] for related results on CATE estimation. Second, the semi-parametric efficiency bound involves certain moments of the ratio $\frac{dg}{d\pi}$. This so-called *importance ratio* measures how well the observational data, as controlled by the behavioral policy π, "covers" the regions of space relevant for estimating the functional. If this coverage is especially bad, then the importance ratio need not have finite moments. This latter cause of breakdown in semi-parametric efficiency is the primary motivation for the theory and methodology put forth in this chapter.

In the literature on causal inference with observational studies, it is common to impose the so-called *strict overlap condition* [76, 41, 197]. The strict overlap condition amounts to imposing a uniform bound on the importance ratio, and so precludes the possibility of infinite moments. Such uniform boundedness conditions also appear frequently in the closely related literature on bandits and reinforcement learning. On one hand, this condition is known to be necessary in a worst-case sense: as shown by Khan and Tamer [100], when neither strict overlap nor structural conditions on the regression function are imposed, then it is no longer possible to obtain $\sqrt{n}$-consistency. It should be noted, however, that the uniform boundedness condition can be quite stringent. For instance, in some recent work, D'Amour et al. [42] show that it rules out many interesting cases of practical interest, especially when the model involves high-dimensional covariates. Motivated by this dilemma, there is a line of past and on-going work (e.g., [38, 78, 134]) that proposes estimators that exploit some kind of structure in the regression function. Despite this progress, we currently have a relatively limited understanding of *optimal methods* for estimating linear functionals based on observational data without imposing the strict overlap condition.

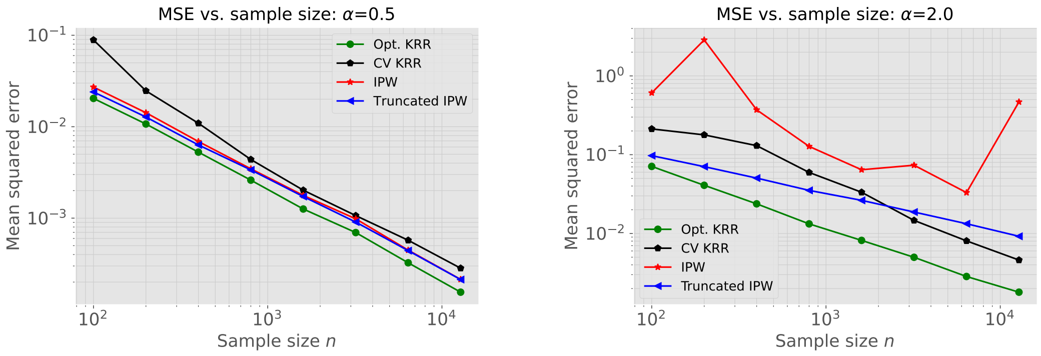
With this context, the main contributions of this chapter are to provide some insight into the nature of optimal methods for estimating linear functionals without (strict) overlap. Our first main result is a general non-asymptotic lower bound on the mean-squared error of any estimator. This lower bound involves a novel variance functional, which depends both on the function class $\mathcal{F}$ used to model the regression function and the behavior of the importance ratio. Turning to upper bounds, we focus on the class of reproducing kernel Hilbert spaces (RKHSs) as models for the regression function, and provide a computationally simple procedure that achieves our local minimax lower bound. Thus, for RKHS-based models of the regression function, we are able to identify the instance-dependent and non-asymptotic local minimax risk up to a constant pre-factor. As we illustrate by a range of examples, this mean-squared error can exhibit a range of scalings, from the classic $\sqrt{n}$-consistency for well-behaved problems to much slower non-parametric rates in cases where the importance ratio is badly behaved.

5.1.1 An illustrative simulation

So as to provide intuition for the results to follow, let us consider a simple family of problems for which the strict overlap assumption is violated, and compare the performance of the estimator proposed in this chapter to other alternatives. More specifically, we consider a missing data problem where the action $a \in \mathbb{A} = \{0, 1\}$ is an indicator of "missingness". With the state space $\mathcal{S} = [0, 1]$, we construct a weight function g and behavioral policy π for which the importance ratio takes the form

$$\frac{dg}{d\pi}(0 \mid x) = 1, \quad \text{and} \quad \frac{dg}{d\pi}(1 \mid x) \sim (1 - x)^{-\alpha} \qquad \text{where } \alpha \geq 0 \text{ is a parameter.} \tag{5.2}$$

The parameter α controls the heaviness of the tails exhibited by the importance ratio: when $\alpha = 0$, the importance ratio is simply a constant, whereas as α increases, its tails become increasingly heavy. Above $\alpha > 1$, it no longer has a finite second moment, and this transition point turns out to be interesting.



(a) Light tails: $\alpha = 0.5$

(b) Heavy tails: $\alpha = 2.0$

Figure 5.1: Log-log plots of mean-squared error versus sample size n for four different estimators fo τ^*: procedure Opt-KRR is analyzed in this chapter, whereas CV-KRR is a related method with regularization parameter chosen by cross-validation. We also compare to the classical IPW estimate along with a truncated version of IPW. (a) Setting $\alpha = 0.5$ yields a propensity score with light tails, and our theory predicts classical n^{-1}-decay of the MSE for Opt-KRR. (b) Setting $\alpha = 2.0$ yields a heavy-tailed problem, and our theory guarantees consistency of Opt-KRR at the rate $n^{-3/4}$.

In Figure 5.1, we compare the performance of four different methods: the classical *Inverse Propensity Weighting* (IPW) estimator [181] (we review this estimator in Section 5.4), a truncated version of IPW [100], the optimal kernel-based procedure proposed in this chapter (Opt-KRR), as well as a sub-optimal kernel-based procedure where the regularization parameter is chosen by cross validation (CV-KRR). In each panel and for each estimate $\widehat{\tau}_n$, we plot the mean-squared error $\mathbb{E}[(\widehat{\tau}_n - \tau^*)^2]$ versus the sample size n on a log-log scale. Panel (a) corresponds to the setting $\alpha = 0.5$: in this case, our theory predicts that the minimax mean-squared error should decay as n^{-1}, as expected for MSE in the classical regime of $\sqrt{n}$-consistency. All four methods are relatively well-behaved for this problem; for our proposed method (Opt-KRR), performing a linear regression of log-MSE on $\log n$ gives a slope estimate of -1.00 ± 0.01. Panel (b), in contrast, exhibits very different behavior: by setting $\alpha = 2.0$, we obtain a much harder problem. Here the IPW performance is very erratic due to the heavy tails of the importance ratio; the truncated version is better behaved, but still has larger error than Opt-KRR. In fact, the theory given in this chapter, when specialized to this family, predicts that for any $\alpha > 1$, the optimal mean-squared error should decay at the rate $n^{-\frac{3}{\alpha+2}}$. Thus, if we set

$\alpha = 2$, then we expect to see an error decay with exponent $-3/4$. In order to estimate the decay rate of Opt-KRR, we again perform a linear regression of the log MSE on $\log(n)$, and obtained an estimated slope -0.76 ± 0.01. Once again, we see excellent agreement with the theory.

5.1.2 Our contributions

We summarize the contributions made in the remainder of the chapter.

- First, working in the setting where the regression function belongs to a known function class, we establish a general non-asymptotic local minimax lower bound for estimating linear functionals from observational data. The lower bound is defined by a variational problem the captures the interplay between the geometry of the function class and the importance ratio. As the proof is based on Le Cam's two point approach, portions of the bound involve a certain modulus of continuity.
- Second, specializing to the regression function belonging to a ball within a reproducing kernel Hilbert space (RKHS), we analyze a class of multi-stage outcome regression estimators. Under certain regularity conditions on the RKHS and the conditional covariance function, we establish a non-asymptotic upper bound that matches our minimax lower bound (up to a constant factor).
- Third, we illustrate our general result by applying it to a range of problems, thereby obtaining a variety of novel minimax rates. For treatment effect estimation when the importance ratio diverges at certain points, we show that minimax risk depends on the interaction between this singularity and the geometry of the Hilbert space. In the setting of contextual bandits with continuous states and actions, we give results on off-policy evaluation of deterministic policies, thereby obtaining novel minimax rates that are adaptive to the complexity of the state-action space.

Notably, our multi-stage kernel-based estimator requires *no knowledge* of the underlying behavioral policy or propensity score π. This property is very attractive from the implementation point of view. At the same time, its performance in terms of MSE matches our minimax lower bound, which applies to a broader family of estimators including those that *know* the behavioral policy. Thus, we see an interesting implication of our results: as long the regression function is a member of a RKHS, knowledge of the behavioral policy plays no role in determining the minimax risk rate. This is in sharp contrast with Hölder classes of the non-Donsker type, where both parts of the model play an important role [179, 96]. Finally, although the value of minimax risk itself depends on the behavioral policy, the tuning parameter in our kernel-based estimator does not.

5.1.3 Related work

Now let us discuss various bodies of related work so as to situate our work within a broader context.

Instance-optimality for non/semi-parametric estimation: For regular parametric models, the classical local asymptotic minimax (LAM) framework of Le Cam and Hajek [117, 72] specifies the instance-optimal behavior of estimators as $n \to \infty$. Levit [120] extended this framework to semi-parametric settings by considering the collection of all finite-dimensional sub-models. For the specific class of linear functional estimation problems considered here, Hahn [71] laid out the asymptotic lower bounds, whereas Chapter 4 studies the same question within a non-asymptotic framework.

Beyond the classical $\sqrt{n}$-regime, instance-dependent optimality for semi-parametric and non-parametric estimation have been established under various settings. In the literature, exact local asymptotic minimax risks are obtained for Sobolev space regression [167, 30], spectral density estimation [101], and shape-constrained estimation [74]. For estimation problems involving linear functionals, Donoho [53] establishes information-theoretic optimality (up to constant factors) of certain class of minimax linear estimators; in the regression setting, this framework applies to fixed design problems as opposed to the random design setting of interest here. Also studying fixed design regression using spline methods, the unpublished work of Speckman [193] is based on a class of under-smoothed estimators. These spline-based estimators are a special case of the more general RKHS set-up considered here for the random design setting, and we also find that a form of under-smoothing is optimal. While all the preceding results are stated as global minimax risks, due to the location-family structure of the underlying model and simplifying noise assumptions, the bounds are also instance-optimal, albeit in a less refined manner.

Overlap and coverage assumptions for off-policy estimation: The *overlap assumption*, first proposed by Rosenbaum and Rubin [182], requires that the behavioral policy or propensity score takes value within the open interval $(0, 1)$.[1] In our general set-up, the overlap assumption is equivalent to requiring that the importance ratio $\frac{dg}{d\pi}$ exists everywhere. Such a condition, along with the unconfoundedness assumption, together imply identifiability of the average treatment effect [182], but could lead to arbitrarily slow rates. In the literature, a popular choice is the much stronger *strict overlap assumption* [76, 41], which requires the importance ratio to be uniformly bounded. Khan and Tamer [100] shows that strictly overlap is a necessary condition for uniform $\sqrt{n}$-consistency in the worst case. On the other hand, recent work [42] revealed that the strict overlap condition can be stringent in some natural high-dimensional problem setups. By making stronger assumptions about the regression function, the strict overlap

[1] In the classical binary treatment setup, the action space is $\mathbb{A} = \{0, 1\}$, and the propensity score is defined as $\pi(x) := \mathbb{P}(A = 1|X = x)$.

condition can be relaxed [38, 80], while still achieving the semi-parametric efficiency bound in the $\sqrt{n}$ regime. The case of infinite semi-parametric efficiency bound, known as the *irregular identification* regime, has been studied by prior works [100, 134], where truncated versions of IPW estimators are proposed and analyzed in some special cases. Moreover, instability in the behavior of these estimators has been documented in both simulation and real-data experimental studies [129, 92, 63].

Uniform boundedness of the importance ratio is also a canonical assumption in the bandit and reinforcement learning literature. Focusing on off-policy evaluation for bandit algorithms, Wang et al. [215] proposed a "switch estimator" that involves truncating the importance ratio. Ma et al. [133] showed that this procedure is worst-case optimal for multi-arm bandits. For off-policy reinforcement learning problems, uniform bounds on the importance ratio, known as coverage or concentrability coefficients, appear in various papers [89, 225, 222]. Most closely related to our results are the bounds in the paper [227], which apply to MDPs with linear function approximation and involve a finer-grained measure of the overlap between the behavioral and target policies.

Kernel and nonparametric methods for off-policy estimation: There is also a line of past work on studying various non-parametric procedures for estimating the average treatment effect under different structural conditions. Under the strict overlap condition combined with Hölder conditions imposed on both the importance ratio and regression function, minimax rates for ATE estimation have been established [179, 175], albeit with pre-factors depending on the instance that need not be optimal. In the Donsker regime considered here, these minimax rates coincide with the classical $\sqrt{n}$-rate, due to the presence of strict overlap. Our results reveal different phenomena that can arise without strict overlap—more specifically, the optimal rate is determined not only by the complexities of the importance ratio and the regression function classes, but also by any *singularity* in the importance ratio, and how it interacts with the functional to be estimated. Additionally, our results also apply to estimation of one-point linear functionals, a generalization of conditional or heterogeneous average treatment effects. Again with the focus on Hölder classes, some recent work [66, 96] has exhibited rate-optimal non-parametric procedures. In recent years, due to their flexibility and computational tractability, kernel-based approaches have been the focus of research in the causal estimation literature [191, 192, 163], where kernel-based estimators have been developed for various functionals.

Recent work on off-policy estimation has explored the use of minimax linear estimators. In the fixed-design setup, the papers [4, 87, 88] apply the classical framework of minimax linear estimators [53, 193] to the estimation problem for the *sample average treatment effect* (SATE), and establish guarantees of both the asymptotic and non-asymptotic flavors. Hirshberg et al. [77] studied a minimax linear estimator for the treatment effect when the regression function belongs to an RKHS; as in the classical work [193], this estimator can be reformulated in terms of a standard kernel ridge regression estimate, as can the two-stage procedure that we analyze in the simpler homoskedastic setting. Under the strict overlap condition and some additional regularity assumptions, they prove asymptotic efficiency as well as non-asymptotic bounds on the empirical loss function. In a more general set-up with general function classes, Hirshberg and Wager [78] proposed an augmented minimax linear estimator andestablished non-asymptotic normal approximation results. When specialized to off-policy estimation, their results yield non-asymptotic normal approximation in the classical regime with finite semi-parametric efficiency bound, but independent of the strict overlap assumption. An important contrast with our results is that these bounds involve both error and approximation error associated with the importance ratio (via the Riesz representer); in contrast, such terms do not arise in our approach.

5.2 Problem set-up and preview

We begin in Section 5.2.1 with a precise formulation of the problem and discussion of some examples. In Section 5.2.2, we describe the classical semi-parametric efficiency bound, and detail how our analysis moves beyond it.

5.2.1 Problem set-up and some examples

Given some probability distribution ξ^* over the state space $\mathbb{X}$, suppose that we observe n i.i.d. triples (X_i, A_i, Y_i) in which $X_i \sim \xi^*$, and

$$A_i \mid X_i \sim \pi(\cdot \mid X_i), \quad \text{and} \quad \mathbb{E}\big[Y_i \mid X_i, A_i\big] = \mu^*(X_i, A_i), \qquad \text{for } i = 1, 2, \ldots, n. \tag{5.3}$$

In addition to the regression function μ^*, our analysis also involves the conditional variance function

$$\sigma^2(x, a) := \mathbb{E}\Big[|Y - \mu^*(X, A)|^2 \mid X = x, A = a\Big], \tag{5.4}$$

which is assumed to exist for any pair $(x, a) \in \mathbb{X} \times \mathbb{A}$.

As previously described, given a collection of signed Radon measures $\omega(\cdot \mid x)$ over the action space $\mathbb{A}$, one for each $x \in \mathbb{X}$, our goal is to estimate the value $\tau^* = \mathscr{L}_\omega(\mu^*, \xi^*)$ of the bilinear functional

$$(\mu, \xi) \mapsto \mathscr{L}_\omega(\mu, \xi) := \int_{\mathcal{S}} \int_{\mathbb{A}} \mu(x, a) d\omega(a \mid x) d\xi(x) \tag{5.5}$$

evaluated at the pair $\mu = \mu^*$ and $\xi = \xi^*$. We require that the signed measures defining $\mathscr{L}_\omega$ satisfy the condition

$$\int_{\mathbb{A}} d\,|\omega(a \mid x)| \leq 1 \qquad \text{for each } x \in \mathcal{S}. \tag{5.6}$$

This holds automatically when each $\omega(\cdot \mid x)$ is a conditional probability distribution, as in off-policy evaluation for contextual bandits.

Various types of weight functions g arise in practice:

Average treatment effect (ATE): This linear functional arises with the binary action action space $\mathbb{A} = \{0, 1\}$, and weight function $\omega(a \mid x) = a - \frac{1}{2}$ for all x. With this choice, we have

$$\tau^* = \frac{1}{2} \mathbb{E}_{X \sim \xi^*} \Big[\mu^*(X, 1) - \mu^*(X, 0) \Big],$$

so that τ^* is proportional to the usual average treatment effect (i.e., equal up to the pre-factor $1/2$ that arises from our choice of normalization).

Off-policy evaluation for multi-arm contextual bandits: In the multi-arm setting of a contextual bandit, we have a finite action space $\mathbb{A}$, and each weight function $\omega(\cdot \mid x)$ defines a conditional probability over the action space, which can be interpreted as a stochastic policy. We say that the weight functions g define the *target policy* whereas the conditional distributions π define the *behavioral policy.*

Contextual bandits with continuous arms: In this case, we take the action space $\mathbb{A}$ to be a compact subset of $\mathbb{R}^d$. For a deterministic target policy $T : \mathbb{X} \to \mathbb{A}$, we let $\omega(\cdot \mid x)$ be the unit atomic mass at $T(x)$.

Also of interest—in addition to the functional (5.5)—is the variant obtained by replacing the expectation over $X \sim \xi^*$ with evaluation at a known state x_0—namely

$$\mathscr{L}_\omega(\mu^*, \delta_{x_0}) := \int_{\mathbb{A}} \mu^*(x_0, a) d\omega(a \mid x_0), \tag{5.7}$$

where δ_{x_0} can be thought of as a point mass at x_0. While the functional is determined by δ_{x_0}, the samples X_i themselves are still drawn from the distribution ξ^* over the state space.

Particular examples of the functional (5.7) include the conditional average treatment effect (CATE) in the causal inference literature, whereas in off-policy reinforcement learning, it includes the problem of evaluating the policy at a fixed state x_0 based on off-policy observations.

5.2.2 Moving beyond classical semi-parametric efficiency

In this section, we explain how this chapter moves beyond classical semi-parametric efficiency.

Recap of classical results: We begin by explaining the usual semi-parametric efficiency bound, which is meaningful when the importance ratio $\frac{dg}{d\pi}$ exists and has suitably controlled moments. Under these conditions, it is possible to obtain estimates $\widehat{\tau}_n$ of τ^* that converge at a $\sqrt{n}$-rate. We can thus ask about the variance associated with the rescaled error $\sqrt{n}(\widehat{\tau}_n - \tau^*)$, and in particular the smallest one that can be achieved.

For estimating $\tau^* = \mathscr{L}_\omega(\mu^*, \xi^*)$, it is known [71] that the smallest variance achievable, in the sense of semi-parametric efficiency, is given by

$$v^2_{\mathrm{semi}}(\mu^*, \tfrac{dg}{d\pi}) = \underbrace{\mathrm{var}_{X \sim \xi^*}\Big(\int_{\mathbb{A}} \mu^*(X, a) d\omega(a \mid X) \Big)}_{V^2_{\xi^*}(\mu^*)} + \underbrace{\mathbb{E}_{\xi^* \cdot \pi}\Big\{ \big[\tfrac{dg}{d\pi}(A \mid X) \big]^2 \, \sigma^2(X, A) \Big\}}_{V^2_\sigma(\frac{dg}{d\pi})}, \tag{5.8}$$

where $\mathbb{E}_{\xi^* \cdot \pi}$ denotes expectation over a pair $X \sim \xi^*$ and $A \sim \pi(\cdot \mid X)$.

This optimal variance consists of two term. The first term $V^2_{\xi^*}(\mu^*)$ captures the fluctuations in an estimate of τ^* due to the randomness in sampling the states from ξ^*. This term depends on the regression function μ^*, but not on the conditional variance function σ^2. In contrast, the second term $V^2_\sigma(\frac{dg}{d\pi})$ depends on both the importance ratio $\frac{dg}{d\pi}$ and the conditional variance function σ^2 but *not* on the regression function: it captures the interaction between the noise and the importance ratio $\frac{dg}{d\pi}$. It is this latter term that can diverge if the importance ratio is ill-behaved, and accordingly, it is the term that takes a more refined form in our analysis.

Non-asymptotic bounds: With this context, our main contributions are to move beyond classical (asymptotic) semi-parametric efficiency in the following ways:

- We use Le Cam's method to prove a general non-asymptotic minimax lower bound on estimating functionals from observational data without the overlap condition, but with μ^* belonging to a convex function class $\mathcal{F}$.
- When $\mathcal{F}$ is a reproducing kernel Hilbert space (RKHS), we show that this lower bound can be achieved by a four-stage kernel regression procedure, and we compute an explicit representation of the minimax risk (sharp up to constant pre-factors).

Let us describe our explicit representation of the non-asymptotic minimax risk in the RKHS setting. Consider an RKHS $\mathscr{H}$ that is a subset of $\mathbb{L}^2(\xi^* \cdot \pi)$, and suppose that the regression function μ^* belongs to the Hilbert ball $\mathbb{B}_{\mathscr{H}}(R)$ of radius R in this space. We show that the non-asymptotic minimax risk replaces the second term $V^2_\sigma(\frac{dg}{d\pi})$ in the classical semi-parametric efficiency bound (5.8) with a novel quantity associated with the eigenvalues and eigenfunctions associated with the RKHS. More precisely, any RKHS of the Mercer type is associated with a sequence $\{\lambda_j\}_{j=1}^\infty$ of positive eigenvalues, and associated eigenfunctions $\{\phi_j\}_{j=1}^\infty$. We let $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_j\}_{j=1}^\infty$ be a diagonal matrix defined by the eigenvalues, and using the eigenfunctions, we define an infinite-dimensional vector $\bar{u} = \bar{u}(\xi^*)$ with elements

$$(\bar{u})_j := \mathbb{E}_{X \sim \xi^*}\Big[\int_{\mathbb{A}} \phi_j(X, a) d\omega(a \mid X) \Big] \qquad \text{for } j = 1, 2, \ldots, \tag{5.9a}$$

along with the infinite-dimensional matrix $\mathbf{\Gamma}_\sigma$ with elements

$$[\mathbf{\Gamma}_\sigma]_{jk} := \mathbb{E}_{\xi^* \cdot \pi}\Big[\tfrac{1}{\sigma^2(X,A)}\phi_j(X,A)\phi_k(X,A)\Big] \qquad \text{for } j,k = 1,2,\ldots. \tag{5.9b}$$

We prove that when the regression function μ^* lies within a ball of radius R within this RKHS, then the minimax mean-squared error for estimating τ^* is proportional to $\frac{1}{n}\big\{V_{\xi^*}^2(\mu^*) + \tilde{V}_{\sigma,n}^2(\pi, g; \mathbb{B}_{\mathscr{H}}(R))\big\}$, where

$$\tilde{V}_{\sigma,n}^2(\pi, g; \mathbb{B}_{\mathscr{H}}(R)) := \bar{u}^\top \Big(\mathbf{\Gamma}_\sigma + \tfrac{1}{R^2 n}\mathbf{\Lambda}^{-1}\Big)^{-1} \bar{u}. \tag{5.10}$$

Note that $\tilde{V}_{\sigma,n}^2(\pi, g; \mathbb{B}_{\mathscr{H}}(R))$ depends (among other quantities) on the sample size n, and it can actually diverge as $n \to \infty$. This type of divergence leads to non-parametric rates for estimating the functional τ^*. Indeed panel (b) in Figure 5.1 provides an illustration of this phenomenon in one particular setting.

Connection to classical semi-parametric efficiency: To understand the connection between our result and the the classical semi-parametric efficiency bound (5.8), let us consider[2] the following special case:

- The importance ratio $\frac{dg}{d\pi}$ exists, and the classical semi-parametric efficiency bound is finite.
- The problem is homoskedastic, with constant conditional variance function $\sigma^2(x,a) = \overline{\sigma}^2$ for all pairs (x,a).

Under homoskedasticity, the matrix $\mathbf{\Gamma}_\sigma$ is diagonal with $\frac{1}{\overline{\sigma}^2}$ along its diagonal, using the fact that the eigenfunctions are orthonormal in $\mathbb{L}^2(\xi^* \cdot \pi)$. Since the matrix $\mathbf{\Lambda}^{-1}$ is also diagonal, we find that

$$\tilde{V}_{\sigma,n}^2(\pi, g; \mathbb{B}_{\mathscr{H}}(R)) = \sum_{j=1}^{\infty} \frac{\bar{u}_j^2}{\frac{1}{\overline{\sigma}^2} + \frac{1}{R^2 n \lambda_j}} \;\leq\; \overline{\sigma}^2 \sum_{j=1}^{\infty} \bar{u}_j^2. \tag{5.11a}$$

When the importance ratio $\frac{dg}{d\pi}$ exists, we can write

$$[\bar{u}]_j := \mathbb{E}_{X \sim \xi^*}\Big[\int_{\mathbb{A}} \phi_j(X,a) d\omega(a \mid X)\Big] \;=\; \mathbb{E}_{\xi^* \cdot \pi}\big[\phi_j(X,A)\tfrac{dg}{d\pi}(A \mid X)\big], \tag{5.11b}$$

so that $\bar{u}_j$ is the basis coefficient of $\frac{dg}{d\pi}$ when expanded in the eigenbasis $\{\phi_j\}_{j \geq 1}$. Thus, by Parseval's theorem, we see that equation (5.11a) implies that

$$\tilde{V}_{\sigma,n}^2(\pi, g; \mathbb{B}_{\mathscr{H}}(R)) \leq \mathbb{E}_{\xi^* \cdot \pi}\Big[\big(\frac{dg}{d\pi}(A \mid X)\big)^2 \overline{\sigma}^2\Big] \;=\; V_\sigma^2(\tfrac{dg}{d\pi}), \tag{5.11c}$$

[2] Note that our theory does not require these assumptions, but imposing them makes clear the connection to classical semi-parametric efficiency.

so that the Hilbert-restricted functional is always upper bounded by the classical semi-parametric quantity $V_\sigma^2(\frac{dg}{d\pi})$. In fact, when the semi-parametric efficiency bound is finite and the RKHS is suitably rich—that is, "universal"—then $\tilde{V}_{\sigma,n}^2(\pi, g; \mathbb{B}_{\mathscr{H}}(R))$ converges to $V_\sigma^2(\frac{dg}{d\pi})$ as n tends to infinity. All of these facts hold more generally for heteroskedastic noise, as we detail in Propositions 5.1 and 5.2 to follow in Section 5.3.1.2.

5.3 Main results and their consequences

We now turn to precise statements of our main results, along with discussion of their consequences for various examples. In Section 5.3.1.1, we state and prove non-asymptotic lower bounds that hold for any convex and symmetric function class $\mathcal{F}$ used to model the regression function. We specialize these lower bounds to reproducing kernel Hilbert spaces in Section 5.3.1.2, where we derive the functional (5.10) discussed in the previous section.

In Section 5.3.2, we turn to the complementary question of deriving upper bounds for reproducing kernel Hilbert spaces. We begin with the simpler homoskedastic case in Section 5.3.2.1 before turning to the more challenging heteroskedastic case in Section 5.3.2.2. Finally, Section 5.3.3 is devoted to the consequences of these results for various specific examples.

5.3.1 Non-asymptotic lower bounds

Suppose that we model the regression function μ^* using a class $\mathcal{F}$ of real-valued functions defined on the state-action space $\mathcal{S} \times \mathbb{A}$. In this section, we prove some non-asymptotic minimax lower bounds for both the averaged quantity $\tau^* = \mathscr{L}_\omega(\mu^*, \xi^*)$ and the one-point quantities $\tau_{x_0}^* = \mathscr{L}_\omega(\mu^*, \delta_{x_0})$. In order to cover both cases in a unified way, for any distribution ν over the state space $\mathcal{S}$, let us define[3]

$$V_{\sigma,n}(\nu, \pi, g; \mathcal{F}) := \sqrt{n} \sup_{f \in \mathcal{F}} \Big\{ \mathscr{L}_\omega(f, \nu) \;\Big|\; \mathbb{E}_{\xi^* \cdot \pi}\big[\tfrac{f^2(X,A)}{\sigma^2(X,A)}\big] \leq \tfrac{1}{4n} \Big\}. \tag{5.12}$$

5.3.1.1 General lower bounds

Local minimax bounds describe the behavior of optimal estimators in a local neighborhood of a given instance. For the problem at hand, we define a given problem instance via the pair $\mathcal{I}^* = (\mu^*, \xi^*)$. The behavioral policy π, conditional variance function σ^2, and the weight function g are shared across all instances. Our local neighborhood of a given instance $\mathcal{I}^*$ is given by

$$\mathfrak{N}_n(\mu^*, \xi^*) := \Big\{ \xi \text{ s.t. } \chi^2(\xi \;||\; \xi^*) \leq \tfrac{1}{n}, \text{ and } \mu \in \mathcal{F} \text{ s.t. } \|\mu - \mu^*\|_{\mathbb{L}^2(\xi^* \cdot \pi)}^2 \leq \tfrac{\bar{\sigma}^2}{n} \Big\}, \tag{5.13}$$

[3] To explain our notational choices, in the special case that $\nu = \xi^*$ and $\mathcal{F} = \mathbb{B}_{\mathscr{H}}(R)$, this functional is proportional to the quantity $\tilde{V}_{\sigma,n}^2(\pi, g; \mathbb{B}_{\mathscr{H}}(R))$ that we defined previously, as shown in the sequel (cf. Proposition 5.1 in Section 5.3.1.2).

and it defines the local minimax risk

$$\mathscr{M}_n(\mathcal{I}^*; \mathcal{F}) := \inf_{\widehat{\tau}_n} \sup_{(\mu,\xi) \in \mathfrak{N}_n(\mu^*,\xi^*)} \mathbb{E}\big[\, |\widehat{\tau}_n - \mathscr{L}_\omega(\mu, \xi)|^2 \,\big], \quad \text{and} \tag{5.14a}$$

$$\mathscr{M}_n(\mathcal{I}^*, x_0; \mathcal{F}) := \inf_{\widehat{\tau}_n} \sup_{\mu \in \mathfrak{N}_n(\mu^*,\xi^*)} \mathbb{E}\big[\, |\widehat{\tau}_n - \mathscr{L}_\omega(\mu, \delta_{x_0})|^2 \,\big]. \tag{5.14b}$$

With a slight abuse of notation, in the definition (5.14b), we have written $\mu \in \mathfrak{N}_n(\mu^*, \xi^*)$ to mean $(\mu, \xi^*) \in \mathfrak{N}_n(\mu^*, \xi^*)$. For the rest of this chapter, we will drop $\mathcal{I}^*$ in the notation when it is clear from the context.

In stating lower bounds for estimating $\mathscr{L}_\omega(\mu^*, \xi^*)$, we require that the effective noise in the observations—namely $Z(X) := \big(\int_{\mathbb{A}} \mu^*(X, a) d\omega(a \mid X) - \tau^* \big)$ for $X \sim \xi^*$—has a bounded kurtosis:

$$\|Z\|_{2\to 4} := \frac{\sqrt{\mathbb{E}[Z^4]}}{\mathbb{E}[Z^2]} \leq M_{2\to 4}(\xi^*) < \infty. \tag{5.15}$$

With this condition in place, we are ready to state a lower bound.

Theorem 5.1. *There exists a universal constant c such that for any problem instance $\mathcal{I}^* = (\mu^*, \xi^*)$ with $\mu^* \in \frac{1}{2}\mathcal{F}$:*

(a) Under the moment condition (5.15) and given a sample size $n \geq 16 M_{2\to 4}^2(\xi^*)$, the local minimax risk (5.14a) is lower bounded as

$$\mathscr{M}_n(\mathcal{I}^*, \mathcal{F}) \geq \frac{c}{n} \Big\{ V_{\xi^*}^2(\mu^*) + V_{\sigma,n}^2(\xi^*, \pi, g; \mathcal{F}) \Big\}. \tag{5.16a}$$

(b) Given a sample size $n \geq 16$, the local minimax risk (5.14b) is lower bounded as

$$\mathscr{M}_n(\mathcal{I}^*, x_0; \mathcal{F}) \geq \frac{c}{n}\, V_{\sigma,n}^2(\delta_{x_0}, \pi, g; \mathcal{F}). \tag{5.16b}$$

See Section 5.5.1 for the proof.

The lower bound (5.16a) consists of two terms. The first term $V_{\xi^*}^2(\mu^*)$ captures uncertainty induced by not knowing the distribution ξ^*; in our lower bound, we obtain it by applying the Le Cam argument to allowable perturbations of ξ^*. The second term captures the effective noise induced by a combination of the additive noise, and potential lack of coverage of the behavioral policy π.

The reader should observe the contrast between the bound (5.16a), applicable to a ξ^*-averaged functional, and the bound (5.16b) that applies to a one-point functional. The latter bound takes a similar form, except that the term $V_{\xi^*}^2(\mu^*)$ no longer appears. Here knowledge of ξ^* is irrelevant, because the functional to be estimated is known, and does not depend on it.

It is worth emphasizing that Theorem 5.1 and its corollaries are all stated for a *fixed* behavior policy π. Accordingly, the stated lower bounds apply even to "oracle" estimators that know the behaviorial policy. In practice, this function often not known, especially for observational studies in causal inference. However, as we show in the following section, when we specialize to reproducing kernel Hilbert spaces (cf. Theorems 5.2 and 5.3 to follow), this lower bound can achieved (up to universal constants) via a simple procedure that operates without any knowledge of the policy π. Thus, a surprising consequence of our theory is that, at least for the RKHS case, knowledge of the behavioral policy π has no effect on the minimax risk. This statement is *not* true in general, as demonstrated by past work on Hölder classes [179, 96].

We also note that the lower bounds in Theorem 5.1 are related to past work for estimating linear functionals in fixed design regression (e.g., [193, 195, 53]). As in this work, the quantity (5.12) can be seen as a modulus of continuity for the functional $f \mapsto \mathscr{L}_\omega(f, \nu)$. Our work deals instead with a random design setting, so that the proof techniques are different. Moreover, it is not always possible to achieve the lower bounds in Theorem 5.1; in particular, as we noted above, for certain types of Hölder classes and unknown behavioral policies, sharp lower bounds require an argument that involves mixtures (as opposed to the two-point Le Cam argument that underlies Theorem 5.1).

5.3.1.2 Explicit representation for reproducing kernels

As noted in Section 5.2.2, when $\mathcal{F}$ is a reproducing kernel Hilbert space (RKHS), our minimax lower bounds are sharp (up to a constant pre-factor), and the optimal risk has an explicit expression. To set up the problem, we consider functions belonging to a subset of $\mathbb{L}^2(\mathbb{P}^*)$ where $d\mathbb{P}^*(x,a) = d\xi^*(x) d\pi(a \mid x)$ is a distribution over $\mathbb{U} := \mathcal{S} \times \mathbb{A}$. In particular, let $\mathcal{K}$ be a real-valued kernel function defined on the Cartesian product space $\mathbb{U} \times \mathbb{U}$. We assume that the kernel function is continuous and positive semi-definite, and we let $\mathbb{H}$ be the associated reproducing kernel Hilbert space (RKHS). Associated with the kernel function is the kernel integral operator

$$f \mapsto \mathcal{K}(f)(z) := \int_{\mathbb{U}} \mathcal{K}(u, u') f(z') d\mathbb{P}^*(u')$$

By Mercer's theorem [141], under mild regularity conditions, this operator has real eigenvalues $\{\lambda_j\}_{j=1}^\infty$, all of which are non-negative due to the assumption of positive semidefiniteness, along with eigenfunctions $\{\phi_j\}_{j=1}^\infty$ that are orthonormal in $\mathbb{L}^2(\mathbb{P}^*)$. Under such notations, the minimax risk can be represented in terms of these sequences.

Recall the definition (5.12) of the quantity $V_{\sigma,n}(\nu, \pi, g; \mathcal{F})$, where $\nu = \xi^*$ or $\nu = \delta_{x_0}$ are the two cases of primary interest in this chapter. For any distribution ν over the state space, we define the infinite-dimensional vector $\bar{u}(\nu)$ with components

$$\bar{u}_j(\nu) := \mathbb{E}_{X \sim \nu} \Big[\int_{\mathbb{A}} \phi_j(X, a) d\omega(a \mid X) \Big]. \tag{5.17a}$$

This vector is a generalization of our previous definition (5.9a), which was specialized to $\nu = \xi^*$. We also recall from equation (5.9b) the infinite-dimensional matrix $\mathbf{\Gamma}_\sigma$ with elements

$$[\mathbf{\Gamma}_\sigma]_{jk} := \mathbb{E}_{\xi^* \cdot \pi}\Big[\tfrac{1}{\sigma^2(X,A)} \phi_j(X,A) \phi_k(X,A)\Big], \tag{5.17b}$$

and the diagonal matrix $\mathbf{\Lambda} = \operatorname{diag}\{\lambda_j\}_{j=1}^\infty$. With these definitions, we have

Proposition 5.1. *For the RKHS ball $\mathbb{B}_{\mathscr{H}}(R) := \big\{f \mid \|f\|_{\mathbb{H}} \leq R\big\}$ and any distribution ν over the state space, we have*

$$\frac{1}{2}\, V^2_{\sigma,n}(\nu, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) \overset{(a)}{\leq} \bar{u}^T(\nu)\Big(\mathbf{\Gamma}_\sigma + \tfrac{1}{R^2 n}\mathbf{\Lambda}^{-1}\Big)^{-1} \bar{u}(\nu) \overset{(b)}{\leq} 4\, V^2_{\sigma,n}(\nu, \pi, g; \mathbb{B}_{\mathscr{H}}(R)). \tag{5.18}$$

See Section 5.5.2.1 for the proof.

As discussed in Section 5.2.2, the functional is closely related to the classical semi-parametric efficiency bound. The following result makes this connection precise:

Proposition 5.2. *Under the setup of Proposition 5.1, if the RKHS $\mathbb{H}$ is dense in $\mathbb{L}^2(\xi^* \cdot \pi)$ and $v_{\mathrm{semi}}(\mu^*, \frac{dg}{d\pi}) < +\infty$, then we have*

$$\lim_{n \to \infty} V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) = \frac{1}{2}\sqrt{\mathbb{E}_{\xi^* \cdot \pi}\Big[\big(\frac{dg}{d\pi}(A \mid X)\big)^2 \cdot \sigma^2(X,A)\Big]}.$$

See Section 5.5.2.2 for the proof.

5.3.2 Achieving the lower bounds for kernel classes

We now show how the lower bounds in Theorem 5.1 can be achieved when the regression function μ^* is assumed to lie within some reproducing kernel Hilbert space (RKHS). The setup for RKHS can be found in Section 5.3.1.2. Our theory involves these eigenvalues and eigenfunctions via the following notion of *effective dimension*:

$$D(\rho) := \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \sum_{j=1}^\infty \frac{\lambda_j \phi_j^2(x,a)}{\lambda_j + \rho} \qquad \text{for any scalar } \rho > 0. \tag{5.19}$$

Similar notions of effective dimension have been used in past work [230, 33]. Roughly speaking, the quantity $D(\rho)$ provides a characterization of the global complexity of the RKHS at the scale $\rho > 0$.

5.3.2.1 Homoskedastic case

Let us warm up by describing a simpler (but possibly sub-optimal) bound that ignores any possible heteroskedasticity. More specifically, we suppose that the conditional variance function is uniformly bounded as $\sigma^2(x,a) \leq \bar{\sigma}^2$ for all pairs (x,a), and prove results in terms of $\bar{\sigma}$.

In this case, the procedure is very simple to describe, and consists of two steps.

Two-stage procedure: Given a data set of size $2n$, we split it evenly into two sets $(X_i^{(\mathrm{I})}, A_i^{(\mathrm{I})}, Y_i^{(\mathrm{I})})_{i \in [n]}$ and $(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}, Y_i^{(\mathrm{II})})_{i \in [n]}$, each of size n. Each step in our procedure uses one of the data splits.

Stage I: Given a regularization parameter $\rho_n > 0$, compute the kernel ridge regression (KRR) estimate on data split I:

$$\widehat{\mu}_n := \arg\min_{f \in \mathbb{H}} \Big\{ \frac{1}{n} \sum_{i=1}^{n} \big(Y_i^{(\mathrm{I})} - f(X_i^{(\mathrm{I})}, A_i^{(\mathrm{I})}) \big)^2 + \rho_n \|f\|_{\mathbb{H}}^2 \Big\}. \tag{5.20a}$$

Stage II: Use the estimate $\widehat{\mu}_n$ and split II to compute the empirical average

$$\widehat{\tau}_n = \frac{1}{n} \sum_{i=1}^{n} \Big\{ \int_{\mathbb{A}} \widehat{\mu}_n(X_i^{(\mathrm{II})}, a) d\omega(a \mid X_i^{(\mathrm{II})}) \Big\} \tag{5.20b}$$

In practice, so as to make most efficient use of the data, one could also perform a form of cross-fitting (e.g., [41]). However, given that our main goal is to show that the lower bounds from Theorem 5.1 are achieved up to constant factors, it suffices to focus attention on the simpler procedure given here.

Assumptions: In our analysis of this method, we assume that the kernel function $\mathcal{K}$ is κ-uniformly bounded:

$$\sup_{(x,a) \in \mathbb{X} \times \mathbb{A}} \mathcal{K}\big((x,a),(x,a)\big) \leq \kappa^2. \tag{Kbou(κ)}$$

This condition is frequently used in the literature on kernel methods. It is satisfied, for instance, for any continuous kernel function $\mathcal{K}$ on a compact domain $\mathbb{X} \times \mathbb{A}$.

In addition, we assume that the zero-mean noise variables $W(x,a) := Y - \mu^*(x,a)$ are uniformly σ-sub-Gaussian, meaning that for all pairs (x,a), we have

$$\mathbb{E}\big[e^{tW(x,a)}\big] \leq e^{\frac{t^2 \sigma^2}{2}} \qquad \text{for all } t \in \mathbb{R}. \tag{subG(σ)}$$

We are now equipped to state our first main upper bound. It requires that the ratio of sample size and effective dimension at scale ρ_n is lower bounded as

$$\frac{n}{D(\rho_n)} \geq c \frac{\sigma^2}{\bar{\sigma}^2} \log\big(\frac{nR\kappa}{\bar{\sigma}\delta}\big) \cdot \log^2(n) \qquad \text{where } \rho_n = \tfrac{\bar{\sigma}^2}{R^2 n}, \tag{5.21a}$$

for a universal constant $c > 0$. We discuss this condition at more length following the statement.

Our result involves the *higher-order term*

$$\mathcal{H}_n(\delta) := c\frac{\log(1/\delta)}{n}\Big\{\kappa R + \sigma\sqrt{D(\rho_n)}\log(n)\Big\}, \tag{5.21b}$$

where $\delta \in (0,1)$ is a user-specified failure probability. It can be verified that under the sample size condition (5.21a), we have $\mathcal{H}_n(\delta) = o(n^{-1/2})$, so this term is of higher order in the analysis. Finally, the dominant term in our upper bound is the quantity

$$V_{\bar{\sigma},n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R))^2 := \bar{\sigma}^2 \sum_{j=1}^{\infty} \frac{\lambda_j \bar{u}_j^2}{\lambda_j + \frac{\bar{\sigma}^2}{R^2 n}}. \tag{5.21c}$$

Theorem 5.2. *Under the* (Kbou(κ)) *and* (subG(σ)) *conditions, suppose that* $\mu^* \in \mathbb{B}_{\mathscr{H}}(R)$*, there exists a universal constant* $c > 0$*, such that for any* $\delta \in (0,1)$ *and sample size* $2n$ *satisfying the bound* (5.21a)*. Then the two-stage estimate* $\widehat{\tau}_n$ *computed with regularization* $\rho_n = \frac{\bar{\sigma}^2}{R^2 n}$ *satisfies*

$$|\widehat{\tau}_n - \tau^*| \le c\Big\{V_{\xi^*}(\mu^*) + V_{\bar{\sigma},n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R))\Big\}\sqrt{\tfrac{\log(1/\delta)}{n}} + \mathcal{H}_n(\delta), \tag{5.22}$$

with probability at least $1 - \delta$.

See Section 5.5.3 for the proof.

Let us make a few comments about this result, and its connection to our lower bounds.

Comparison with Theorem 5.1: As noted, the dominant term in the bound (5.22) is the first one. When the noise is homoskedastic (i.e., constant conditional variance), then this first term matches the lower bound given in Theorem 5.1 up to constants and the logarithmic factor[4] in the failure probability δ. When the conditional variance function is not constant, then our bound (5.22) no longer matches Theorem 5.1. We rectify this shortcoming in Section 5.3.2.2, where we analyze a more refined four-stage procedure that adapts to heteroskedasticity.

It should be emphasized that the two-stage estimator analyzed in Theorem 5.2 does not require any knowledge of the behavioral policy π. At the same time, as we just described, for homoskedastic noise, it matches the lower bound from Theorem 5.1, which applies even to oracle estimators that know the policy. Thus, we conclude that at least in the special case of an RKHS, knowledge of the behavior policy does not alter minimax risks (apart from possibly in constant factors).

[4] While we have stated a high probability guarantee, a simple modification yields an estimator with mean-squared error guarantees. In particular, since $|\tau^*| \le \sup_{x,a} |\mu^*(x,a)| \le R\sqrt{\kappa}$ by the Cauchy–Schwarz inequality, we can construct a truncated estimator

$$\widetilde{\tau}_n := \mathrm{sgn}\big(\widehat{\tau}_n\big) \cdot \min\big\{|\widehat{\tau}_n|, R\sqrt{\kappa}\big\}.$$

By construction, we have $|\widetilde{\tau}_n - \tau^*| \le |\widehat{\tau}_n - \tau^*|$ almost surely, and since $\widetilde{\tau}_n$ is a bounded random variable, the high-probability bounds established in Theorem 5.2 can be converted to a MSE bound whose leading term matches Theorem 5.1 up to a constant factor.

Tuning parameter: The only tuning parameter in the estimator is the regularization weight $\rho_n = \frac{\bar{\sigma}^2}{R^2 n}$. This choice depends on the signal-to-noise-ratio, as measured by the ratio $R^2/\bar{\sigma}^2$, but does not depend on kernel eigenvalues or other aspects of the problem. We note that this choice also appears in the classical work on linear functional estimation in fixed design settings [193, 195], but the analysis leading to it in our random design setting is quite different.

The decay rate $\rho_n \asymp n^{-1}$ of the regularization parameter is much faster than the standard one required to achieve optimal mean-squared error when estimating the full regression function (c.f. [213], Chapter 13). Consequently, the first stage of our procedure outputs an *under-smoothed* estimate of the regression function μ^*, and using this estimate in the second stage produces an optimal estimate of the functional. This difference arises because the bias-variance trade-off that underlies estimating the functional of μ^* is very different from that associated with estimating the full regression function μ^*. In particular, when estimating a functional, we pay for variance only at the direction of the target functional, whereas the bias induced by regularization wholly appears in the estimation error.

Lower bound on sample size: Finally, let us comment on the required lower bound (5.21a) on the sample size. There are various conditions that ensure (5.21a). For example, in various examples, it is possible to show that the effective dimension satisfies the bound

$$D(\rho) \leq \frac{D_0}{\rho^{1-\omega}} \qquad \text{for some scalar } \omega \in (0, 1]. \tag{5.23}$$

In Section 5.3.3, we discuss various concrete applications in which this growth condition holds. Under the bound (5.23), the sample size condition (5.21a) is satisfied as long as

$$\frac{n}{\log^{3/\omega}\left(\frac{nR\kappa}{\bar{\sigma}\delta}\right)} \geq c D_0^{1/\omega} R^{\frac{2}{\omega}-2} \sigma^{\frac{2}{\omega}} \bar{\sigma}^{2-\frac{4}{\omega}}.$$

In Appendix D.1, we present various conditions under which the effective dimension satisfies a growth condition that ensures the sample size condition (5.21a) can be satisfied. Moreover, in Appendix D.2, we present alternative guarantees that do not rely on any additional growth conditions.

5.3.2.2 Extension to heteroskedasticity

We now turn to the more challenging problem of achieving the minimax optimal risk in the heteroskedastic case. In this case, we propose and analyze a four-stage procedure. Since the conditional variance function is non-constant and unknown, we need to estimate it, and the first two steps of our four-stage procedure are devoted to this task.

Let us provide a high-level perspective. The first stage generates a rough estimate $\widetilde{\mu}_n$ of the regression function μ^*. In the second stage, we first use $\widetilde{\mu}_n$ to compute estimates

$Z_i := \big\{Y_i - \widetilde{\mu}_n(X_i, A_i)\big\}^2$ of the squared noise associated with a new set $\mathcal{S}$ of triples $\{(Y_i, X_i, A_i)\}_{i \in \mathcal{S}}$. We then compute an estimate of the conditional variance function of the form

$$\widehat{\sigma}_n^2 := \mathcal{A}\Big(\{(X_i, A_i, Z_i)\}_{i \in \mathcal{S}}\Big), \tag{5.24}$$

for a suitably chosen estimator $\mathcal{A}$. We allow the conditional variance estimator $\mathcal{A}$ to take different forms depending on the application, so our set-up provides a family of possible procedures, indexed by this choice. In our theory, we require only a relatively mild form of accuracy from the estimator, which we refer to as *robust pointwise accuracy.*

Now let us specify all four stages in more detail. Given sample size $4n$, we split the data evenly into four pieces, and perform the following four steps:

Stage I: Using the first dataset $(X_i^{\mathrm{(I)}}, A_i^{\mathrm{(I)}}, Y_i^{\mathrm{(I)}})_{i=1}^n$ and regularization parameter $\rho_n > 0$, compute the pilot estimate

$$\widetilde{\mu}_n := \arg\min_{\mu \in \mathbb{H}} \Big\{ \frac{1}{n} \sum_{i=1}^n \big(Y_i^{\mathrm{(I)}} - \mu(X_i^{\mathrm{(I)}}, A_i^{\mathrm{(I)}})\big)^2 + \rho_n^{\mathrm{(I)}} \|\mu\|_{\mathbb{H}}^2 \Big\}. \tag{5.25a}$$

Stage II: Using the second dataset $(X_i^{\mathrm{(II)}}, A_i^{\mathrm{(II)}}, Y_i^{\mathrm{(II)}})_{i=1}^n$ and the procedure $\mathcal{A}$, compute the squared noise estimates $Z_i^{\mathrm{(II)}} := (Y_i^{\mathrm{(II)}} - \widetilde{\mu}_n(X_i^{\mathrm{(II)}}, A_i^{\mathrm{(II)}}))^2$ based on the pilot estimate (5.25a), and then compute the estimate

$$\widehat{\sigma}_n^2 := \mathcal{A}\Big(\Big\{X_i^{\mathrm{(II)}}, A_i^{\mathrm{(II)}}, Z_i^{\mathrm{(II)}}\Big\}_{i \in [n]}\Big) \qquad \text{of the conditional variance function.} \tag{5.25b}$$

Stage III: Using the third dataset $(X_i^{\mathrm{(III)}}, A_i^{\mathrm{(III)}}, Y_i^{\mathrm{(III)}})_{i=1}^n$, regularization parameter $\rho_n^{\mathrm{(III)}} > 0$, and the estimated function $\widehat{\sigma}_n^2$, compute the weighted regression estimate

$$\widehat{\mu}_n := \arg\min_{\mu \in \mathbb{H}} \Big\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{\widehat{\sigma}_n^2(X_i^{\mathrm{(III)}}, A_i^{\mathrm{(III)}})} \big(Y_i^{\mathrm{(III)}} - \mu(X_i^{\mathrm{(III)}}, A_i^{\mathrm{(III)}})\big)^2 + \rho_n^{\mathrm{(III)}} \|\mu\|_{\mathbb{H}}^2 \Big\}. \tag{5.25c}$$

Stage IV: Using the fourth dataset $(X_i^{\mathrm{(IV)}}, A_i^{\mathrm{(IV)}}, Y_i^{\mathrm{(IV)}})_{i=1}^n$ and the weighted regression estimate (5.25c), compute the empirical average

$$\widehat{\tau}_n = \frac{1}{n} \sum_{i=1}^n \int \widehat{\mu}_n(X_i^{\mathrm{(IV)}}, a) d\omega(a \mid X_i^{\mathrm{(IV)}}) \tag{5.25d}$$

We remark that the idea of re-weighting with estimated conditional variance has been utilized in literature, in the context of parameter estimation for linear models.

See the paper [180] and references therein for detailed discussion. We now turn to the analysis of the 4-stage procedure. Rather than analyze a particular estimator $\mathcal{A}$ of the conditional variance function, let us lay out an abstract condition that handles a variety of different estimators

Robust pointwise variance estimators: This property is a way of certifying that the estimator $\mathcal{A}$ provides an ε-accurate estimate in a pointwise sense: for any fixed pair (s_0, a_0), with probability at least $1 - \delta$, we have

$$\Big| \mathcal{A}\Big(\big\{ X_i, A_i, Z_i \big\}_{i \in [n]} \Big)(s_0, a_0) - \sigma^2(s_0, a_0) \Big| \leq \varepsilon. \tag{5.26}$$

The key is to quantify how errors in the inputs Z_i as approximations of the squared noise $(Y_i - \mu^*(X_i, A_i))^2$ affect this guarantee. We do so via a pair of functions on the inputs (ε, δ), known as the tolerance function t and sample threshold M respectively.

Definition 1. *The procedure $\mathcal{A}$ is (t, M)-pointwise-robust if for any pair $(\varepsilon, \delta) \in [0, 1]^2$, any dataset $\{Z_i\}_{i=1}^n$ of size $n \geq M(\varepsilon, \delta)$, consisting of variables such that for each $i \in [n]$*[5]

$$\big| \mathbb{E}[Z_i \mid X_i, A_i] - \sigma^2(X_i, A_i) \big| \leq t(\varepsilon, \delta) \quad \textit{and} \quad \| Z_i \mid X_i, A_i \|_{\psi_1} \leq 4\big(\sigma^2 + t(\varepsilon, \delta) \big), \tag{5.27}$$

then for any fixed pair (x_0, a_0), the bound (5.26) holds with probability $1 - \delta$.

There are various estimators that satisfy the robust pointwise risk property; see Appendix D.4 for further discussion.

In order to analyze the 4-stage procedure, we require one additional condition on the conditional variance function: there are scalars $0 < \underline{\sigma} \leq \bar{\sigma} < \infty$ such that

$$\sigma(x, a) \in [\underline{\sigma}, \bar{\sigma}] \quad \text{for all } (x, a). \tag{σ-INT}$$

We also require that the sample size satisfies the lower bounds

$$\frac{n}{D(\rho_n^{(\mathrm{I})})} \geq \Big\{ c \frac{\sigma^2}{\underline{\sigma}^2} \log \big(\frac{n R \kappa}{\underline{\sigma} \delta} \big) \cdot \log^2 n \Big\}, \tag{5.28a}$$

$$n \geq M\big(\underline{\sigma}/2, \delta/(2n) \big) \quad \text{and} \quad \frac{n}{\sigma D(\rho_n^{(\mathrm{I})}) \log(n/\delta)} \geq \frac{1}{t\big(\frac{\underline{\sigma}^2}{2}, \frac{\delta}{2n} \big)}. \tag{5.28b}$$

With this set-up, we are now ready to state a guarantee on our 4-state procedure. Note that it has two tuning parameters: the regularization parameter $\rho_n^{(\mathrm{I})}$ from the first stage regression, and the regularization parameter $\rho_n^{(\mathrm{III})}$ from the weighted regression in the third stage. Our guarantee applies to the procedure using the parameters

$$\rho_n^{(\mathrm{I})} = \frac{\bar{\sigma}^2}{R^2 n} \quad \text{and} \quad \rho_n^{(\mathrm{III})} = \frac{1}{R^2 n}. \tag{5.29}$$

[5] In the definition (5.27), the quantity σ is the sub-Gaussian parameter (cf. condition (subG(σ))), whereas $\| \cdot \|_{\psi_1}$ is the Orlicz(1)-norm, or sub-exponential parameter.

Theorem 5.3. *In addition to the assumptions of Theorem 5.2, suppose that the conditional variance function satisfies the interval condition (σ-INT), the estimator $\mathcal{A}$ is (t, M)-robust, and the sample size $4n$ satisfies the lower bounds (5.28a) and (5.28b). Then using regularization parameters from equation (5.29), the 4-stage procedure yields an estimate $\widehat{\tau}_n$ such that*

$$|\widehat{\tau}_n - \tau^*| \leq c\Big\{V_{\xi^*}(\mu^*) + V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R))\Big\}\sqrt{\frac{\log(1/\delta)}{n}} + \mathcal{H}_n(\delta), \tag{5.30}$$

with probability $1-\delta$, where the higher-order term $\mathcal{H}_n(\delta)$ was previously defined (5.21b).[6]

See Section 5.5.4 for the proof.

A few remarks are in order. First, Theorem 5.3 is adaptive to the heteroskedastic nature of the observation noise — the term $V^2_{\sigma,n}(\xi^*, \pi, g; \mathcal{F})$ involves the actual conditional variance σ, instead of its uniform upper bound $\bar{\sigma}$. With such a fine-grained variance, the bound (5.30) achieves the instance-dependent optimality result in Theorem 5.1, up to universal constants and high-order terms. (See footnote 4 for the connection between high-probability bounds and mean-squared error bounds.)

We note that the sample size condition (5.28a) is slightly stronger than the condition (5.21a) used in Theorem 5.2, with the variance upper bound $\bar{\sigma}^2$ in the denominator replaced by the lower bound $\underline{\sigma}^2$. Theorem 5.3 further requires an additional sample size condition (5.28b), which comes from the sample complexity of the robust pointwise estimator $\widehat{\sigma}_n$.

Extension to estimating one-point functionals: Now we extend our results to estimating the one-point functional $\tau^*(x_0) := \mathscr{L}_\omega(\mu^*, \delta_{x_0})$. In this case, the target functional is known, so that the fourth stage of the four-stage procedure is not necessary. It suffices to split the data into three folds in total, and we plug in the regression function $\widehat{\mu}_n$ directly to obtain the estimate

$$\widehat{\tau}_n(x_0) := \mathscr{L}_\omega(\widehat{\mu}_n, \delta_{x_0}) = \int_{\mathbb{A}} \widehat{\mu}_n(x_0, a) d\omega(a \mid x_0). \tag{5.31}$$

This estimate satisfies optimal guarantees matching Theorem 5.1(b) up to a constant factor. In particular, under the setup of Theorem 5.3, for any $x_0 \in \mathbb{X}$, we have

$$|\widehat{\tau}_n(x_0) - \tau^*(x_0)| \leq c \cdot V_{\sigma,n}(\delta_{x_0}, \pi, g; \mathbb{B}_{\mathscr{H}}(R))\sqrt{\frac{\log(1/\delta)}{n}}, \tag{5.32}$$

with probability $1 - \delta$. See Section 5.5.4.4 for the proof.

A few remarks are in order. First, the upper bound in equation (5.32) matches the local minimax lower bound in Theorem 5.1(b) up to universal constant factors, exhibiting

[6] We take $\rho_n = \rho_n^{(\mathrm{I})}$ in its expression.

its optimality in an instance-dependent sense.[7] As opposed to Theorems 5.2 and 5.3, the optimal instance-dependent risk is achieved (up to universal constants) without additional high-order terms. The choice of parameters and sample size requirement in (5.32) is exactly the same as the one in Theorem 5.3, and does not depend on the query point x_0. Such an adaptive property makes the estimator useful in practice, allowing for a plug-and-play approach: one only needs to run stages I–III of the four-stage framework (5.25), and generate an estimator $\widehat{\mu}_n$. By substituting such an estimator in equation (5.25d) using another fold of data, or in equation (5.31) for any query point x_0, optimal and adaptive guarantees can always be achieved.

5.3.3 Consequences for some concrete examples

In this section, we develop some consequences of our general theory for some specific classes of problems, including the missing data problem without overlap (Section 5.3.3.1), for which we presented an illustrative simulation in Section 5.1.

5.3.3.1 A missing data example without overlap assumption

Consider the classical missing data setting, where the action space is $\mathbb{A} = \{0, 1\}$ and the weight is given by $\omega(a \mid x) = a$ for any $x \in \mathbb{X}$. Assume without loss of generality that $Y = 0$ whenever $A = 0$. We slightly abuse the notation to use $\mu^* : \mathbb{X} \to \mathbb{R}$ to denote the outcome function $\mu^*(\cdot, 1)$ and use $\pi : \mathbb{X} \to [0, 1]$ to denote the propensity score $\pi(1 \mid \cdot)$. Similarly, we use $\mathcal{K}(x, x')$ to denote $\mathcal{K}(x, 1), (x', 1))$, and let the kernel function be 0 if one of the arguments has action equal to 0. Under this simplified notation, the inner product of $\mathbb{L}^2(\xi^* \cdot \pi)$ takes the form

$$\langle f_1,\, f_2 \rangle := \int_{\mathbb{X}} f_1(x) f_2(x) \pi(x) d\xi^*(x),$$

and we are interested in estimating the average treatment effect and its conditional analogue

$$\tau^* := \mathbb{E}_\xi \big[\mu^*(X)\big], \quad \text{and} \quad \mathscr{L}_\omega(\mu^*, \delta_{x_0}) := \mu^*(x_0).$$

For concreteness, we let the state space be a unit interval $\mathbb{X} = [0, 1]$ and take the input distribution ξ be the uniform distribution on $\mathbb{X}$. In order to illustrate the effect of the lack of the overlap condition on the risk, given a scalar $\alpha > 0$, we construct the following propensity score function

$$\pi(x) = (1 - x)^\alpha \quad \text{for any } x \in [0, 1]. \tag{5.33}$$

Our goal is to understand the effect of a singularity in the importance ratio with local α-th order polynomial growth. The specific location of such singularity, and any

[7] Following the discussion in footnote 4, the high-probability bound can be readily converted into a mean-squared error bound using a simple truncation method.

properties apart from the existence of this α-th order singularity are not germane to our comparison, so that we have chosen the particular form (5.33) for technical convenience.

We consider an RKHS $\mathbb{H}$ corresponding to the first-order Sobolev space on $[0, 1]$ (see [213], Chapter 12). Its kernel function is given by $\mathcal{K}(x, x) = \min\{x, x'\}$, and the corresponding RKHS $\mathbb{H}$ consists of functions f satisfying $f(0) = 0$, and

$$\|f\|_{\mathbb{H}}^2 := \int_0^1 \big(f'(x)\big)^2 dx < +\infty.$$

We assume that the regression function $\mu^* : \mathbb{X} \to \mathbb{R}$ belongs to this RKHS, with $\|\mu^*\|_{\mathbb{H}} \leq 1/2$. Finally, we take the conditional variance as $\sigma^2(x, a) \equiv 1$ for any state-action pair (x, a), and assume that the sub-Gaussian parameter σ is of order one.

With this set-up, we are ready to compute minimax rates for various linear functionals. Throughout this section, we use the notation $a_n \asymp b_n$ to denote that the ratio a_n/b_n satisfies finite positive upper and lower bounds depending on the constants (α, x_0) but independent of n. As mentioned before, we omit the problem instance $\mathcal{I}^*$ in the notations $\mathscr{M}_n\big(\mathcal{I}^*, \mathcal{F}\big)$ and $\mathscr{M}_n\big(\mathcal{I}^*, x_0; \mathcal{F}\big)$ for minimax risk rates.

Corollary 5.1. *Under the above set-up, for any function $\mu^* \in \mathbb{B}_{\mathbb{H}}(1/2)$, the minimax risk for estimating the linear functional τ^* is given by*

$$\mathscr{M}_n\big(\mathbb{B}_{\mathbb{H}}(1)\big) \asymp \begin{cases} n^{-1} & \alpha < 1, \\ n^{-1}\log n & \alpha = 1, \\ n^{-\frac{3}{\alpha+2}} & \alpha > 1. \end{cases} \tag{5.34a}$$

For the one-point functional $\mathscr{L}_\omega(\mu^*, \delta_{x_0})$, we have

$$\mathscr{M}_n\big(x_0; \mathbb{B}_{\mathbb{H}}(1)\big) \asymp \begin{cases} 0 & x_0 = 0, \\ n^{-1/2} & x_0 \in (0, 1), \\ n^{\frac{-1}{2+\alpha}} & x_0 = 1, \end{cases} \tag{5.34b}$$

See Appendix D.5.1 for the proof.

A few remarks are in order. For the average treatment effect τ^*, the optimal rate of estimation exhibits a phase transition depending on the local growth exponent α. In the regime $\alpha \in [0, 1)$, the importance ratio is sufficiently well-behaved that the classical quantity v_{semi} is finite, so that we obtain convergence at the classical $\sqrt{n}$-rate. Slower rates arise once $\alpha \geq 1$, where the variance v_{semi}^2 is infinite. A large value of α yields fewer observations in the neighborhood of $x = 1$, which in turn leads to slower rate of convergence. Note that even if the target τ^* is defined as a global average over the interval $[0, 1]$, the optimal rate of convergence is still affected by the singularity within the interval.[8] Finally, we note that although Corollary 5.1 exhibits a wide spectrum of

[8] The proofs in Appendix D.5.1 can be easily extended to propensity score functions with zeros at any finite subset of $[0, 1]$, with arbitrary behavior except for the local growth conditions around the zeros.

rates, they all can be achieved adaptively—that is, using an estimator that requires no knowledge of the behavioral policy π nor the exponent α.

Let us make a few comments on the conditional average treatment effect $\mathscr{L}_\omega(\mu^*, \delta_{x_0})$. In this case, the problem becomes trivial at $x_0 = 0$, as the functions in the Sobolev space $\mathbb{H}$ satisfy $\mu^*(0) = 0$. The optimal rate is $n^{-1/2}$ for any $x_0 \in (0, 1)$, which corresponds to the minimax one-point rate for Sobolev regression in literature [206]. A much slower minimax rate is observed at $x_0 = 1$, where the scarcity of outcome observations is controlled by the exponent α. These rates, just as in the ATE case, can be achieved using an estimator without any knowledge of the function π.

5.3.3.2 Off-policy evaluation with continuous actions

Now we consider a continuum-arm bandit setup. For simplicity, we work with the state space $\mathbb{X} = [0, 1]^{d_x}$ and the action space $\mathbb{A} = [0, 1]^{d_a}$, and let the distributions ξ, $\pi(\cdot \mid x)$ be the uniform distribution on the spaces $\mathbb{X}$ and $\mathbb{A}$, respectively, for any $x \in \mathbb{X}$. Given a scalar $s > (d_x + d_a)/2$, we let the RKHS $\mathbb{H} = \mathbb{H}^s$ be the Sobolev space of order s, with periodic boundary conditions (so that the state-actions spaces are seen as tori).

Under this setup, the eigenfunctions are given by the standard (complex) Fourier bases on the torus $\mathbb{T}^{d_x+d_a}$, which can be written in a product form

$$\Big\{(x, a) \mapsto \phi_j(x)\psi_k(a)\Big\}_{j,k\geq 0},$$

where $\{\phi_j\}_{j\geq 0}$ and $\{\psi_k\}_{k\geq 0}$ are the Fourier bases on the tori $\mathbb{T}^{d_x}$ and $\mathbb{T}^{d_a}$, respectively. Note that these eigenfunctions are uniformly bounded in sup norm.

Throughout this section, we view the problem parameters (d_x, d_a, s) as universal constants, and suppress any constant factor depending only on them. For the Sobolev space $\mathbb{H}^s$, let $\lambda_{j,k}$ be the eigenvalue associated to the eigenfunction indexed by j, k, which satisfies the decay condition (see [12])

$$\lambda_{j,k} \asymp \min\Big\{j^{-2s/d_x}, k^{-2s/d_a}\Big\}. \tag{5.35}$$

Combining the eigendecay assumption and the boundedness condition on the eigenfunctions, we can verify that condition (Kbou(κ)) holds; in particular, we have

$$\kappa^2 := \sup_{x,a} \sum_{j,k\geq 0} \lambda_{j,k}\phi_j^2(x)\psi_k^2(a) \leq \sup_{x,a} \sum_{j,k\geq 0} \lambda_{j,k} < \infty,$$

where the last inequality follows from the fact $s > (d_a + d_x)/2$.

Given a deterministic target policy $T : \mathbb{X} \to \mathbb{A}$, we let $\omega(\cdot \mid x)$ be the atomic measure on $T(x)$ for $x \in \mathcal{S}$, so that the linear functionals of interest take the following form:

$$\tau^* := \int_{\mathbb{X}} \mu^*(x, T(x))dx, \quad \text{and} \quad \mathscr{L}_\omega(\mu^*, \delta_{x_0}) := \mu^*(x_0, T(x_0)).$$

Finally, we let the conditional variance function be unity $\sigma^2 \equiv 1$, and assume that the sub-Gaussian parameter σ is of order one. Let μ^* be any function lying in the Hilbert ball $\mathbb{B}_{\mathbb{H}}(1/2)$.

Note that in this example, the importance ratio $\frac{dg}{d\pi}$ is not well-defined, as the measure $\omega(\cdot \mid x)$ is atomic, for any $x \in \mathcal{S}$. Nevertheless, estimation is still possible, and our general frameworks provide precise characterization of the minimax risks, stated as follows.

Corollary 5.2. *Under the above setup, we have*

$$\mathscr{M}_n\big(\mathbb{B}_{\mathbb{H}}(1)\big) \asymp \frac{\mathrm{var}_\xi\left(\mu^*(X, T(X))\right)}{n} + \sum_{j,k \geq 1} \frac{\left|\langle \phi_j,\, \psi_k \circ T\rangle\right|^2}{n + j^{2s/d_x} + k^{2s/d_a}}, \tag{5.36a}$$

$$\mathscr{M}_n\big(x_0; \mathbb{B}_{\mathbb{H}}(1)\big) \asymp n^{\frac{d_a+d_x}{2s}-1}, \quad \textit{for any } x_0 \in \mathbb{T}^{d_x}. \tag{5.36b}$$

Furthermore, under the worst-case target policy, we have

$$\sup_T \mathscr{M}_n\big(\mathbb{B}_{\mathbb{H}}(1)\big) \asymp n^{\frac{d_a}{2s}-1}. \tag{5.36c}$$

See Appendix D.5.2 for the proof.

A few remarks are in order. For the one-point functional $\mathscr{L}_\omega(\mu^*, \delta_{x_0})$, the optimal rate given by equation (5.36b) is exactly the optimal rate for estimating a $(d_a + d_x)$-dimensional Sobolev function at the point $(x_0, T(x_0))$. For the averaged functional τ^*, in the worst case, we only need to pay for the dimension d_a of the action space, due to the averaging effect in the state space. Moreover, the precise complexity for estimation is characterized by equation (5.36a), which depends on the behavior of the target policy T. Such an instance-optimal risk is achieved by the estimator $\widehat{\tau}_n$. Finally, we remark that though the statement of Corollary 5.2 is for a deterministic target policy T, the result naturally extends to general randomized target policies.

5.4 Simulation studies

In this section, we present the results of some simulation studies in which we compare our procedures with other methods. In particular, we perform experiments on two classes of missing data problems, one defined by the family of singular importance ratios discussed in Section 5.3.3.1 and the heavy-tailed example proposed by Khan and Tamer [100], with some generalizations. These two examples allow us to explore two different ways in which unbounded importance ratios can arise.

Concretely, we perform experiments in which the goal is to estimate the treatment effect based on missing data. Let the state space be the real line $\mathbb{X} = \mathbb{R}$, and let the action space be binary, $\mathbb{A} = \{0, 1\}$. We use the action $a \in \mathbb{A}$ to model missingness, so that we only observe the outcome Y if and only if $A = 1$. For simplicity, we slightly abuse notation, and let μ^* denote the function $\mu^*(\cdot, 1)$. Similarly, we use π to denote the function $\pi(1 \mid \cdot)$. The goal is to estimate the linear functional (5.1) with $\omega(a \mid x) = a$, i.e.,

$$\tau^* = \mathbb{E}_\xi\big[\mu^*(X)\big].$$

Throughout this section, we consider the homoskedastic case with $\sigma^2(x, a) \equiv 1$. By equation (5.8), the semi-parametric efficiency bound for this problem takes the form

$$v_{\mathrm{semi}}^2 = \mathrm{var}_{\xi^*}\big(\mu^*(X)\big) + \int_{-\infty}^{\infty} \frac{\xi(x)}{\pi(x)} dx, \tag{5.37}$$

which may or may not be finite.

For the rest of this section, we describe and discuss the construction of simulation problem instances, as well as various choices of estimators under our consideration. We then present the simulation results.

Four possible estimators: We compare the performance of four possible estimators for the average treatment effect — two of which are based on inverse propensity weights, while the other two (including our estimator) are based on outcome regression.

First, we consider the naïve inverse propensity weighting (IPW) estimator, defined as

$$\widehat{\tau}_{n,\mathrm{ipw}} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i A_i}{\pi(X_i)}. \tag{5.38}$$

Note that the estimator $\widehat{\tau}_{n,\mathrm{ipw}}$ always has finite expectation, with $\mathbb{E}[\widehat{\tau}_{n,\mathrm{ipw}}] = \tau^*$. Assuming that the outcome functions are bounded, the variance of $\widehat{\tau}_{n,\mathrm{ipw}}$, if exists, is given by

$$\mathbb{E}\big[\,|\widehat{\tau}_{n,\mathrm{ipw}} - \tau^*|^2\,\big] \asymp n^{-1}\Big(v_{\mathrm{semi}}^2 + \mathbb{E}_\xi\big[\frac{1 + [\mu^*(X)]^2}{\pi(X)}\big]\Big) \asymp \frac{1}{n}\int_{-\infty}^{\infty} \frac{\xi(x)}{\pi(x)} dx.$$

In general, if the second moment does not exist, the naïve IPW estimator may converge to a heavy-tailed stable law, at a rate slower than $\sqrt{n}$. (c.f. [86], Chapter 14)

Khan and Tamer [100] suggested improving the naïve IPW by removing data with extremely small propensity scores. Given a truncation level γ_n, we define the estimator

$$\widehat{\tau}_{n,\mathrm{trunc}} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i A_i}{\pi(X_i)} \mathbf{1}[\pi(X_i) \geq \gamma_n], \tag{5.39}$$

where $\mathbf{1}[\pi(X_i) \geq \gamma_n]$ is equal to 1 when $\pi(X_i) \geq \gamma_n$, and zero otherwise.

Now we turn to the outcome-regression estimators based on kernel ridge regression, as defined in the two-stage framework (5.20). In order to improve the universal constant factors (which are not covered by our theory), we use a cross-fit procedure, i.e., we generate an estimator $\widehat{\tau}_n^{(\mathrm{I})}$ from the framework (5.20). By switching the role of $(X_i^{(\mathrm{I})}, A_i^{(\mathrm{I})}, Y_i^{(\mathrm{I})})$

and $(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}, Y_i^{(\mathrm{II})})$ and applying the same two-stage framework, we obtain another estimator $\widehat{\tau}_n^{(\mathrm{II})}$, and the final estimator is given by

$$\widehat{\tau}_n = \frac{1}{2}\big(\widehat{\tau}_n^{(\mathrm{I})} + \widehat{\tau}_n^{(\mathrm{II})}\big). \tag{5.40}$$

In terms of the regularization parameter ρ_n, we consider two possible choices:

- Optimal choice: based on the optimal theoretical prediction in equation (5.21a), we set $\rho_n = \frac{0.5}{n}$ for any $n > 0$. We call this estimator $\widehat{\tau}_{n,\mathrm{opt}}$.
- Cross validation: for each sample size n, we use cross validation to find the regularization parameter that minimizes the mean-squared error in predicting μ^*. We call this estimator $\widehat{\tau}_{n,\mathrm{cv}}$.

It is worth noticing that the optimal choice of the regularization parameter ρ_n for estimating the scalar τ^* does *not* correspond to the optimal choice in estimating the regression function μ^*. Indeed, as we will see in the simulation results, the common cross-validation approach in non-parametric estimation leads to sub-optimal semi-parametric performance under our framework, and under-smoothing is crucial to the optimal guarantees.

5.4.1 Simulation results with heavy-tailed covariates

We first present the simulation setup and results on the heavy-tailed covariate examples proposed by Khan and Tamer [100].

Model set-up: We consider the following choices for the distribution ξ^* over data:

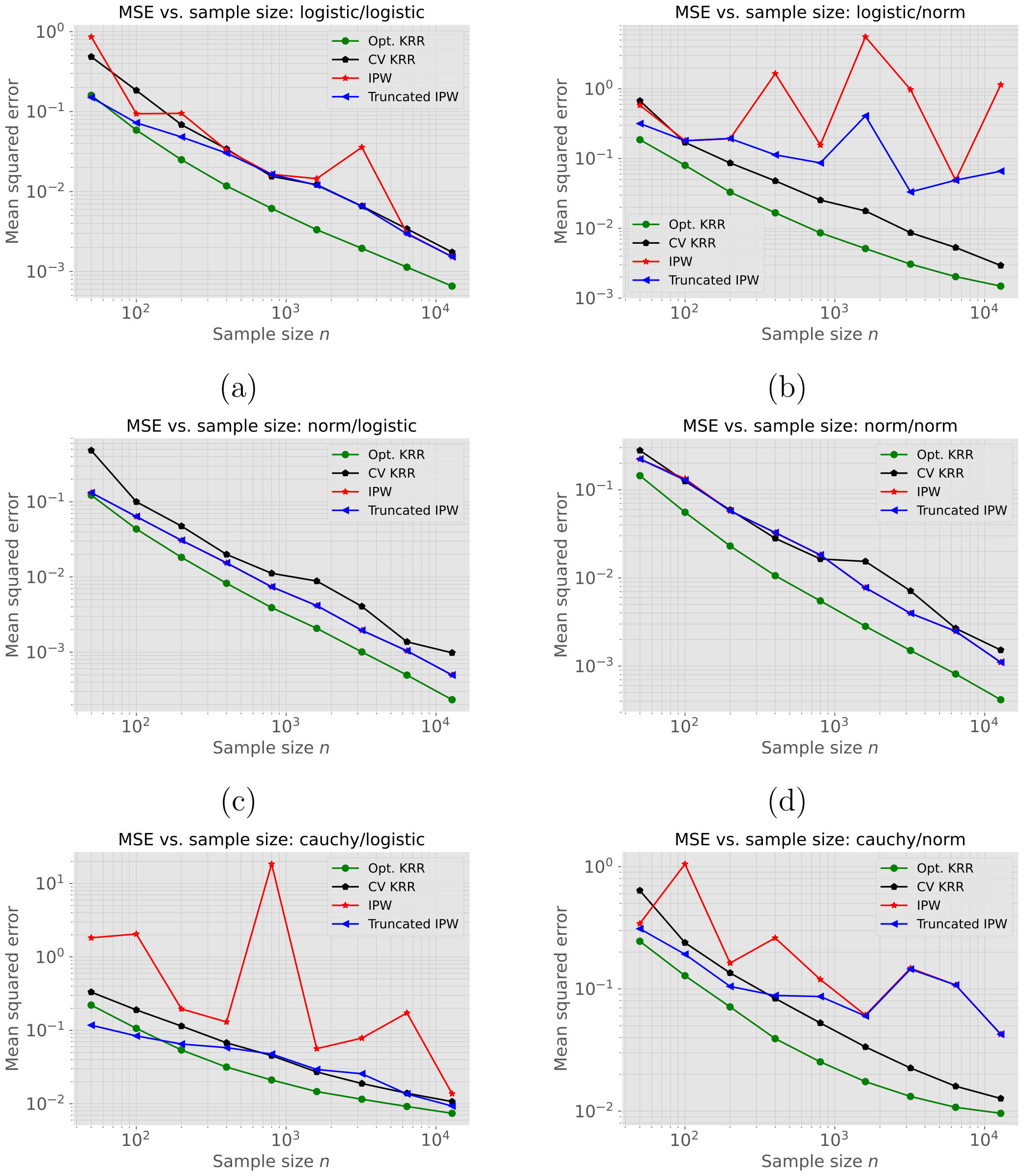
$$\begin{aligned}
\text{Standard normal:} \quad & \xi^{\mathrm{N}}(x) = \frac{1}{\sqrt{2\pi}} \exp\big(-x^2/2\big),\\
\text{Standard logistic:} \quad & \xi^{\mathrm{L}}(x) = \big(e^{x/2} + e^{-x/2}\big)^{-2},\\
\text{Standard Cauchy:} \quad & \xi^{\mathrm{C}}(x) = \frac{1}{\pi(1+x^2)}.
\end{aligned}$$

Among these choices, the normal distribution possesses the lightest tail, while the tail of the Cauchy distribution is the heaviest.

We carry out our simulation studies using the regression function

$$\mu^*(x) = 1 + \cos(x) \quad \text{for } x \in \mathbb{R}.$$

This specific choice is not essential to our study; we have simply chosen a bounded and smooth regression function. Note that many estimators under our consideration involve shrinkage, regularization, or truncation steps, which make the output contract towards 0. In order to ensure a fair comparison, we include offset 1 so that the regression function is non-negative, and the target functional is bounded away from zero.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 5.2: Plots of the mean-squared error $\mathbb{E}\big[\,|\widehat{\tau}_{n,\diamond} - \tau^*|^2\,\big]$ versus sample size n. Each curve corresponds to a different algorithm $\diamond \in \big\{\text{ipw}, \text{trunc}, \text{opt}, \text{cv}\big\}$. Each marker corresponds to a Monte Carlo estimate based on the empirical average of 2000 independent runs. For the cross-validated estimator $\widehat{\tau}_{n,\text{cv}}$, the choice of regularization parameter is based on averaging the cross validation results of the first 50 runs. As indicated by the subfigure titles, each panel corresponds to a problem setup $(\xi, \pi) \in \big\{\xi^{\text{L}}, \xi^{\text{N}}, \xi^{\text{C}}\big\} \times \big\{\pi^{\text{L}}, \pi^{\text{N}}\big\}$. Both axes in the plots are given by logarithmic scales. Some of the curves may overlap with each other.

In order to implement the kernel-based procedures, we use a Laplacian kernel

$$\mathcal{K}(u, u') := \exp\big(-2|u - u'|\big), \quad \text{for } u, u' \in \mathbb{R}.$$

For the behavioral policy π, we use the cumulative distribution functions of logistic and normal distributions, respectively.

$$\pi^{\mathrm{N}}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\big(-y^2/2\big) dy, \quad \text{and} \quad \pi^{\mathrm{L}}(x) = \frac{1}{1 + e^x}.$$

Note that for both choices, the value $\pi(x)$ approaches 0 as x increases; the rate of decay is faster under the normal model than the logistic model.

The paper [100] considers the density function $\xi = \xi^{\mathrm{L}}$, along with propensity score functions $\pi \in \big\{\pi^{\mathrm{N}}, \pi^{\mathrm{L}}\big\}$. Under both setups, the semi-parametric efficiency bound derived in equation (5.37) are infinite, while certain rates of convergence are still achieved via truncation-based estimators (see Section 4.1 of [100] for details). Khan and Tamer proposed truncating at the threshold $X_i \leq \sqrt{\log n}$ for $\pi = \pi^{\mathrm{N}}$, and $X_i \leq \log n$ for $\pi = \pi^{\mathrm{L}}$. Indeed, they are the thresholds that ensures that the truncated inverse propensity weight is uniformly bounded by a polynomial of n, under propensity scores π^{L} and π^{N}, respectively. In our simulation studies, we consider all possible combinations of $\xi \in \big\{\xi^{\mathrm{N}}, \xi^{\mathrm{L}}, \xi^{\mathrm{C}}\big\}$ and $\pi \in \big\{\pi^{\mathrm{N}}, \pi^{\mathrm{L}}\big\}$. We choose $\gamma_n = \pi(\log n)$ under the logistic propensity score π^{L}, and $\gamma_n = \pi(\sqrt{\log n})$ under the normal propensity score π^{N}, which yield near-optimal truncation levels, regardless of the choice of data distribution ξ. Intuitively, heavier tail of the distribution ξ and lighter tail of the propensity score π together lead to less regular behavior for estimators based on important weighting.

Among our simulation setups, the only case that yields finite v_{semi} is that of $(\pi = \pi^{\mathrm{L}}, \xi = \xi^{\mathrm{N}})$; other recent work [197, 82] has also studied this particular configuration. For the other five setups, the classical theories for $\sqrt{n}$-consistency and semi-parametric efficiency are not available, due to the singular behavior of propensity scores.

Simulation results: In Figure 5.2, we demonstrate the simulation results for different estimators under aforementioned setups. The sample size varies within the range $n \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$, and the mean-squared error is estimated through empirical average over 2000 independent runs.

From our simulation results, it can be observed that our estimator $\widehat{\tau}_{n,\mathrm{opt}}$ consistently outperforms other three baselines. When tuning the regularization parameter using cross validation, however, the estimator $\widehat{\tau}_{n,\mathrm{cv}}$ performs significantly worse, over all the simulation instances. This shows that under-smoothing is crucial to the performance of outcome-regression estimators, and that the optimal bias-variance trade-off in function and scalar estimation problems are drastically different. The truncated IPW estimator also yields a reasonable and robust performance, but in most settings, its rate of convergence (represented as the slope of the curve in log-log plot) is worse than $\widehat{\tau}_{n,\mathrm{opt}}$. The naïve IPW estimator, on the other hand, can be highly unstable, especially for heavy-tailed data distributions ξ^{L} and ξ^{C}. Finally, we remark that the two classes of estimators are not comparable in general, as they use different information — the IPW-based estimators $\widehat{\tau}_{n,\mathrm{ipw}}$ and $\widehat{\tau}_{n,\mathrm{trunc}}$ use the information of the true propensity score π, which is not needed for $\widehat{\tau}_{n,\mathrm{opt}}$ and $\widehat{\tau}_{n,\mathrm{cv}}$; on the other hand, the outcome regression estimators $\widehat{\tau}_{n,\mathrm{opt}}$ and $\widehat{\tau}_{n,\mathrm{cv}}$ require the treatment effect function to lie in an RKHS, while the truncated IPW estimator $\widehat{\tau}_{n,\mathrm{trunc}}$ only requires it to be bounded.

It is also useful to discuss the difference in the performance of estimators under various setups. In the classical $\sqrt{n}$-regime with $\pi = \pi^{\mathrm{L}}$ and $\xi = \xi^{\mathrm{N}}$, the truncation does not happen with high probability, and the naïve IPW estimator yields the same MSE as the truncated one, as shown in Figure 5.2(c). In other five cases, the estimation error of $\widehat{\tau}_{n,\mathrm{ipw}}$ is unstable, and worse than the truncated analogue. It can be observed from that the slopes of the green curves are around 1 in the log-log plots in panels (a)–(d) of Figure 5.2, but are much flatter in panels (e) and (f). This observation suggests that the optimal rate of convergence may be near-parametric under the logistic and normal model, while a slower minimax rate could be unavoidable in the Cauchy setting.

5.4.2 Simulation results with singular importance ratio

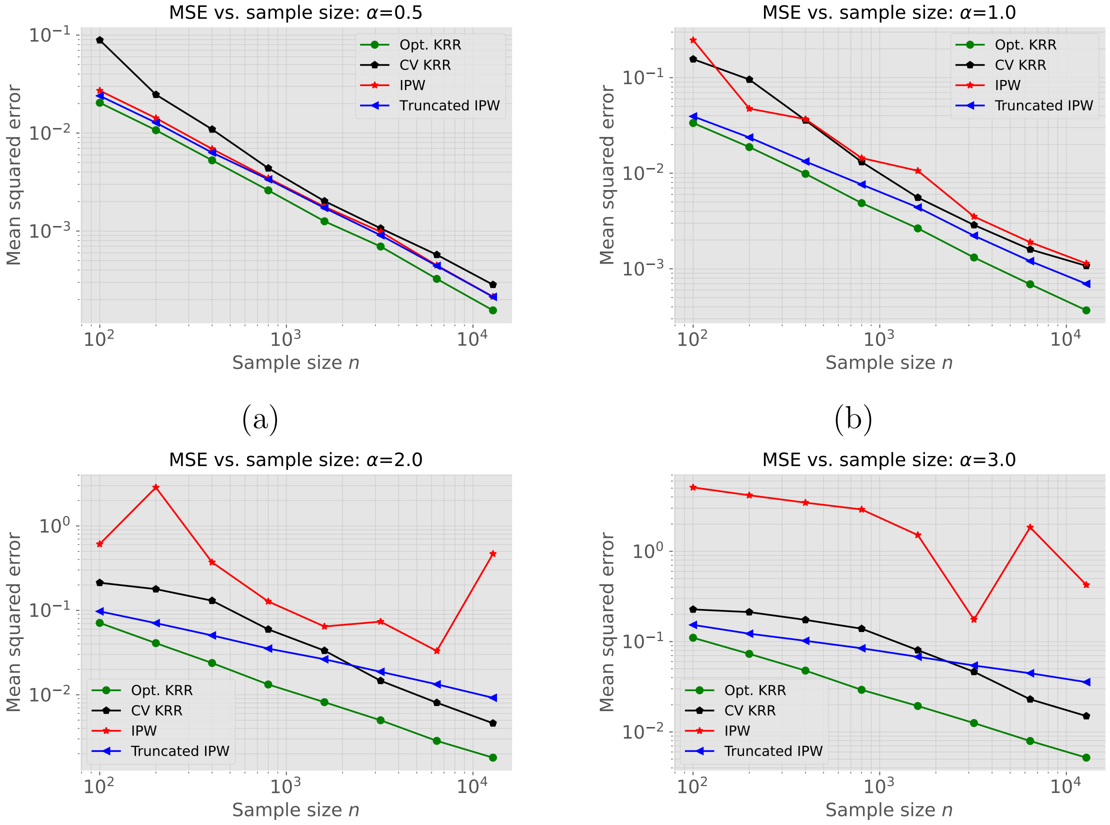
In this section, we report complete simulation results for the missing data problem, but with singular importance ratios, as previously described in Section 5.1.1—in particular, see equation (5.2). We run the four estimators discussed above, and compare their performance. The simulation setup is essentially the same as Section 5.4.1, with the only difference being that the sample size varies within the range $n \in \{100, 200, 400, 800, 1600, 3200, 6400, 12800\}$.[9] In defining the truncation-based estimator $\widehat{\tau}_{n,\mathrm{trunc}}$, we use the truncation level $\gamma_n = 1/\sqrt{n}$; this choice yields the optimal rate of convergence among truncated IPW estimators.

In Figure 5.3, we present the results of our simulations. The problem instances are generated from the family of singular models (5.2) with exponents $\alpha \in \{0.5, 1, 2, 3\}$. It can be seen that the simulation results match well with our theoretical prediction: in the classical regime with $\alpha = 0.5$, all the four estimators yield the same rate, while $\widehat{\tau}_{n,\mathrm{opt}}$ achieves slightly better instance-dependent behavior; in the critical regime $\alpha = 1$, the four estimators start to exhibit diverging behavior; in the harder regimes of $\alpha \in \{2, 3\}$, the optimal estimator $\widehat{\tau}_{n,\mathrm{opt}}$ achieves the sharpest slope, significantly outperforming the other three alternatives.

5.5 Proofs

In this section, we collect the proofs of our main results, with some auxiliary results deferred to the appendices.

[9] We made this slight modification so as to avoid the rare event that no outcome is observed.



(a)

(b)

(c)

(d)

Figure 5.3: Plots of the mean-squared error $\mathbb{E}\left[|\widehat{\tau}_{n,\diamond} - \tau^*|^2\right]$ versus sample size n for $\diamond \in \{\text{ipw}, \text{trunc}, \text{opt}, \text{cv}\}$. The simulation parameters are exactly the same as Figure 5.2, except for the underlying problem instances. As indicated by the sub-figure titles, each panel corresponds to an exponent $\alpha \in \{0.5, 1, 2, 3\}$. We have already presented part of the results (the cases of $\alpha = 0.5$ and $\alpha = 2$) in Section 5.1.

5.5.1 Proof of Theorem 5.1

Throughout this section, we adopt the shorthand $\mathscr{M}_n(\mathcal{I}^*, \mathcal{F}) \equiv \mathscr{M}_n(\mathcal{F})$, since $\mathcal{I}^*$ remains fixed throughout.

5.5.1.1 Proof of Theorem 5.1(a)

This proof exploits some techniques introduced in Chapter 4. Recalling that c is a universal constant, it suffices to prove the following two claims:

$$\mathscr{M}_n(\mathcal{I}^*, \mathcal{F}) \overset{(a)}{\geq} \frac{c}{n} V_{\xi^*}^2(\mu^*), \quad \text{and} \quad \mathscr{M}_n(\mathcal{I}^*, \mathcal{F}) \overset{(b)}{\geq} \frac{c}{n} V_{\sigma,n}^2(\xi^*, \pi, g; \mathcal{F}). \tag{5.41}$$

Beginning with the bound (5.41)(a), we first observe that the minimax risk over the class $\mathfrak{N}_n(\mu^*, \xi^*)$ is lower bounded by the risk with fixed outcome function μ^* and underlying distribution in the neighborhood of ξ^*, i.e.,

$$\mathscr{M}_n(\mathcal{F}) \geq \mathscr{M}_n(\{\mu^*\}) = \inf_{\widehat{\tau}_n} \sup_{(\xi, \mu^*) \in \mathfrak{N}_n(\mu^*, \xi^*)} \mathbb{E}\big[\left| \widehat{\tau}_n - \tau(\mu^*, \xi) \right|^2 \big].$$

But by Theorem 4.3 in Chapter 4, this minimax risk is lower bounded by $\frac{c}{n} V_{\xi^*}^2(\mu^*)$, which establishes the claim.

We now turn to proving the bound (5.41)(b), and we do so via a version of Le Cam's two point lower bound. More precisely, for a fixed underlying distribution ξ^*, we construct a pair (μ_+, μ_-) of outcome functions within the neighborhood $\big\{\mu : \|\mu - \mu^*\|_{\mathbb{L}^2(\xi^* \cdot \pi)} \leq \frac{\bar{\sigma}^2}{n}\big\} \cap \mathcal{F}$ such that if we let $\mathbb{P}_{\mu, \xi}$ be the distribution of observations under the ground truth (μ, ξ), there is

$$d_{\mathrm{TV}}\Big(\mathbb{P}_{\mu_+, \xi^*}^{\otimes n}, \mathbb{P}_{\mu_-, \xi^*}^{\otimes n} \Big) \overset{(a)}{\leq} \frac{1}{2}, \quad \text{and} \quad \mathscr{L}_\omega(\mu_+, \xi^*) - \mathscr{L}_\omega(\mu_-, \xi^*) \overset{(b)}{\geq} \frac{c}{\sqrt{n}} V_{\sigma, n}(\xi^*, \pi, g; \mathcal{F}) \tag{5.42}$$

Le Cam's two-point lemma (see e.g. [213], Chapter 15) then implies

$$\begin{aligned} \mathscr{M}_n(\mathcal{F}) &\geq \frac{1}{4} \Big\{ 1 - d_{\mathrm{TV}}\Big(\mathbb{P}_{\mu_+, \xi^*}^{\otimes n}, \mathbb{P}_{\mu_-, \xi^*}^{\otimes n} \Big) \Big\} \cdot \big\{ \mathscr{L}_\omega(\mu_+, \xi^*) - \mathscr{L}_\omega(\mu_-, \xi^*) \big\}^2 \\ &\geq \frac{c^2}{8n} V_{\sigma, n}^2(\xi^*, \pi, g; \mathcal{F}), \end{aligned}$$

completing the proof of equation (5.41)(b)

In order to prove the two bounds in line (5.42), we first need to specify the problem instances.

Construction of problem instances: We consider the noisy observation model

$$Y_i \mid X_i, A_i \sim \mathcal{N}\Big(\mu^*(X_i, A_i), \sigma^2(X_i, A_i) \Big) \qquad \text{for } i = 1, 2, \ldots, n. \tag{5.43}$$

We may assume that $V_{\sigma, n}(\xi^*, \pi, g; \mathcal{F}) > 0$ without loss of generality (otherwise the lower bound is trivial). By the defining equation (5.1) and (5.12), and the symmetry of the function class $\mathcal{F}$, there exists a function $q_0 : \mathbb{X} \times \mathbb{A} \rightarrow \mathbb{R}$ such that

$$\mathbb{E}_{\xi^*}\Big[\int_{\mathbb{A}} q_0(X, a) dg(a \mid X) \Big] \geq \tfrac{V_{\sigma, n}(\xi^*, \pi, g; \mathcal{F})}{2}, \quad \text{and}$$
$$\tfrac{q_0}{\sqrt{n}} \in \mathcal{F}, \quad \mathbb{E}_{\xi^* \cdot \pi}\Big[\tfrac{q_0^2(X, A)}{\sigma^2(X, A)} \Big] \leq \tfrac{1}{4}.$$

Using this function, we construct the outcome functions

$$\mu_+ := \mu^* + \tfrac{1}{2\sqrt{n}} q_0, \quad \text{and} \quad \mu_- := \mu^* - \tfrac{1}{2\sqrt{n}} q_0.$$

Since $\mu^* \in \frac{1}{2}\mathcal{F}$ and $\frac{1}{2\sqrt{n}} q_0 \in \frac{1}{2}\mathcal{F}$, we have $\mu_+, \mu_- \in \mathcal{F}$ by convexity and symmetry. On the other hand, we have the distance bound

$$\|\mu^* - \mu_+\|_{\mathbb{L}^2(\xi^* \cdot \pi)}^2 = \frac{1}{4n}\mathbb{E}_{\xi^* \cdot \pi}\big[q_0^2(X, A)\big] \leq \frac{\bar{\sigma}^2}{4n}\mathbb{E}_{\xi^* \cdot \pi}\Big[\frac{q_0^2(X, A)}{\sigma^2(X, A)}\Big] \leq \frac{\bar{\sigma}^2}{16n}.$$

Consequently, we have $\mu_+ \in \mathfrak{N}_n^{val}(\mu^*) \cap \mathcal{F}$. Similarly, we also have $\mu_- \in \mathfrak{N}_n^{val}(\mu^*) \cap \mathcal{F}$.

Proof of equation (5.42)**(a):** We bound the KL divergence between the product measures. Let $\mathcal{L}(Y|X, A)$ denote the conditional law of Y given the pair (X, A), we note that

$$\begin{aligned} D_{\mathrm{KL}}\left(\mathbb{P}_{\mu_+, \xi^*}^{\otimes n} \parallel \mathbb{P}_{\mu_-, \xi^*}^{\otimes n}\right) &\overset{(i)}{=} n \cdot D_{\mathrm{KL}}\left(\mathbb{P}_{\mu_+, \xi^*} \parallel \mathbb{P}_{\mu_-, \xi^*}\right) \\ &\overset{(ii)}{\leq} n \cdot \mathbb{E}\Big[D_{\mathrm{KL}}\left(\mathcal{L}(Y|X, A)\big|_{\mu_+} \parallel \mathcal{L}(Y|X, A)\big|_{\mu_-}\right)\Big] \\ &= n \cdot \frac{1}{4n} \cdot \mathbb{E}\Big[\frac{q_0^2(X, A)}{\sigma^2(X, A)}\Big] \leq \frac{1}{4}, \end{aligned}$$

where we use tensorization of KL divergence in step (i), and use convexity of KL divergence in step (ii).

Applying Pinsker's inequality yields

$$d_{\mathrm{TV}}\left(\mathbb{P}_{\mu_+, \xi^*}^{\otimes n}, \mathbb{P}_{\mu_-, \xi^*}^{\otimes n}\right) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}\left(\mathbb{P}_{\mu_+, \xi^*}^{\otimes n} \parallel \mathbb{P}_{\mu_-, \xi^*}^{\otimes n}\right)} \leq \frac{1}{2\sqrt{2}},$$

which proves equation (5.42)(a).

Proof of equation (5.42)**(b):** Straightforward calculation yields

$$\mathscr{L}_\omega(\mu_+, \xi^*) - \mathscr{L}_\omega(\mu_-, \xi^*) = \frac{1}{\sqrt{n}}\mathbb{E}_{\xi^*}\Big[\int_{\mathbb{A}} q_0(X, a) dg(a \mid X)\Big] \geq \frac{1}{2\sqrt{n}} V_{\sigma, n}(\xi^*, \pi, g; \mathcal{F}).$$

5.5.1.2 Proof of Theorem 5.1(b)

Similar to the proof of Theorem 5.1, we use Le Cam's two-point lemma. By the definition (5.12) of the variance functional $V_{\sigma, n}(\delta_{x_0}, \pi, g; \mathcal{F})$, there exists a function $q_0 : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$, such that

$$\int_{\mathbb{A}} q_0(x_0, a) dg(a \mid x_0) \geq \frac{V_{\sigma, n}(\delta_{x_0}, \pi, g; \mathcal{F})}{2}, \quad \frac{q_0}{\sqrt{n}} \in \mathcal{F}, \quad \text{and} \quad \mathbb{E}_{\xi^* \cdot \pi}\Big[\frac{q_0^2(X, A)}{\sigma^2(X, A)}\Big] \leq \frac{1}{4}.$$

Using this function, we construct the outcome functions

$$\mu_+ := \mu^* + \frac{1}{2\sqrt{n}} q_0, \quad \text{and} \quad \mu_- := \mu^* - \frac{1}{2\sqrt{n}} q_0.$$

Under the construction (5.43), following the derivation of equation (5.42)(a), we have

$$d_{\text{TV}}\Big(\mathbb{P}^{\otimes n}_{\mu_+,\xi^*}, \mathbb{P}^{\otimes n}_{\mu_-,\xi^*}\Big) \leq \frac{1}{2}.$$

On the other hand, the gap satisfies

$$\mathscr{L}_\omega(\mu_+, \delta_{x_0}) - \mathscr{L}_\omega(\mu_-, \delta_{x_0}) = \frac{1}{\sqrt{n}} \int_{\mathbb{A}} q_0(x_0, a) dg(a \mid x_0) \geq \frac{1}{2\sqrt{n}} V_{\sigma,n}(\delta_{x_0}, \pi, g; \mathcal{F}).$$

Applying Le Cam's lemma yields the claim.

5.5.2 Proof of Propositions 5.1 and 5.2

In this section, we prove our two propositions that characterize the variance functional in the case of an RKHS.

5.5.2.1 Proof of Proposition 5.1

The claim consists of two inequalities, and we split our proof accordingly.

Proof of inequality (5.18)(b): By definition, we have

$$\begin{aligned} & V_{\sigma,n}(\nu, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) \\ & = \sqrt{n} \sup_{f \in \mathscr{H}} \left\{ \left| \mathbb{E}_\nu\Big[\int_{\mathbb{A}} f(X,a) dg(a \mid X) \Big] \right| \;\mid\; \|f\|_{\mathbb{H}} \leq R \text{ and } \mathbb{E}_{\xi^* \cdot \pi}\Big[\tfrac{f^2(X,A)}{\sigma^2(X,A)}\Big] \leq \tfrac{1}{4n} \right\} \\ & \geq \sup_{q \in \mathscr{H}} \left\{ \mathbb{E}_\nu\Big[\int_{\mathbb{A}} q(X,a) dg(a \mid X) \Big] \;\mid\; \tfrac{\|q\|_{\mathbb{H}}^2}{R^2 n} + 4\mathbb{E}_{\xi^* \cdot \pi}\Big[\tfrac{q^2(X,A)}{\sigma^2(X,A)}\Big] \leq 1 \right\}, \end{aligned} \tag{5.44}$$

where we have made the change of variable $f = q/\sqrt{n}$.

Any function $q \in \mathbb{H}$ has a basis expansion of the form $q = \sum_{j=1}^\infty \theta_j \phi_j$, whence

$$\begin{aligned} \mathbb{E}_\nu\Big[\int_{\mathbb{A}} q(X,a) dg(a \mid X) \Big] &= \langle \theta, \, \bar{u} \rangle_{\ell^2}, \quad \text{where } \bar{u} \equiv \bar{u}(\nu), \text{ and} \\ \tfrac{1}{R^2 n} \|q\|_{\mathbb{H}}^2 + 4\mathbb{E}_{\xi^* \cdot \pi}\Big[\tfrac{q^2(X,A)}{\sigma^2(X,A)}\Big] &= \theta^\top \Big\{ (R^2 n)^{-1} \mathbf{\Lambda}^{-1} + 4\, \mathbf{\Gamma}_\sigma \Big\} \theta, \end{aligned}$$

where we use the eigen-value representation $\|q\|_{\mathbb{H}}^2 = \theta^\top \mathbf{\Lambda}^{-1} \theta$.

We make the choice

$$\theta = \big\{(R^2 n)^{-1} \mathbf{\Lambda}^{-1} + 4\mathbf{\Gamma}_\sigma\big\}^{-1} \bar{u} / \| \big\{(R^2 n)^{-1} \mathbf{\Lambda}^{-1} + 4\mathbf{\Gamma}_\sigma\big\}^{-1/2} \bar{u} \|_{\ell^2}.$$

Substituting this choice into equation (5.44) yields

$$V^2_{\sigma,n}(\nu, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) \geq \bar{u}^\top \big\{(R^2 n)^{-1} \mathbf{\Lambda}^{-1} + 4\mathbf{\Gamma}_\sigma\big\}^{-1} \bar{u} \geq \frac{1}{4} \bar{u}^\top \big\{(R^2 n)^{-1} \mathbf{\Lambda}^{-1} + \mathbf{\Gamma}_\sigma\big\}^{-1} \bar{u},$$

which establishes inequality (b).

Proof of inequality (5.18)**(a):** Turning to the other inequality in the claim, the same change of variable and followed by basis expansion yields

$$
\begin{aligned}
& V_{\sigma,n}(\nu, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) \\
&= \sup_{q \in \mathscr{H}} \Big\{ \Big| \mathbb{E}_\nu \Big[\int_{\mathbb{A}} q(X,a) dg(a \mid X) \Big] \Big| \;\mid\; \tfrac{1}{R\sqrt{n}} \|q\|_{\mathbb{H}} \leq 1,\ \mathbb{E}_{\xi^* \cdot \pi} \Big[\tfrac{q^2(X,A)}{\sigma^2(X,A)} \Big] \leq \tfrac{1}{4} \Big\} \\
&\leq \sup_{q \in \mathscr{H}} \Big\{ \mathbb{E}_\nu \Big[\int_{\mathbb{A}} q(X,a) dg(a \mid X) \Big] \;\mid\; \tfrac{1}{R^2 n} \|q\|_{\mathbb{H}}^2 + \mathbb{E}_{\xi^* \cdot \pi} \Big[\tfrac{q^2(X,A)}{\sigma^2(X,A)} \Big] \leq \tfrac{5}{4} \Big\} \\
&= \sup_{\theta \in \ell^2} \Big\{ \langle \theta,\, \bar{u} \rangle_{\ell^2} \;\mid\; \theta^\top \Big((R^2 n)^{-1} \mathbf{\Lambda}^{-1} + \mathbf{\Gamma}_\sigma \Big) \theta \leq \tfrac{5}{4} \Big\} \\
&\leq \tfrac{\sqrt{5}}{2} \cdot \sqrt{ \bar{u}^\top \big\{ (R^2 n)^{-1} \mathbf{\Lambda}^{-1} + \mathbf{\Gamma}_\sigma \big\}^{-1} \bar{u} },
\end{aligned}
$$

which completes the proof of inequality (a).

5.5.2.2 Proof of Proposition 5.2

Throughout this proof, we use $\| \cdot \|_2$ as a shorthand for the $\mathbb{L}^2(\xi^* \cdot \pi)$-norm. The variational formulation (5.12) can be re-written as

$$
V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) = \sup \Big\{ \langle f,\, \tfrac{dg}{d\pi} \rangle \;\mid\; \mathbb{E}_{\xi^* \cdot \pi} \big[\tfrac{f^2(X,A)}{\sigma^2(X,A)} \big] \leq \tfrac{1}{4}, \|f\|_{\mathbb{H}} \leq R\sqrt{n} \Big\}. \tag{5.45}
$$

Clearly, the function $n \mapsto V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R))$ is non-decreasing, and since $v_{\mathrm{semi}}(\mu^*, \frac{dg}{d\pi}) < +\infty$, it is uniformly bounded from above. Therefore, by taking $n \to +\infty$, the limit exists. Moreover, we have

$$
\lim_{n \to \infty} V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) \leq \tfrac{1}{2} \sqrt{ \mathbb{E}_{\xi^* \cdot \pi} \Big[\big(\tfrac{dg}{d\pi}(A \mid X) \big)^2 \cdot \sigma^2(X,A) \Big] }.
$$

Now the function $(x,a) \mapsto \sigma(x,a)\ \frac{dg}{d\pi}(a \mid x)$ belongs to $\mathbb{L}^2(\xi^* \cdot \pi)$. Combined with the uniform upper bound $\sup_{x,a} \sigma^2(x,a) \leq \bar{\sigma}^2$, it follows that the function $\sigma^2\ \frac{dg}{d\pi}$ also belongs to $\mathbb{L}^2(\xi^* \cdot \pi)$. Since the Hilbert space $\mathbb{H}$ is universal (and hence dense in $\mathbb{L}^2(\xi^* \cdot \pi)$), it follows that for any $\varepsilon > 0$, we can find a function $h_\varepsilon \in \mathbb{H}$ such that $\|h_\varepsilon - \sigma^2 \frac{dg}{d\pi}\|_2 \leq \varepsilon$.

Now define the rescaled function $q_\varepsilon := h_\varepsilon / (2\|\sigma \frac{dg}{d\pi}\|_2)$. With this definition, we have

$$
\begin{aligned}
\mathbb{E}\Big[\tfrac{q_\varepsilon^2(X,A)}{\sigma^2(X,A)} \Big] &\leq \tfrac{1}{4} \|\sigma \tfrac{dg}{d\pi}\|_2^{-2} \mathbb{E}\Big[\Big\{ \big| \sigma(X,A) \cdot \tfrac{dg}{d\pi}(A \mid X) \big| + \frac{\big| h_\varepsilon - \sigma^2 \frac{dg}{d\pi} \big|}{\sigma}(X,A) \Big\}^2 \Big] \\
&\leq \|\sigma \tfrac{dg}{d\pi}\|_2^{-2} \Big\{ \tfrac{1+\varepsilon}{4} \mathbb{E}\big[\sigma^2(X,A) \cdot \big(\tfrac{dg}{d\pi}(A \mid X) \big)^2 \big] + \tfrac{1}{\varepsilon} \mathbb{E}\big[\big(\tfrac{h_\varepsilon - \sigma^2 \frac{dg}{d\pi}}{\sigma} \big)^2 (X,A) \big] \Big\} \\
&\leq \frac{1+\varepsilon}{4} + \frac{\varepsilon}{\underline{\sigma}^2},
\end{aligned}
$$

along with the bound $\|q_\varepsilon\|_{\mathbb{H}} \leq \|\sigma \frac{dg}{d\pi}\|_2^{-1}\ \|h_\varepsilon\|_{\mathbb{H}} < \infty$.

We now define the rescaled function $q := \frac{1}{1+\varepsilon+\frac{4\varepsilon}{\underline{\sigma}^2}} q_\varepsilon$. Given a sample size lower bounded as $n \geq \|h_\varepsilon\|_{\mathbb{H}}^2 / (R^2 \|\sigma \frac{dg}{d\pi}\|_{\mathbb{L}^2(\xi^* \cdot \pi)}^2)$, the above inequalities imply that the rescaled function q satisfies the constraints in the optimization problem (5.45). Substituting this choice into the objective function, we find that

$$\begin{aligned}\mathbb{E}_{\xi^*}\Big[\int_{\mathbb{A}} q(X,a) dg(a \mid X)\Big] &\geq \frac{1-\varepsilon-4\varepsilon/\underline{\sigma}^2}{2\|\sigma \frac{dg}{d\pi}\|_2} \mathbb{E}_{\xi^*}\Big[\int_{\mathbb{A}} h_\varepsilon(X,a) dg(a \mid X)\Big] \\ &\geq \frac{1-\varepsilon-4\varepsilon/\underline{\sigma}^2}{2\|\sigma \frac{dg}{d\pi}\|_2} \Big\{ \|\tfrac{dg}{d\pi}\sigma\|_2^2 - \mathbb{E}\Big[\big(h_\varepsilon - \sigma^2 \tfrac{dg}{d\pi}\big) \cdot \tfrac{dg}{d\pi}(A \mid X)\Big]\Big\}.\end{aligned}$$

The Cauchy–Schwarz inequality implies that

$$\mathbb{E}\Big[\big(h_\varepsilon - \sigma^2 \tfrac{dg}{d\pi}\big) \cdot \frac{dg}{d\pi}(A \mid X)\Big] \leq \|h_\varepsilon - \sigma^2 \frac{dg}{d\pi}\|_2 \cdot \frac{1}{\underline{\sigma}} \, \|\sigma \tfrac{dg}{d\pi}\|_2,$$

where we have used the fact that $\sigma(x,a) \geq \underline{\sigma}$ for all pairs (x,a).

Combining the two bounds together and taking the limit, we conclude that

$$\begin{aligned}\lim_{n \to \infty} V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) &\geq \mathbb{E}_{\xi^*}\Big[\int_{\mathbb{A}} q(X,a) dg(a|X)\Big] \\ &\geq \tfrac{1-\varepsilon-\varepsilon/\underline{\sigma}^2}{2} \, \|\sigma \tfrac{dg}{d\pi}\|_2 - \tfrac{1-\varepsilon-\varepsilon/\underline{\sigma}^2}{2\underline{\sigma}} \varepsilon.\end{aligned}$$

Since the choice of ε is arbitrary, this concludes the proof of this proposition.

5.5.3 Proof of Theorem 5.2 and variants

Let us first introduce some notation used in the proof. Our proof involves the diagonal operator $\mathbf{\Lambda}^{-1} := \mathrm{diag}\big(\{\lambda_k^{-1}\}_{k=1}^\infty\big)$, and the weighted ℓ^2-norms

$$\|z\|_\lambda^2 := \sum_{j=1}^\infty \lambda_j z_j^2, \quad \text{and} \quad \|z\|_{\lambda^{-1}}^2 := \sum_{j=1}^\infty \lambda_j^{-1} z_j^2. \tag{5.46a}$$

5.5.3.1 Set-up for auxiliary results

We define the empirical feature vector $\widehat{u}_n := \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{A}} \phi(X_i^{(\mathrm{II})}, a) dg(a \mid X_i^{(\mathrm{II})})$. By the kernel boundedness assumption Kbou(κ), we have $\|\bar{u}\|_\lambda < \infty$ and $\|\widehat{u}_n\|_\lambda < \infty$ almost surely. We also define a linear operator Ψ from the Hilbert space $\mathbb{H}$ to the sequence space ℓ^2 with components $[\Psi(f)]_j = \langle f, \, \phi_j \rangle_{\mathbb{L}^2(\xi^* \cdot \pi)}$. Since μ^* and $\widehat{\mu}_n$ belong to the Hilbert space $\mathbb{H}$, it is meaningful to define

$$\beta_* := \Psi(\mu^*), \quad \text{and} \quad \widehat{\beta}_n := \Psi(\widehat{\mu}_n).$$

Note that for any function $f \in \mathbb{H}$, we have

$$\|\Psi f\|_{\lambda^{-1}}^2 = \sum_{j=1}^{+\infty} \lambda_j^{-1} \langle f,\, \phi_j \rangle^2_{\mathbb{L}^2(\xi^* \cdot \pi)} = \sum_{j=1}^{\infty} \langle f,\, \sqrt{\lambda_j} \phi_j \rangle_{\mathbb{H}}^2 = \|f\|_{\mathbb{H}}^2 < \infty.$$

Consequently, the inner products $\langle u,\, \beta \rangle \leq \|u\|_\lambda \cdot \|\beta\|_{\lambda^{-1}}$ are well-defined for $u \in \{\bar{u}, \widehat{u}_n\}$ and $\beta \in \{\beta_*, \widehat{\beta}_n\}$. Fubini's theorem guarantees that target functional τ^* and the estimator $\widehat{\tau}_n$ can be written as

$$\tau^* = \sum_{k=0}^{\infty} \langle \mu^*,\, \phi_k \rangle_{\mathbb{L}^2(\xi^* \cdot \pi)} \cdot \mathbb{E}_{\xi^*} \left[\int_{\mathbb{A}} \phi_k(X, a) dg(a \mid X) \right] = \langle \bar{u},\, \beta_* \rangle, \quad \text{and}$$

$$\widehat{\tau}_n = n^{-1} \sum_{k=0}^{\infty} \langle \widehat{\mu}_n^{(1)},\, \phi_k \rangle_{\mathbb{L}^2(\xi^* \cdot \pi)} \sum_{i=1}^{n} \int_{\mathbb{A}} \phi_k(X_i^{(\mathrm{II})}, a) dg(a \mid X_i^{(\mathrm{II})}) \;=\; \langle \widehat{u}_n,\, \widehat{\beta}_n \rangle.$$

We therefore have the following error decomposition:

$$\widehat{\tau}_n - \tau^* = \langle \widehat{u}_n - \bar{u},\, \beta_* \rangle + \langle \bar{u},\, \widehat{\beta}_n - \beta_* \rangle + \langle \widehat{u}_n - \bar{u},\, \widehat{\beta}_n - \beta_* \rangle. \tag{5.47}$$

The rest of this section is devoted to bounds on each terms appearing in this decomposition. In particular, we require two auxiliary results. Recall the shorthand notation $V_{\xi^*}^2(f) := \mathrm{var}_{X \sim \xi^*} \big(\int_{\mathbb{A}} f(X, a) d\omega(a \mid X) \big)$.

Lemma 5.1. *Under Assumptions (Kbou(κ)) and (subG(σ)), for any function $f \in \mathbb{H}$, we have*

$$|\langle \widehat{u}_n - \bar{u},\, \Psi f \rangle| \leq 2 V_{\xi^*}(f) \sqrt{\frac{\log(1/\delta)}{n}} + 6\kappa \|f\|_{\mathbb{H}} \frac{\log(1/\delta)}{n}. \tag{5.48a}$$

with probability at least $1 - \delta$. Furthermore, given a sample size $n \geq \log(1/\delta)$ and a scalar $\rho > 0$, we have

$$\|(I + \rho \mathbf{\Lambda}^{-1})^{-1/2} \big(\widehat{u}_n - \bar{u} \big)\|_{\ell^2} \leq \sqrt{\frac{D(\rho)}{n} \log(1/\delta)} \tag{5.48b}$$

with probability $1 - \delta$.

See Section 5.5.3.3 for the proof.

Lemma 5.2. *Suppose that the kernel bound (Kbou(κ)) and tail condition (subG(σ)) are in force. Then for any infinite-dimensional vector z and scalar $\delta \in (0, 1)$, with the regularization parameter $\rho_n = \frac{\bar{\sigma}^2}{R^2 n}$, and under the sample-size condition (5.21a), we have*

$$\Big| \langle z,\, \widehat{\beta}_n - \beta_* \rangle \Big| \leq c \| \big(I + \rho_n \mathbf{\Lambda}^{-1} \big)^{-1/2} z \|_{\ell^2} \cdot \Big\{ \bar{\sigma} \sqrt{\frac{\log(1/\delta)}{n}} + \sigma \sqrt{D(\rho_n)} \cdot \frac{\log(1/\delta) \log n}{n} \Big\}, \tag{5.49}$$

with probability at least $1 - \delta$.

See Section 5.5.3.4 for the proof.

5.5.3.2 Main argument

Taking these two lemmas as given, we now prove Theorem 5.2 by bounding each term in the decomposition result (5.47). First, recalling that $\beta_* = \Psi\mu^*$, we can apply the bound (5.48a) to find that

$$|\langle \widehat{u}_n - \bar{u},\, \beta_* \rangle| \leq 2V_{\xi^*}(\mu^*)\sqrt{\frac{\log(1/\delta)}{n}} + 6\kappa R \frac{\log(1/\delta)}{n} \tag{5.50}$$

with probability at least $1 - \delta/3$. Second, by the boundedness of basis functions, we have $\bar{u} \in \ell^\infty$, so that Lemma 5.2 can be applied to obtain

$$\begin{aligned} \left|\langle \bar{u},\, \widehat{\beta}_n - \beta_* \rangle\right| &\leq c\|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2}\bar{u}\|_{\ell^2} \cdot \Big\{\bar{\sigma}\sqrt{\frac{\log(1/\delta)}{n}} + \sigma\sqrt{D(\rho_n)} \cdot \frac{\log(1/\delta)\log n}{n}\Big\} \\ &\leq 2c\|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2}\bar{u}\|_{\ell^2}\bar{\sigma}\sqrt{\frac{\log(1/\delta)}{n}}, \end{aligned} \tag{5.51}$$

with probability $1-\delta/3$, where the last step follows from the sample-size condition (5.21a), as the condition ensures that $\sigma\sqrt{D(\rho_n)} \cdot \frac{\log(1/\delta)\log n}{n} \leq \bar{\sigma}\sqrt{\frac{\log(1/\delta)}{n}}$.

Next, we apply equation (5.48b) in combination with Lemma 5.2, and obtain the following inequality with probability $1 - \delta/3$

$$\begin{aligned} &\left|\langle \widehat{u}_n - \bar{u},\, \widehat{\beta}_n - \beta_* \rangle\right| \\ &\leq c\|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2}\big(\widehat{u}_n - \bar{u}\big)\|_{\ell^2} \cdot \Big\{\bar{\sigma}\sqrt{\frac{\log(1/\delta)}{n}} + \sigma\sqrt{D(\rho_n)} \cdot \frac{\log(1/\delta)\log n}{n}\Big\} \\ &\leq c\sqrt{\frac{D(\rho_n)}{n}\log(1/\delta)} \cdot \Big\{\bar{\sigma}\sqrt{\frac{\log(1/\delta)}{n}} + \sigma\sqrt{D(\rho_n)} \cdot \frac{\log(1/\delta)\log n}{n}\Big\} \\ &\overset{(i)}{\leq} c\sigma\sqrt{D(\rho_n)}\frac{\log(1/\delta)\log n}{n}\Big\{1 + \sqrt{\frac{D(\rho_n)\log(1/\delta)}{n}}\Big\} \\ &\overset{(ii)}{\leq} 2c\sigma\sqrt{D(\rho_n)}\frac{\log(1/\delta)\log n}{n}. \end{aligned} \tag{5.52}$$

where step (i) follows from the relation $\bar{\sigma} \leq \sigma$, whereas the final step follows from the sample size condition (5.21a), as it ensures $\frac{D(\rho_n)\log(1/\delta)}{n} \leq 1$.

Combining the inequalities (5.50), (5.51), and (5.52) completes the proof of Theorem 5.2.

5.5.3.3 Proof of Lemma 5.1

We simplify notation by omitting the superscript $^{(\mathrm{II})}$, and using (X_i, A_i, Y_i) to denote the data.

Proof of the directional bound (5.48a) : By definition, we have

$$\begin{aligned}\langle \widehat{u}_n, \, \Psi f \rangle &= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{A}} \langle \phi(X_i, a), \, \Psi f \rangle d\omega(a \mid X_i) \\ &= \frac{1}{n}\sum_{i=1}^n Z_i \qquad \text{where } Z_i := \textstyle\int_{\mathbb{A}} f(X_i, a) d\omega(a \mid X_i).\end{aligned}$$

Similarly, the population-level vector $\bar{u}$ satisfies $\langle \bar{u}, \, \Psi f \rangle = \mathbb{E}_{\xi^*}[Z]$, so that our problem amounts to bounding the fluctuations of the sample average $\frac{1}{n}\sum_{i=1}^n Z_i$ around its mean. Our approach is via Bernstein's inequality, and applying it requires control on both the variance and absolute value of Z_i. By inspection, we have $\mathrm{var}(Z_i) = V_{\xi^*}^2(f) = \mathrm{var}\big(\int_{\mathbb{A}} f(X, a) d\omega(a \mid X)\big)$ and moreover, we claim that

$$|Z_i| \leq \kappa \cdot \|f\|_{\mathbb{H}}. \tag{5.53}$$

With these two bounds in hand, invoking Bernstein's inequality (see e.g. [136]) yields

$$\mathbb{P}\left(|\langle \widehat{u}_n - \bar{u}, \, \Psi f \rangle| \geq t\right) \leq 2 \exp\left(\frac{-nt^2}{2V_{\xi^*}^2(f) + 3\kappa \|f\|_{\mathbb{H}} t}\right),$$

setting the right hand side as δ and solving for t, we complete the proof of the directional bound (5.48a).

It remains to prove the claim (5.53). Using our boundedness condition (5.6) on g, we have

$$\begin{aligned}|Z_i| = \left| \int_{\mathbb{A}} f(X_i, a) d\omega(a \mid X_i) \right| &\leq \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} |f(x, a)| \\ &\overset{(i)}{\leq} \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \sum_{k \geq 1} \lambda_k \, |\phi_k(x, a)| \cdot |\langle f, \, \phi_k \rangle_{\mathbb{H}}| \\ &\overset{(ii)}{\leq} \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \Big(\sum_{k \geq 1} \lambda_k \phi_k^2(x, a) \Big)^{1/2} \cdot \Big(\sum_{k \geq 1} \lambda_k \langle f, \, \phi_k \rangle_{\mathbb{H}}^2 \Big)^{1/2}\end{aligned}$$

where step (i) follows by expanding f into a basis representation $f = \sum_{k=1}^{+\infty} \lambda_k \langle f, \, \phi_k \rangle_{\mathbb{H}} \phi_k$; and step (ii) follows from the Cauchy–Schwarz inequality. Now by Mercer's theorem, we have the relation

$$\big(\sum_{k \geq 1} \lambda_k \langle f, \, \phi_k \rangle_{\mathbb{H}}^2 \big)^{1/2} = \|f\|_{\mathbb{H}},$$

whereas the boundedness condition (Kbou(κ)) implies that $\sum_{k \geq 1} \lambda_k \phi_k^2(x, a) \leq \mathcal{K}\big((x, a), (x, a)\big) \leq \kappa^2$, for any $(x, a) \in \mathbb{X} \times \mathbb{A}$. Putting together the pieces yields the claimed bound (5.53).

Proof of the preconditioned bound (5.48b)**:** Defining $U_i := \int_{\mathbb{A}} \phi(X_i, a) d\omega(a \mid X_i)$ so that $\widehat{u}_n = \frac{1}{n}\sum_{i=1}^n U_i$, the norm on the left-hand-side of Eq (5.48b) can be equivalently written as an empirical process supremum

$$\|(I + \rho \mathbf{\Lambda}^{-1})^{-1/2}\big(\widehat{u}_n - \bar{u}\big)\|_{\ell^2} = \sup_{z^\top (I+\rho\mathbf{\Lambda}^{-1}) z \leq 1} \frac{1}{n}\sum_{i=1}^n z^\top (U_i - \bar{u}) =: H_n$$

In order to bound the expected supremum, we simply use the Cauchy–Schwarz inequality to arrive at the bound

$$\mathbb{E}\big[H_n\big] \leq \Big\{\mathbb{E}\Big[\|(I + \rho\mathbf{\Lambda}^{-1})^{-1/2} \cdot \frac{1}{n}\sum_{i=1}^n (U_i - \bar{u})\|_{\ell^2}^2\Big]\Big\}^{1/2} \leq \sqrt{n^{-1}\mathbb{E}\big[\|(I+\rho\mathbf{\Lambda}^{-1})^{-1/2} U_i\|_{\ell^2}^2\big]},$$

where the second inequality comes from the fact that U_i's are i.i.d.

In order to bound this quantity, we use Talagrand's concentration inequality (c.f. [213], Theorem 3.8 and remarks). With probability $1 - \delta$, we have

$$\begin{aligned} H_n \leq 2\mathbb{E}\big[H_n\big] + c\Big(\sup_{z^\top (I+\rho\mathbf{\Lambda}^{-1}) z \leq 1} \mathbb{E}[(z^\top (U_i - \bar{u}))^2] \tfrac{\log(1/\delta)}{n}\Big)^{1/2} \\ + c \sup_{(x,a)\in \mathbb{X}\times\mathbb{A}} \|(I + \rho\mathbf{\Lambda}^{-1})^{-1/2}\phi(x, a)\|_{\ell^2} \cdot \tfrac{\log(1/\delta)}{n}. \end{aligned} \tag{5.54}$$

Since the Radon measure g satisfies the bound (5.6), the summand U_i satisfies the almost sure upper bound

$$\|(I + \rho\mathbf{\Lambda}^{-1})^{-1/2} U_i\|_{\ell^2} \leq \sup_{(x,a)\in\mathcal{S}\times\mathbb{A}} \|(I + \rho\mathbf{\Lambda}^{-1})^{-1/2}\phi(x, a)\|_{\ell^2} = \sqrt{D(\rho)}$$

The bound (5.54) then becomes

$$H_n \leq c\sqrt{D(\rho)} \cdot \Big\{\sqrt{\frac{\log(1/\delta)}{n}} + \frac{\log(1/\delta)}{n}\Big\},$$

which completes the proof of equation (5.48b).

5.5.3.4 Proof of Lemma 5.2

For notational simplicity, we omit the supscript $^{(\mathrm{I})}$ in (X_i, A_i, Y_i). Using the basis expansion $\mu = \sum_{k\geq 0} \beta(k)\phi_k$, we have the equivalence

$$\begin{aligned} \widehat{\beta}_n &= \Psi \cdot \arg\min_{\mu\in\mathbb{H}} \Big\{\frac{1}{n}\sum_{i=1}^n \big(Y_i - \mu(X_i, A_i)\big)^2 + \rho_n \|\mu\|_{\mathbb{H}}^2\Big\} \\ &= \arg\min_{\beta\in\ell^2(\mathbb{N})} \Big\{\frac{1}{n}\sum_{i=1}^n \big(Y_i - \langle \beta,\, \phi(X_i, A_i)\rangle\big)^2 + \rho_n \|\beta\|_{\lambda^{-1}}^2\Big\}. \end{aligned}$$

Define the noise variable $\varepsilon_i := Y_i - \mu^*(X_i, A_i)$ along with the empirical covariance operator $\widehat{\mathbf{\Gamma}}_n := \frac{1}{n}\sum_{i=1}^n \phi(X_i, A_i)\phi(X_i, A_i)^\top$. Using this notation, we can write

$$\widehat{\beta}_n - \beta_* = \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} \cdot \frac{1}{n}\sum_{i=1}^n \Big\{\varepsilon_i \phi(X_i, A_i) - \rho_n \mathbf{\Lambda}^{-1}\beta_*\Big\}.$$

Our approach to controlling the projection of this quantity in any fixed direction z consists of two steps:

- First, conditionally on the state-action pairs $(X_i, A_i)_{i=1}^n$, we exhibit a high-probability upper bound on the error $z^\top\big(\widehat{\beta}_n - \beta_*\big)$ with respect to the randomness in the outcomes Y_i. The bound depends on the behavior of the empirical covariance operator $\widehat{\mathbf{\Gamma}}_n$ of feature vectors; see Lemma 5.3 for details.
- Second, we relate the empirical covariance operator $\widehat{\mathbf{\Gamma}}_n$ with its population analogue (which is the identity operator I, since $(\phi_j)_{j=1}^\infty$ forms an orthonormal basis). The form of infinite-dimensional concentration results is exactly the form required in the first step. See Lemma 5.4 for details.

Let us give precise statements of the two auxiliary results needed in the proof:

Lemma 5.3. *Conditionally on the state-action sequence $(X_i, A_i)_{i=1}^n$, for any $z \in \ell^\infty(\mathbb{N})$, we have*

$$\begin{aligned}&\Big|z^\top(\widehat{\beta}_n - \beta_*)\Big| \leq c\big\|\big(\widehat{\mathbf{\Gamma}}_n + \rho_n\mathbf{\Lambda}^{-1}\big)^{-1/2} z\big\|_{\ell^2} \\ &\times \Big\{\sqrt{\rho_n}\|\mu^*\|_{\mathbb{H}} + \bar{\sigma}\sqrt{\tfrac{\log(1/\delta)}{n}} + \sigma \sup_{(x,a)} \big\|\big(\widehat{\mathbf{\Gamma}}_n + \rho_n\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(x,a)\big\|_{\ell^2} \cdot \tfrac{\log(1/\delta)\log n}{n}\Big\}, \qquad (5.55)\end{aligned}$$

with probability at least $1 - \delta$.

See Appendix D.3.1 for the proof.

Our next auxiliary result relates the sample covariance operator $\widehat{\mathbf{\Gamma}}_n$ with the population one. Here we state a somewhat general result, since we use it both here and in our later proof of Theorem 5.3.

Consider a weight function $(x, a) \mapsto q(x, a) \in [\underline{q}, \overline{q}]$, where $(\underline{q}, \overline{q})$ are a pair of positive scalars. Define the empirical operator

$$\widehat{\mathbf{\Gamma}}_{n,q} := n^{-1}\sum_{i=1}^n q(X_i, A_i)\phi(X_i, A_i)\phi(X_i, A_i)^\top,$$

along with its its population version $\mathbf{\Gamma}_{*,q} := \mathbb{E}\big[\widehat{\mathbf{\Gamma}}_{n,q}\big]$. For the current proof, it suffices to take $q(x, a) = 1$.

Lemma 5.4. *For scalars $\delta, \omega \in (0, 1)$, consider a regularization parameter ρ_n satisfying the relation*

$$(\overline{q}/\underline{q}) \log\big(\tfrac{\kappa^2}{\rho_n \delta}\big) \cdot \tfrac{D\big(\rho_n/\underline{q}\big)}{n} \leq \tfrac{\omega}{16}. \tag{5.56}$$

Then we have

$$(1-\omega)\big(\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1}\big) \preceq \widehat{\mathbf{\Gamma}}_{n,q} + \rho_n \mathbf{\Lambda}^{-1} \preceq (1+\omega)\big(\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1}\big) \tag{5.57}$$

with probability at least $1-\delta$.

See Appendix D.3.2 for the proof.

Taking these two lemmas as given, we now proceed with the proof of Lemma 5.2. We define the event

$$\mathscr{E} := \Big\{ \widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1} \succeq \tfrac{1}{2}(I + \rho_n \mathbf{\Lambda}^{-1}) \Big\}.$$

With the given choice $\rho_n = \frac{\bar{\sigma}^2}{R^2 n}$, for a sample size n satisfying the requirement (5.21a), we have

$$\log\big(\frac{\kappa^2}{\rho_n \delta}\big) \frac{D(\rho_n)}{n} \leq \frac{1}{32}.$$

By applying Lemma 5.4 with $q(x,a) \equiv 1$, we are guaranteed that $\mathbb{P}(\mathscr{E}) \geq 1-\delta$.

Conditioned on the event $\mathscr{E}$, the definition of effective dimension guarantees that

$$\sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \|\big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x,a)\|_{\ell^2} \\ \leq \sqrt{2} \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x,a)\|_{\ell^2} \leq \sqrt{2D(\rho_n)}.$$

Consequently, conditioned on the event $\mathscr{E}$, Lemma 5.3 guarantees that

$$\begin{aligned} &\Big| z^\top (\widehat{\beta}_n - \beta_*) \Big| \\ &\leq c \|\big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} z\|_{\ell^2} \cdot \Big\{ \sqrt{\rho_n} \|\mu^*\|_{\mathbb{H}} + \bar{\sigma} \sqrt{\tfrac{\log(1/\delta)}{n}} + \sigma \sqrt{2D(\rho_n)} \cdot \tfrac{\log(1/\delta) \log n}{n} \Big\} \\ &\leq 2c \|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} z\|_{\ell^2} \cdot \Big\{ \sqrt{\rho_n} \|\mu^*\|_{\mathbb{H}} + \bar{\sigma} \sqrt{\tfrac{\log(1/\delta)}{n}} + \sigma \sqrt{D(\rho_n)} \cdot \tfrac{\log(1/\delta) \log n}{n} \Big\}, \end{aligned}$$

with probability at least $1-\delta$.

Substituting the choice $\rho_n = \frac{\bar{\sigma}^2}{R^2 n}$, we note that $\sqrt{\rho_n} \|\mu^*\|_{\mathbb{H}} \leq \sqrt{\rho_n} R \leq \frac{\bar{\sigma}}{\sqrt{n}}$, leading to the bound with probability $1-\delta$

$$\Big| z^\top (\widehat{\beta}_n - \beta_*) \Big| \leq 2c \|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} z\|_{\ell^2} \cdot \Big\{ \bar{\sigma} \sqrt{\tfrac{\log(1/\delta)}{n}} + \sigma \sqrt{D(\rho_n)} \cdot \tfrac{\log(1/\delta) \log n}{n} \Big\},$$

which completes the proof of Lemma 5.2.

5.5.4 Proof of Theorem 5.3 and corollaries

The proof consists of three parts: we first establish guarantees on the auxiliary estimators $\widetilde{\mu}_n$ and $\widehat{\sigma}_n^2$, and then use these guarantees to bound the error of the two-stage estimator $\widehat{\tau}_n$. Concretely, we prove the following claims in turn.

- For any fixed state-action pair (s_0, a_0) and any $\delta \in (0,1)$, the first-stage estimator $\widetilde{\mu}_n$ satisfies the bound
$$|\widetilde{\mu}_n(s_0, a_0) - \mu^*(s_0, a_0)| \leq c\sigma \sqrt{\frac{D(\rho_n^{(I)})}{n} \log(1/\delta)}, \tag{5.58a}$$
with probability $1 - \delta$. See Section 5.5.4.1 for the proof.
- For any fixed state-action pair (s_0, a_0), the second-stage estimator $\widehat{\sigma}_n$ satisfies the bound
$$\frac{1}{2}\sigma^2(s_0, a_0) \leq \widehat{\sigma}_n^2(s_0, a_0) \leq 2\sigma^2(s_0, a_0), \tag{5.58b}$$
with probability $1 - \delta/n$. See Section 5.5.4.2 for the proof.
- Using an approach analogous to that in the proof of Theorem 5.2, we represent the target functionals using basis functions, and recall the error decomposition
$$\widehat{\tau}_n - \tau^* = \langle \widehat{u}_n - \bar{u},\, \beta_* \rangle + \langle \bar{u},\, \widehat{\beta}_n - \beta_* \rangle + \langle \widehat{u}_n - \bar{u},\, \widehat{\beta}_n - \beta_* \rangle, \tag{5.58c}$$
where we denote $\beta_* := \Psi \mu^*$ and $\widehat{\beta}_n := \Psi \widehat{\mu}_n$. The errors in the sample average feature vector $\widehat{u}_n$ can be controlled using Lemma 5.1 just as in the proof of Theorem 5.2, while bounding the error for the weighted least-square estimator $\widehat{\beta}_n$ requires new ingredients; see Lemma 5.5 to follow.

Lemma 5.5. *Under the conditions of Theorem 5.3, with probability* $1 - \delta$, *for any infinite-dimensional vector* z, *we have*
$$\left| z^\top (\widehat{\beta}_n - \beta_*) \right| \leq c \| \left(\mathbf{\Gamma}_\sigma + \rho_n^{(III)} \mathbf{\Lambda}^{-1} \right)^{-1/2} z \|_{\ell^2} \sqrt{\frac{\log(1/\delta)}{n}}, \tag{5.59}$$
where $c > 0$ *is a universal constant.*

See Section 5.5.4.3 for the proof.

Having set up the basic ingredients, we are now ready to prove the main claims of Theorem 5.3. We bound each terms in the decomposition result (5.58c) as follows.

Applying the bound (5.48a) from Lemma 5.1 with $f = \mu^*$, with probability $1 - \delta$, we have
$$|\langle \widehat{u}_n - \bar{u},\, \beta_* \rangle| \leq 2V_{\xi^*}(\mu^*) \sqrt{\frac{\log(1/\delta)}{n}} + 6\kappa R \frac{\log(1/\delta)}{n}. \tag{5.60}$$

Applying Lemma 5.5 with $z = u$ yields the bound

$$|\langle \bar{u},\, \widehat{\beta}_n - \beta_* \rangle| \leq c \| \big(\mathbf{\Gamma}_\sigma + \rho_n^{(\mathrm{III})} \mathbf{\Lambda}^{-1}\big)^{-1/2} \bar{u} \|_{\ell^2} \sqrt{\tfrac{\log(1/\delta)}{n}} \leq 2c V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) \sqrt{\tfrac{\log(1/\delta)}{n}}. \tag{5.61}$$

valid with probability $1-\delta$. Here the second step follows from Proposition 5.1.

Finally, applying Lemma 5.5 with $z = \widehat{u}_n - \bar{u}$,[10] as well as equation (5.48b) in Lemma 5.1, with probability $1-\delta$, we have the upper bound

$$\begin{aligned}
\Big| \langle \widehat{u}_n - \bar{u},\, \widehat{\beta}_n - \beta_* \rangle \Big| &\leq c \| \big(\mathbf{\Gamma}_\sigma + \rho_n^{(\mathrm{III})} \mathbf{\Lambda}^{-1}\big)^{-1/2} (\widehat{u}_n - \bar{u}) \|_{\ell^2} \sqrt{\frac{\log(1/\delta)}{n}} \\
&\leq c\bar{\sigma} \| \big(I + \bar{\sigma}^2 \rho_n^{(\mathrm{III})} \mathbf{\Lambda}^{-1}\big)^{-1/2} (\widehat{u}_n - \bar{u}) \|_{\ell^2} \sqrt{\frac{\log(1/\delta)}{n}} \\
&\leq c\bar{\sigma} \sqrt{D(\bar{\sigma}^2 \rho_n^{(\mathrm{III})}) \frac{\log(1/\delta)}{n}} \\
&= c\bar{\sigma} \sqrt{D(\rho_n^{(\mathrm{I})}) \frac{\log(1/\delta)}{n}}.
\end{aligned} \tag{5.62}$$

Combining equations (5.60), (5.61), and (5.62) completes the proof of Theorem 5.3.

5.5.4.1 Proof of equation (5.58a)

Define the infinite-dimensional vectors

$$\widetilde{\beta}_n := \Psi \widetilde{\mu}, \quad \text{and} \quad \beta_* := \Psi \mu^*.$$

The error can be written in the form of the basis function representation

$$\widetilde{\mu}_n(s_0, a_0) - \mu^*(s_0, a_0) = \langle \widetilde{\beta}_n - \beta_*,\, \phi(s_0, a_0) \rangle.$$

Invoking Lemma 5.2 with $z = \phi(s_0, a_0)$, we have

$$\begin{aligned}
&\Big| \langle \phi(s_0, a_0),\, \widehat{\beta}_n - \beta_* \rangle \Big| \\
&\leq c \| \big(I + \rho_n^{(\mathrm{I})} \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(s_0, a_0) \|_{\ell^2} \cdot \Big\{ \bar{\sigma} \sqrt{\frac{\log(1/\delta)}{n}} + \sigma \phi_{\max} \sqrt{D(\rho_n^{(\mathrm{I})})} \cdot \frac{\log(1/\delta) \log n}{n} \Big\},
\end{aligned}$$

holding true with probability $1-\delta$.

Recall the definition (5.19) of effective dimension, we have the uniform upper bound

$$\| \big(I + \rho_n^{(\mathrm{I})} \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x_0, a_0) \|_{\ell^2} \leq \sup_{(x,a)} \| \big(I + \rho_n^{(\mathrm{I})} \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x, a) \|_{\ell^2} \leq \sqrt{D(\rho_n^{(\mathrm{I})})}$$

Substituting back and using the sample-size condition (5.28a) completes the proof of equation (5.58a).

[10] Note that the vector $\widehat{u}_n$ is independent of $(X_i^{(\mathrm{III})}, A_i^{(\mathrm{III})}, Y_i^{(\mathrm{III})})_{i=1}^n$, so that Lemma 5.5 is applicable.

5.5.4.2 Proof of equation (5.58b)

Define the σ-field $\mathcal{B}_1 := \sigma\big(\{X_i^{(\mathrm{I})}, A_i^{(\mathrm{I})}, Y_i^{(\mathrm{I})}\}_{i=1}^n\big)$. Clearly, the first-stage regression function $\widetilde{\mu}_n$ is measurable in $\mathcal{B}_1$. Since each data-point $(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}, Y_i^{(\mathrm{II})})$ at the second stage is independent of $\mathcal{B}_1$, equation (5.58a) guarantees that

$$\mathbb{P}\left\{\big|\widetilde{\mu}_n(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}) - \mu^*(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})\big| \geq 2c\sigma\sqrt{\frac{D(\rho_n^{(\mathrm{I})})}{n}\log(n/\delta)}\right\} \leq \frac{\delta}{2n^2}.$$

Applying union bound and the tower property yields

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{P}\Big\{\max_{i\in[n]}\big|\widetilde{\mu}_n(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}) - \mu^*(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})\big| \geq 2c\sigma\sqrt{\frac{D(\rho_n^{(\mathrm{I})})}{n}\log(n/\delta)}\Big\} \mid \mathcal{B}_1\right] \\
&\leq \sum_{i=1}^n \mathbb{P}\left\{\big|\widetilde{\mu}_n(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}) - \mu^*(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})\big| \geq 2c\sigma\sqrt{\frac{D(\rho_n^{(\mathrm{I})})}{n}\log(n/\delta)}\right\} \leq \frac{\delta}{2n}. \qquad (5.63)
\end{aligned}$$

For the noisy observation we construct in this step, the conditional expectation takes the form

$$\mathbb{E}\big[(Y_i^{(\mathrm{II})} - \widetilde{\mu}_n(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}))^2 \mid X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}, \mathcal{B}_1\big] = \sigma^2(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}) + \underbrace{(\widetilde{\mu}_n - \mu^*)^2(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})}_{=:b(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})}.$$

We further note that for any $p > 0$, we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\big(Y_i^{(\mathrm{II})} - \widetilde{\mu}_n(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})\big)^2\Big|^p \mid X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}, \mathcal{B}_1\Big] \\
&\leq 2^{2p}\mathbb{E}\Big[\Big|\big(Y_i^{(\mathrm{II})} - \mu^*(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})\big)^2\Big|^p \mid X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}, \mathcal{B}_1\Big] + 2^{2p}\big|\widetilde{\mu}_n(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}) - \mu^*(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})\big|^{2p} \\
&\leq 4^{2p}p^p\sigma^{2p} + 2^{2p}b^p(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}),
\end{aligned}$$

which verifies the tail assumption $\|Y_i^{(\mathrm{II})} - \widetilde{\mu}_n(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}))^2\|_{\psi_1} \leq 4(\sigma^2 + b(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}))$ conditionally on $X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})}$ and $\mathcal{B}_1$.

Having verified the observation assumption (5.27), we are ready to apply the robust pointwise risk property satisfied by the estimating procedure $\mathcal{A}$. By definition, given a sample size $n \geq m\big(\underline{\sigma}^2/2, \delta/(2n)\big)$, for any $\varepsilon > 0$ and $\delta \in (0, 1)$, we have

$$\begin{aligned}
&\mathbb{P}\Big\{\,\big|\widehat{\sigma}_n^2(s_0, a_0) - \sigma^2(s_0, a_0)\big| \leq \underline{\sigma}^2/2 \mid \mathcal{B}_1\Big\} \\
&\qquad\qquad \geq \mathbb{P}\Big\{\max_{i\in[n]}|b(X_i^{(\mathrm{II})}, A_i^{(\mathrm{II})})| \leq \bar{b}(\underline{\sigma}^2/2, \delta/(2n)) \mid \mathcal{B}_1\Big\} - \frac{\delta}{2n}, \qquad (5.64)
\end{aligned}$$

almost surely.

Given a sample size satisfying the requirement in equation (5.28b), we have

$$\begin{aligned}
&\mathbb{P}\Big\{ \max_{i\in[n]} \big|b(X_i^{\text{(II)}}, A_i^{\text{(II)}})\big| \leq \bar{b}(\underline{\sigma}^2/2, \delta/(2n)) \mid \mathcal{B}_1 \Big\} \\
&\qquad \geq \mathbb{P}\Big\{ \max_{i\in[n]} \big|\widetilde{\mu}_n(X_i^{\text{(II)}}, A_i^{\text{(II)}}) - \mu^*(X_i^{\text{(II)}}, A_i^{\text{(II)}})\big| \leq c\sigma\sqrt{\frac{D(\rho_n^{\text{(I)}})}{n}\log(n/\delta)} \mid \mathcal{B}_1 \Big\}, \qquad (5.65)
\end{aligned}$$

almost surely.

Combining equations (5.63), (5.64), and (5.65) and taking expectations with respect to $(X_i^{\text{(I)}}, A_i^{\text{(I)}}, Y_i^{\text{(I)}})$, we conclude that

$$\begin{aligned}
&\mathbb{P}\Big\{ \big|\widehat{\sigma}_n^2(s_0, a_0) - \sigma^2(s_0, a_0)\big| \leq \underline{\sigma}^2/2 \Big\} \\
&= \mathbb{E}\Big[\mathbb{P}\Big\{ \big|\widehat{\sigma}_n^2(s_0, a_0) - \sigma^2(s_0, a_0)\big| \leq \underline{\sigma}^2/2 \mid \mathcal{B}_1 \Big\}\Big] \\
&\geq \mathbb{E}\left[\mathbb{P}\Big\{ \max_{i\in[n]} \big|\widetilde{\mu}_n(X_i^{\text{(II)}}, A_i^{\text{(II)}}) - \mu^*(X_i^{\text{(II)}}, A_i^{\text{(II)}})\big| \geq 2c\sigma\sqrt{\frac{D(\rho_n^{\text{(I)}})}{n}\log(n/\delta)}\Big\} \mid \mathcal{B}_1 \right] - \frac{\delta}{2n} \\
&\geq 1 - \frac{\delta}{n}.
\end{aligned}$$

On the event $|\widehat{\sigma}_n^2(s_0, a_0) - \sigma^2(s_0, a_0)| \leq \underline{\sigma}^2/2$, we have

$$\frac{1}{2}\sigma^2(s_0, a_0) \leq \sigma^2(s_0, a_0) - \underline{\sigma}^2/2 \leq \widehat{\sigma}_n^2(s_0, a_0) \leq \sigma^2(s_0, a_0) + \underline{\sigma}^2/2 \leq 2\sigma^2(s_0, a_0),$$

completing the proof of equation (5.58b).

5.5.4.3 Proof of Lemma 5.5

First, by the guarantee (5.58b) from the second stage and a union bound, we have

$$\mathbb{P}\Big\{\exists i \in [n],\ \frac{\widehat{\sigma}_n^2}{\sigma^2}(X_i^{\text{(III)}}, A_i^{\text{(III)}}) \notin \big(1, 2\big)\Big\} \leq \sum_{i=1}^{n} \mathbb{P}\Big\{\frac{\widehat{\sigma}_n^2}{\sigma^2}(X_i^{\text{(III)}}, A_i^{\text{(III)}}) \notin \big(1, 2\big)\Big\} \leq \delta. \qquad (5.66)$$

Defining the event

$$\mathscr{E}^{\text{(III)}} := \Big\{\frac{1}{2}\sigma^2(X_i^{\text{(III)}}, A_i^{\text{(III)}}) \leq \widehat{\sigma}_n^2(X_i^{\text{(III)}}, A_i^{\text{(III)}}) \leq 2\sigma^2(X_i^{\text{(III)}}, A_i^{\text{(III)}}), \quad \text{for any } i \in [n]\Big\},$$

we have $\mathbb{P}\big(\mathscr{E}^{\text{(III)}}\big) \geq 1 - \delta$, with respect to the randomness of both the state-action pairs $(X_i^{\text{(III)}}, A_i^{\text{(III)}})_{i=1}^n$ and the function $\widehat{\sigma}_n^2$.

The remainder of the proof is analogous to that of Lemma 5.2. For notational simplicity, we omit the supscript $^{\text{(III)}}$ in (X_i, A_i, Y_i). Under the basis function representation, we have

$$\begin{aligned}
\widehat{\beta}_n &= \Psi \cdot \arg\min_{\mu\in\mathbb{H}} \Big\{\frac{1}{n}\sum_{i=1}^{n} \widehat{\sigma}_n^{-2}(X_i, A_i)\big(Y_i - \mu(X_i, A_i)\big)^2 + \rho_n^{\text{(III)}}\|\mu\|_{\mathbb{H}}^2\Big\} \\
&= \arg\min_{\beta\in\ell^2(\mathbb{N})} \Big\{\frac{1}{n}\sum_{i=1}^{n} \widehat{\sigma}_n^{-2}(X_i, A_i)\big(Y_i - \langle\beta,\, \phi(X_i, A_i)\rangle\big)^2 + \rho_n^{\text{(III)}}\|\beta\|_{\lambda^{-1}}^2\Big\}.
\end{aligned}$$

Defining the noise $\varepsilon_i := Y_i - \mu^*(X_i, A_i)$ and the empirical covariance operator

$$\widehat{\mathbf{\Gamma}}_n^\sigma := \frac{1}{n} \sum_{i=1}^n \widehat{\sigma}_n^{-2}(X_i, A_i) \phi(X_i, A_i) \phi(X_i, A_i)^\top,$$

the error vector admits the representation

$$\widehat{\beta}_n - \beta_* = \big(\widehat{\mathbf{\Gamma}}_n^\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1} \frac{1}{n} \sum_{i=1}^n \Big\{ \widehat{\sigma}_n^{-2} \varepsilon_i \phi(X_i, A_i) - \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1} \beta_* \Big\}$$

We can bound such an error conditionally on the state-action pairs $(X_i, A_i)_{i=1}^n$ and the estimated conditional covariance function $\widehat{\sigma}_n$, as stated in the following lemma.

Lemma 5.6. *Under the set-up above, conditionally on the function $\widehat{\sigma}_n$ and the state-action pairs $(X_i, A_i)_{i=1}^n$ such that the event $\mathscr{E}^{\text{(III)}}$ happens, with probability $1 - \delta$, we have the upper bound*

$$\begin{aligned} \Big| z^\top (\widehat{\beta}_n - \beta_*) \Big| &\leq c \big\| \big(\widehat{\mathbf{\Gamma}}_n^\sigma + \rho_{n,*} \mathbf{\Lambda}^{-1}\big)^{-1/2} z \big\|_{\ell^2} \\ &\times \left\{ 2 \sqrt{\frac{\log(1/\delta)}{n}} + \frac{\log n \log(1/\delta) \sigma}{n \underline{\sigma}^2} \sup_{(x,a)} \big\| \big(\widehat{\mathbf{\Gamma}}_n^\sigma + \rho_{n,*} \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x, a) \big\|_{\ell^2} \right\} \end{aligned} \tag{5.67a}$$

See Appendix D.3.4 for the proof.

Taking this lemma as given, we proceed with the proof of Lemma 5.5. Define the truncated variance function and the corresponding reweighted operator.

$$\widetilde{\sigma}_n(x, a) := \begin{cases} \widehat{\sigma}_n^2(x, a) & \text{if } \frac{\widehat{\sigma}_n^2(x,a)}{\sigma^2(x,a)} \in \big(1/2, 2\big), \\ \sigma^2(x, a) & \text{otherwise} \end{cases} \quad \text{and}$$

$$\widetilde{\mathbf{\Gamma}}_n^\sigma := \frac{1}{n} \sum_{i=1}^n \frac{1}{\widetilde{\sigma}_n^2(X_i, A_i)} \phi(X_i, A_i) \phi(X_i, A_i)^\top.$$

Conditioned on the event $\mathscr{E}^{\text{(III)}}$, we have $\widetilde{\mathbf{\Gamma}}_n^\sigma = \widehat{\mathbf{\Gamma}}_n^\sigma$. On the other hand, we invoke Lemma 5.4 with the weight function $q = \widetilde{\sigma}_n^{-2}$ and $\omega = 1/2$. Note that the condition (5.56) becomes

$$(\bar{\sigma}^2 / \underline{\sigma}^2) \log \big(\frac{\kappa^2}{\rho_n^{\text{(III)}} \delta} \big) \cdot \frac{D\big(\rho_n^{\text{(III)}} \bar{\sigma}^2\big)}{n} \leq \frac{1}{32},$$

which is satisfied under the sample size requirement (5.28a) and regularization parameter choice (5.29). Therefore, on the event $\mathscr{E}^{\text{(III)}}$, with probability $1 - \delta$, we have

$$\frac{1}{2} \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1} \big) \preceq \widetilde{\mathbf{\Gamma}}_n^\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1} = \widehat{\mathbf{\Gamma}}_n^\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1} \preceq 2 \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1} \big). \tag{5.68}$$

Substituting equation (5.68) into the guarantee from Lemma 5.3, we find that

$$\Big|\langle z,\, \widehat{\beta}_n - \beta_* \rangle\Big| \leq 4c \| \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1/2} z \|_{\ell^2} \cdot \Bigg\{ \sqrt{\frac{\log(1/\delta)}{n}} + \frac{\log n \log(1/\delta) \sigma}{n \underline{\sigma}^2} \sup_{(x,a)} \| \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x,a) \|_{\ell^2} \Bigg\}.$$

By the definition (5.19) of effective dimension, we have

$$\sup_{(x,a)} \| \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x,a) \|_{\ell^2} \leq \bar{\sigma} \sup_{(x,a)} \| \big(I + \bar{\sigma}^2 \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x,a) \|_{\ell^2} = \bar{\sigma} \sqrt{D(\rho_n^{\text{(I)}})}.$$

Moreover, given the condition (5.28a) on the sample size, it follows that

$$\sqrt{\tfrac{\log(1/\delta)}{n}} \geq \tfrac{\log n \log(1/\delta) \sigma^2}{n \underline{\sigma}^2} \sqrt{D(\rho_n^{\text{(I)}})} \geq \tfrac{\log n \log(1/\delta) \sigma}{n \underline{\sigma}^2} \sup_{(x,a)} \| \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x,a) \|_{\ell^2}.$$

Thus, we conclude that

$$\Big|\langle z,\, \widehat{\beta}_n - \beta_* \rangle\Big| \leq 8c \| \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1/2} z \|_{\ell^2} \sqrt{\frac{\log(1/\delta)}{n}}$$

with probability $1 - \delta$, with establishes the claim in Lemma 5.5.

5.5.4.4 Proof of equation (5.32)

Recall the definition (5.17a) of the infinite-dimensional vector $\bar{u}(\delta_{x_0})$

$$\bar{u}(\delta_{x_0}) = \int_{\mathbb{A}} \phi(x_0, a) d\omega(a \mid x_0).$$

Using this definition, the estimation error admits a basis-function representation

$$\widehat{\tau}_n(x_0) - \tau^*(x_0) = \langle \widehat{\beta}_n - \beta_*,\, \widetilde{u}(x_0) \rangle,$$

where the vectors $\widehat{\beta}_n$ and β_* are defined in Section 5.5.3.1.

Applying Lemma 5.5 with $z = \widetilde{u}(x_0)$ yields

$$\Big|\langle \widetilde{u}(x_0),\, \widehat{\beta}_n - \beta_* \rangle\Big| \leq c \| \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1/2} \widetilde{u}(x_0) \|_{\ell^2} \sqrt{\frac{\log(1/\delta)}{n}}. \tag{5.69}$$

By Proposition 5.1, we have

$$\| \big(\mathbf{\Gamma}_\sigma + \rho_n^{\text{(III)}} \mathbf{\Lambda}^{-1}\big)^{-1/2} \widetilde{u}(x_0) \|_{\ell^2}^2 = \widetilde{u}^\top(x_0) \big(\mathbf{\Gamma}_\sigma + \tfrac{1}{R^2 n} \mathbf{\Lambda}^{-1}\big)^{-1} \widetilde{u} \leq 4 V_{\sigma,n}^2(\delta_{x_0}, \pi, g; \mathcal{F}).$$

Substituting back completes the proof of the claim (5.32).

5.6 Discussion

In this chapter, we studied the problem of estimating linear functionals based on observational data. Our main focus was the challenging setting in which the importance ratio is poorly behaved. In such settings, the classical semi-parametric efficiency bound—based on a presumptive $\sqrt{n}$-rate of convergence—can be infinite, and so fail to characterize the problem. So as to remedy this deficiency, the main contribution of this chapter was to propose a modified risk functional, defined as the optimal value of a variational problem that respects the geometry of the function class. The resulting minimax risks interpolate between the classical regimes of semi-parametric efficiency with the $\sqrt{n}$-rate, and nonparametric rates for functional estimation. Focusing on the case of RKHS, we analyze an outcome-based regression estimator, and showed that it achieves our instance-dependent lower bound (up to a universal constant pre-factor). This estimator is attractive in not requiring any knowledge of the behavioral policy. Nonetheless, despite its agnostic nature, it matches our lower bound that applies even to oracle estimators that have full knowledge of the policy. When applied to various off-policy estimation problems with singularities in the importance ratio, our results uncover a novel class of minimax rates, as well as instance-dependent optimality, adaptively achieved by our estimators.

While this chapter takes an initial step in characterizing instance-dependent optimality for off-policy estimation beyond semi-parametric efficiency, there are many open directions.

- Our optimality results impose assumptions on the conditional variance function σ^2. We either require it to be uniformly bounded (for achieving the worst-case variance bound $V_\sigma^2(\frac{dg}{d\pi})$), or require additional structure that allows for consistent estimation (for optimal adaptation to the conditional variance structure). It is not clear if such requirements are necessary. In the classical $\sqrt{n}$-regime of semi-parametric efficiency, regime, AIPW estimators adapt to the conditional variance structure without knowledge of σ^2; for instance, see the paper [41]. An important open question, therefore, is whether such adaptivity is possible in the more challenging regime considered by this chapter without additional assumptions on the conditional variance.

- In this chapter, we established achievability of our lower bounds only for reproducing kernel Hilbert spaces. Thus, an important question is to what extent our results can be extended to more general function classes. We conjecture that the minimax linear estimation strategy [53, 78] could yield an optimal estimator–in the same sense as the results presented in this chapter—for any function class $\mathcal{F}$ satisfying the Donsker property. For non-Donsker classes, it is known from past work [179] that knowledge of the behavior policy plays a role. An important direction of future research, therefore, is to identify the optimal risk for estimation, jointly determined by the structural assumptions on the treatment effect function, the behavior policy function, and singularities in the importance ratio function.

- Our results focus on the classical off-policy contextual bandit setup, where the data $(X_i, A_i, Y_i)_{i=1}^n$ are independent and identically distributed. However, many decision-making problems involve collecting data in an adaptive manner (e.g., by running a bandit algorithm), or following a Markov chain (e.g., in reinforcement learning). The importance ratio can easily grow unbounded in these settings, leading to practical challenges [225, 97]. Being optimally agnostic to the singular behavior of the importance ratio, we suspect that our estimation framework and risk functional should be helpful for problems of this type.

Bibliography

[1] A. Abadie and G. W. Imbens. “Matching on the estimated propensity score”. In: *Econometrica* 84.2 (2016), pp. 781–807.

[2] R. Adamczak. “A tail inequality for suprema of unbounded empirical processes with applications to Markov chains”. In: *Electronic Journal of Probability* 13 (2008), pp. 1000–1034.

[3] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. “Scale-sensitive dimensions, uniform convergence, and learnability”. In: *Journal of the ACM (JACM)* 44.4 (1997), pp. 615–631.

[4] T. B. Armstrong and M. Kolesár. “Finite-sample optimal estimation and inference on average treatment effects under unconfoundedness”. In: *Econometrica* 89.3 (2021), pp. 1141–1177.

[5] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb. “Solving inverse problems using data-driven models”. In: *Acta Numerica* 28 (2019), pp. 1–174.

[6] O. Ashenfelter. “Estimating the effect of training programs on earnings”. In: *The Review of Economics and Statistics* (1978), pp. 47–57.

[7] S. Athey and S. Wager. “Policy learning with observational data”. In: *Econometrica* 89.1 (2021), pp. 133–161.

[8] P. L. Bartlett, O. Bousquet, and S. Mendelson. “Local Rademacher complexities”. In: *The Annals of Statistics* 33.4 (2005), pp. 1497–1537.

[9] P. L. Bartlett, P. M. Long, and R. C. Williamson. “Fat-shattering and the learnability of real-valued functions”. In: *Proceedings of the seventh annual conference on Computational learning theory*. 1994, pp. 299–310.

[10] P. C. Bellec. “Sharp oracle inequalities for least squares estimators in shape restricted regression”. In: *The Annals of Statistics* 46.2 (2018), pp. 745–780.

[11] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations*. Vol. 22. Springer Science & Business Media, 2012.

[12] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media, 2011.

[13] D. P. Bertsekas. “Proximal algorithms and temporal differences for large linear systems: extrapolation, approximation, and simulation”. In: *arXiv preprint arXiv:1610.05427* (2016).

[14] D. P. Bertsekas. *Reinforcement Learning and Optimal Control*. Athena Scientific Belmont, MA, 2019.

[15] D. P. Bertsekas. “Temporal difference methods for general projected equations”. In: *IEEE Transactions on Automatic Control* 56.9 (2011), pp. 2128–2139.

[16] J. Bhandari, D. Russo, and R. Singal. “A finite time analysis of temporal difference learning with linear function approximation”. In: *Operations Research* 69.3 (2021), pp. 950–973.

[17] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Vol. 4. Springer, 1993.

[18] Patrick Billingsley. “Statistical methods in Markov chains”. In: *The Annals of Mathematical Statistics* (1961), pp. 12–40.

[19] V. S. Borkar. “A concentration bound for contractive stochastic approximation”. In: *Systems & Control Letters* 153 (2021), p. 104947.

[20] V. S. Borkar. *Stochastic Approximation: a Dynamical Systems Viewpoint*. Vol. 48. Springer, 2009.

[21] L. Bottou, F. E. Curtis, and J. Nocedal. “Optimization methods for large-scale machine learning”. In: *SIAM Review* 60.2 (2018), pp. 223–311.

[22] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: a Nonasymptotic Theory of Independence*. Oxford university press, 2013.

[23] O. Bousquet, D. Kane, and S. Moran. “The optimal approximation factor in density estimation”. In: *Conference on Learning Theory*. 2019, pp. 318–341.

[24] J. A. Boyan. “Technical update: Least-squares temporal difference learning”. In: *Machine Learning* 49.2-3 (2002), pp. 233–246.

[25] J. Bradic, V. Chernozhukov, W. K. Newey, and Y. Zhu. “Minimax semiparametric learning with approximate sparsity”. In: *arXiv preprint arXiv:1912.12213* (2019).

[26] J. Bradic, S. Wager, and Y. Zhu. “Sparsity double robust inference of average treatment effects”. In: *arXiv preprint arXiv:1905.00744* (2019).

[27] S. J. Bradtke and A. G. Barto. “Linear least-squares algorithms for temporal difference learning”. In: *Machine Learning* 22.1-3 (1996), pp. 33–57.

[28] S. Brenner and R. Scott. *The Mathematical Theory of Finite Element Methods*. Vol. 15. Springer Science & Business Media, 2007.

[29] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Science & Business Media, 2009.

[30] L. D. Brown, M. G. Low, and L. H. Zhao. “Superefficiency in nonparametric function estimation”. In: *The Annals of Statistics* 25.6 (1997), pp. 2607–2625.

[31] F. Bunea, A. B. Tsybakov, and M. H. Wegkamp. "Aggregation for Gaussian regression". In: *The Annals of Statistics* 35.4 (2007), pp. 1674–1697.

[32] F. Bunea, A. B. Tsybakov, and M. H. Wegkamp. "Sparsity oracle inequalities for the Lasso". In: *Electronic Journal of Statistics* 1 (2007), pp. 169–194.

[33] A. Caponnetto and E. De Vito. "Optimal rates for the regularized least-squares algorithm". In: *Foundations of Computational Mathematics* 7.3 (2007), pp. 331–368.

[34] I. Castillo. "Semi-parametric second-order efficient estimation of the period of a signal". In: *Bernoulli* 13.4 (2007), pp. 910–932.

[35] J. Céa. "Approximation variationnelle des problèmes aux limites". In: *Annales de l'institut Fourier*. Vol. 14. 1964, pp. 345–444.

[36] S. O. Chan, I. Diakonikolas, R. A. Servedio, and X. Sun. "Near-optimal density estimation in near-linear time using variable-width histograms". In: *Advances in Neural Information Processing Systems*. 2014, pp. 1844–1852.

[37] S. Chen, A. Devraj, A. Busic, and S. Meyn. "Explicit mean-square error bounds for Monte-Carlo and linear stochastic approximation". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 4173–4183.

[38] X. Chen, H. Hong, and A. Tarozzi. "Semiparametric efficiency in GMM models of nonclassical measurement errors, missing data and treatment effects". In: *Technical report* (2004).

[39] X. Chen, J. D. Lee, X. T. Tong, and Y. Zhang. "Statistical inference for model parameters in stochastic gradient descent". In: *The Annals of Statistics* 48.1 (2020), pp. 251–273.

[40] Z. Chen, S. T. Maguluri, S. Shakkottai, and K. Shanmugam. "A Lyapunov theory for finite-sample guarantees of asynchronous Q-learning and TD-learning variants". In: *arXiv preprint arXiv:2102.01567* (2021).

[41] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. "Double/debiased machine learning for treatment and structural parameters". In: *The Econometrics Journal* 21.1 (2018).

[42] A. D'Amour, P. Ding, A. Feller, L. Lei, and J. Sekhon. "Overlap in observational studies with high-dimensional covariates". In: *Journal of Econometrics* 221.2 (2021), pp. 644–654.

[43] A. S. Dalalyan, G. K. Golubev, and A. B. Tsybakov. "Penalized maximum likelihood and semiparametric second-order efficiency". In: *The Annals of Statistics* 34.1 (2006), pp. 169–201.

[44] A. S. Dalalyan and J. Salmon. "Sharp oracle inequalities for aggregation of affine estimators". In: *The Annals of Statistics* 40.4 (2012), pp. 2327–2355.

[45] A. S. Dalalyan and M. Sebbar. “Optimal Kullback–Leibler aggregation in mixture density estimation by maximum likelihood”. In: *Mathematical Statistics and Learning* 1.1 (2018), pp. 1–35.

[46] S. Darolles, Y. Fan, J.-P. Florens, and E. Renault. “Nonparametric instrumental regression”. In: *Econometrica* 79.5 (2011), pp. 1541–1565.

[47] C. Daskalakis, N. Dikkala, and N. Gravin. “Testing symmetric Markov chains from a single trajectory”. In: *Conference On Learning Theory*. PMLR. 2018, pp. 385–409.

[48] P. Dayan and T. J. Sejnowski. “TD(λ) converges with probability 1”. In: *Machine Learning* 14.3 (1994), pp. 295–301.

[49] V. Debavelaere, S. Durrleman, and S. Allassonnière. “On the convergence of stochastic approximations under a subgeometric ergodic Markov dynamic”. In: *Electronic Journal of Statistics* 15.1 (2021), pp. 1583–1609.

[50] A. Dieuleveut, A. Durmus, and F. Bach. “Bridging the gap between constant step size stochastic gradient descent and Markov chains”. In: *The Annals of Statistics* 48.3 (2020), pp. 1348–1382.

[51] S. Dirksen. “Tail bounds via generic chaining”. In: *Electronic Journal of Probability* 20 (2015), pp. 1–29.

[52] T. T. Doan, L. M. Nguyen, N. H. Pham, and J. Romberg. “Finite-time analysis of stochastic gradient descent under Markov randomness”. In: *arXiv preprint arXiv:2003.10973* (2020).

[53] D. L. Donoho. “Statistical estimation and optimal recovery”. In: *The Annals of Statistics* 22.1 (1994), pp. 238–270.

[54] Y. Duan, M. Wang, and M. J. Wainwright. “Optimal policy evaluation using kernel-based temporal difference methods”. In: *arXiv preprint arXiv:2109.12002* (2021).

[55] J. Duchi and F. Ruan. “Asymptotic optimality in stochastic optimization”. In: *The Annals of Statistics* 49.1 (2021), pp. 21–48.

[56] A. Durmus, É Moulines, A. Naumov, S. Samsonov, and H-T. Wai. “On the stability of random matrix product with Markovian noise: application to linear stochastic approximation and TD learning”. In: *Conference on Learning Theory*. PMLR. 2021, pp. 1711–1752.

[57] R. Durrett. *Random Graph Dynamics*. Vol. 200. Citeseer, 2007.

[58] L. C. Evans. *Partial differential equations*. Vol. 19. American Mathematical Soc., 2010.

[59] E. Even-Dar and Y. Mansour. “Learning rates for Q-learning”. In: *Journal of Machine Learning Research* 5 (2003), pp. 1–25.

[60] C. A. J. Fletcher. *Computational Galerkin Methods*. Springer, 1984.

[61] G. Fort. "Central limit theorems for stochastic approximation with controlled Markov chain dynamics". In: *ESAIM: Probability and Statistics* 19 (2015), pp. 60–80.

[62] D. J. Foster and V. Syrgkanis. "Orthogonal statistical learning". In: *arXiv preprint arXiv:1901.09036* (2019).

[63] M. Frölich. "Finite-sample properties of propensity-score matching and weighting estimators". In: *Review of Economics and Statistics* 86.1 (2004), pp. 77–90.

[64] S. Gadat and F. Panloup. "Optimal non-asymptotic bound of the Ruppert–Polyak averaging without strong convexity". In: *arXiv preprint arXiv:1709.03342* (2017).

[65] B. G. Galerkin. "Series solution of some problems of elastic equilibrium of rods and plates". In: *Vestnik inzhenerov i tekhnikov* 19.7 (1915), pp. 897–908.

[66] Z. Gao and Y. Han. "Minimax optimal nonparametric estimation of heterogeneous treatment effects". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 21751–21762.

[67] S. Ghadimi and G. Lan. "Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization I: A generic algorithmic framework". In: *SIAM Journal on Optimization* 22.4 (2012), pp. 1469–1492.

[68] R. D. Gill and B. Ya. Levit. "Applications of the van Trees inequality: a Bayesian Cramér-Rao bound". In: *Bernoulli* 1.1-2 (1995), pp. 59–79.

[69] M. Giordano and R. Nickl. "Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem". In: *Inverse Problems* 36.8 (2020), p. 085001.

[70] P. E. Greenwood and W. Wefelmeyer. "Efficiency of empirical estimators for Markov chains". In: *The Annals of Statistics* (1995), pp. 132–143.

[71] J. Hahn. "On the role of the propensity score in efficient semiparametric estimation of average treatment effects". In: *Econometrica* (1998), pp. 315–331.

[72] J. Hájek. "Local asymptotic minimax and admissibility in estimation". In: *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability.* Vol. 1. 1972, pp. 175–194.

[73] J. D. Hamilton. *Time series analysis.* Princeton university press, 2020.

[74] Q. Han and C.-H. Zhang. "Limit distribution theory for block estimators in multiple isotonic regression". In: *The Annals of Statistics* 48.6 (2020), pp. 3251–3282.

[75] T. Hastie, R. Tibshirani, and M. J. Wainwright. *Statistical learning with sparsity: The Lasso and generalizations.* New York: CRC Press, Chapman and Hall, 2015.

[76] K. Hirano, G. W. Imbens, and G. Ridder. "Efficient estimation of average treatment effects using the estimated propensity score". In: *Econometrica* 71.4 (2003), pp. 1161–1189.

[77] D. A. Hirshberg, A. Maleki, and J. R. Zubizarreta. "Minimax linear estimation of the retargeted mean". In: *arXiv preprint arXiv:1901.10296* (2019).

[78] D. A. Hirshberg and S. Wager. "Augmented minimax linear estimation". In: *The Annals of Statistics* 49.6 (2021), pp. 3206–3227.

[79] K. Hitomi, Y. Nishiyama, and R. Okui. "A puzzling phenomenon in semiparametric estimation problems with infinite-dimensional nuisance parameters". In: *Econometric Theory* 24.6 (2008), pp. 1717–1728.

[80] H. Hong, M. P. Leung, and J. Li. "Inference on finite-population treatment effects under limited overlap". In: *The Econometrics Journal* 23.1 (2020), pp. 32–47.

[81] D. Hsu, A. Kontorovich, D. A. Levin, Y. Peres, Cs. Szepesvári, and G. Wolfer. "Mixing time estimation in reversible Markov chains from a single sample path". In: *The Annals of Applied Probability* 29.4 (2019), pp. 2439–2480.

[82] K. Jiang, R. Mukherjee, S. Sen, and P. Sur. "A new central limit theorem for the augmented IPW estimator: variance inflation, cross-Fit covariance and beyond". In: *arXiv preprint arXiv:2205.10198* (2022).

[83] N. Jiang and L. Li. "Doubly robust off-policy value evaluation for reinforcement learning". In: *International Conference on Machine Learning*. PMLR. 2016, pp. 652–661.

[84] R. Johnson and T. Zhang. "Accelerating stochastic gradient descent using predictive variance reduction". In: *Advances in Neural Information Processing Systems* 26 (2013), pp. 315–323.

[85] M. Kaledin, É. Moulines, A. Naumov, V. Tadic, and H.-T. Wai. "Finite time analysis of linear two-timescale stochastic approximation with Markovian noise". In: *Conference on Learning Theory*. PMLR. 2020, pp. 2144–2203.

[86] O. Kallenberg. *Foundations of Modern Probability*. Vol. 2. Springer, 1997.

[87] N. Kallus. "Balanced policy evaluation and learning". In: *Advances in Neural Information Processing Systems* 31 (2018).

[88] N. Kallus. "Generalized optimal matching methods for causal inference". In: *Journal of Machine Learning Research* 21 (2020), pp. 62–1.

[89] N. Kallus and M. Uehara. "Double reinforcement learning for efficient off-policy evaluation in markov decision processes". In: *Journal of Machine Learning Research* 21.167 (2020).

[90] N. Kallus and M. Uehara. "Efficiently breaking the curse of horizon in off-policy evaluation with double reinforcement learning". In: *Operations Research* (2022).

[91] B. Kaltenbacher, A. Kirchner, and B. Vexler. "Adaptive discretizations for the choice of a Tikhonov regularization parameter in nonlinear inverse problems". In: *Inverse Problems* 27.12 (2011), p. 125008.

[92] J. D. Y. Kang and J. L. Schafer. "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data". In: *Statistical Science* 22.4 (2007), pp. 523–539.

[93] B. Karimi, B. Miasojedow, É Moulines, and H.-T. Wai. "Non-asymptotic analysis of biased stochastic approximation scheme". In: *Conference on Learning Theory*. PMLR. 2019, pp. 1944–1974.

[94] P. Karmakar and S. Bhatnagar. "Two time-scale stochastic approximation with controlled Markov noise and off-policy temporal-difference learning". In: *Mathematics of Operations Research* 43.1 (2018), pp. 130–151.

[95] M. J. Kearns and R. E. Schapire. "Efficient distribution-free learning of probabilistic concepts". In: *Journal of Computer and System Sciences* 48.3 (1994), pp. 464–497.

[96] E. H. Kennedy, S. Balakrishnan, and L. Wasserman. "Minimax rates for heterogeneous causal effect estimation". In: *arXiv preprint arXiv:2203.00837* (2022).

[97] K. Khamaru, Y. Deshpande, L. Mackey, and M. J. Wainwright. "Near-optimal inference in adaptive linear regression". In: *arXiv preprint arXiv:2107.02266* (2021).

[98] K. Khamaru, A. Pananjady, F. Ruan, M. J. Wainwright, and M. I. Jordan. "Is temporal difference learning optimal? An instance-dependent analysis". In: *SIAM Journal on Mathematics of Data Science* 3.4 (2021), pp. 1013–1040.

[99] K. Khamaru, E. Xia, M. J. Wainwright, and M. I. Jordan. "Instance-optimality in optimal value estimation: Adaptivity via variance-reduced Q-learning". In: *arXiv preprint arXiv:2106.14352* (2021).

[100] S. Khan and E. Tamer. "Irregular identification, support conditions, and inverse weight estimation". In: *Econometrica* 78.6 (2010), pp. 2021–2042.

[101] R. Z. Khas'minskii and I. A. Ibragimov. "Asymptotically efficient nonparametric estimation of functionals of a spectral density function". In: *Probability Theory and Related Fields* 73.3 (1986), pp. 447–461.

[102] O. Klopp, A. B. Tsybakov, and N. Verzelen. "Oracle inequalities for network models and sparse graphon estimation". In: *The Annals of Statistics* 45.1 (2017), pp. 316–354.

[103] A. N. Kolmogorov and V. M. Tikhomirov. "ε-entropy and ε-capacity of sets in function spaces". Russian. In: *Usp. Mat. Nauk* 14.2(86) (1959), pp. 3–86. ISSN: 0042-1316.

[104] V. Koltchinskii. "Local Rademacher complexities and oracle inequalities in risk minimization". In: *The Annals of Statistics* 34.6 (2006), pp. 2593–2656.

[105] V. Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008*. Vol. 2033. Springer Science & Business Media, 2011.

[106] V. Koltchinskii and S. Mendelson. “Bounding the smallest singular value of a random matrix without concentration”. In: *International Mathematics Research Notices* 2015.23 (2015), pp. 12991–13008.

[107] V. R. Konda and J. N. Tsitsiklis. “Actor-critic algorithms”. In: *Advances in Neural Information Processing Systems*. 2000, pp. 1008–1014.

[108] G. Kotsalis, G. Lan, and T. Li. “Simple and optimal methods for stochastic variational inequalities, II: Markovian noise and policy evaluation in reinforcement learning”. In: *SIAM Journal on Optimization* 32.2 (2022), pp. 1120–1155.

[109] S. Kowshik, D. Nagaraj, P. Jain, and P. Netrapalli. “Streaming linear system identification with reverse experience replay”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 30140–30152.

[110] M. A. Krasnosel’skiı, G. M. Vaınikko, R. P. Zabreıko, Ya. B. Ruticki, and V. Ya. Stet’senko. *Approximate Solution of Operator Equations*. Translated from Russian by D. Louvish. Wolters-Noordhoff Publishing, Groningen, 1972.

[111] H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Vol. 35. Springer Science & Business Media, 2003.

[112] Harold J. Kushner and Dean S. Clark. *Stochastic approximation methods for constrained and unconstrained systems*. Applied Mathematical Sciences. Springer-Verlag, New York-Berlin, 1978.

[113] Y. A. Kutoyants. “Efficiency of the empirical distribution for ergodic diffusion”. In: *Bernoulli* (1997), pp. 445–456.

[114] T. L. Lai. “Stochastic Approximation”. In: *The Annals of Statistics* 31.2 (2003), pp. 391–406.

[115] C. Lakshminarayana and Cs. Szepesvári. “Linear stochastic approximation: How far does constant step-size and iterate averaging go?” In: *International Conference on Artificial Intelligence and Statistics*. 2018, pp. 1347–1355.

[116] S. Larsson and V. Thomée. *Partial Differential Equations with Numerical Methods*. Vol. 45. Springer Science & Business Media, 2008.

[117] L. Le Cam. “Locally asymptotically normal families of distributions”. In: *Univ. California Publ. Statist.* 3 (1960), pp. 37–98.

[118] O. V. Lepskii. “On a problem of adaptive estimation in Gaussian white noise”. In: *Theory of Probability & Its Applications* 35.3 (1991), pp. 454–466.

[119] B. Ya. Levit. “Conditional estimation of linear functionals”. In: *Problemy Peredachi Informatsii* 11.4 (1975), pp. 39–54.

[120] B. Ya. Levit. “Infinite-dimensional informational lower bounds”. In: *Theor. Prob. Appl* 23 (1978), pp. 388–394.

[121] C. J. Li, W. Mou, M. Wainwright, and M. Jordan. “ROOT-SGD: Sharp Nonasymptotics and Asymptotic Efficiency in a Single Algorithm”. In: *Conference on Learning Theory*. PMLR. 2022, pp. 909–981.

[122] G. Li, Y. Wei, Y. Chi, Y. Gu, and Y. Chen. "Breaking the Sample Size Barrier in Model-Based Reinforcement Learning with a Generative Model". In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 12861–12872.

[123] L. Li, R. Munos, and Cs. Szepesvári. "Toward minimax off-policy value estimation". In: *Artificial Intelligence and Statistics*. PMLR. 2015, pp. 608–616.

[124] S. Li, T. T. Cai, and H. Li. "Transfer learning for high-dimensional linear regression: Prediction, estimation, and minimax optimality". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.1 (2022), pp. 149–173.

[125] T. Li, G. Lan, and A. Pananjady. "Accelerated and instance-optimal policy evaluation with linear function approximation". In: *SIAM Journal on Mathematics of Data Science* 5.1 (2023), pp. 174–200.

[126] X. Li, M. Wang, and A. Zhang. "Estimation of Markov chain via rank-constrained likelihood". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 3033–3042.

[127] L. Ljung. "Analysis of recursive stochastic algorithms". In: *IEEE Transactions on Automatic Control* 22.4 (1977), pp. 551–575.

[128] L. Ljung. "On positive real transfer functions and the convergence of some recursive schemes". In: *IEEE Transactions on Automatic Control* 22.4 (1977), pp. 539–551.

[129] J. K. Lunceford and M. Davidian. "Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study". In: *Statistics in Medicine* 23.19 (2004), pp. 2937–2960.

[130] R. Lung, Y. Wu, D. Kamilis, and N. Polydorides. "A sketched finite element method for elliptic models". In: *Computer Methods in Applied Mechanics and Engineering* 364 (2020), p. 112933.

[131] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. New York: Springer, 2005.

[132] C. Ma, R. Pathak, and M. J. Wainwright. "Optimally tackling covariate shift in RKHS-based nonparametric regression". In: *The Annals of Statistics* 51.2 (2023), pp. 738–761.

[133] C. Ma, B. Zhu, J. Jiao, and M. J. Wainwright. "Minimax off-Policy evaluation for multi-armed bandits". In: *IEEE Transactions on Information Theory* (2022).

[134] X. Ma and J. Wang. "Robust inference using inverse probability weighting". In: *Journal of the American Statistical Association* 115.532 (2020), pp. 1851–1860.

[135] L. Mackey, V. Syrgkanis, and I. Zadik. "Orthogonal machine learning: Power and limitations". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 3375–3383.

[136] P. Massart. *Concentration Inequalities and Model Selection.* Vol. 6. Springer, 2007.

[137] P. Massart and É. Nédélec. "Risk bounds for statistical learning". In: *The Annals of Statistics* 34.5 (2006), pp. 2326–2366.

[138] S. Mendelson. "Learning without concentration". In: *Journal of the ACM (JACM)* 62.3 (2015), pp. 1–25.

[139] S. Mendelson and J. Neeman. "Regularization in kernel learning". In: *The Annals of Statistics* 38.1 (2010), pp. 526–565.

[140] S. Mendelson and R. Vershynin. "Entropy, combinatorial dimensions and random averages". In: *International Conference on Computational Learning Theory.* Springer. 2002, pp. 14–28.

[141] J. Mercer. "Functions of positive and negative type and their connection with the theory of integral equations". In: *Philos. Trans. Royal Soc* 209 (1909), pp. 4–415.

[142] M. Métivier and P. Priouret. "Applications of a Kushner and Clark lemma to general classes of stochastic algorithms". In: *IEEE Transactions on Information Theory* 30.2 (1984), pp. 140–151.

[143] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Springer Science & Business Media, 2012.

[144] S. Minsker. "On some extensions of Bernstein's inequality for self-adjoint operators". In: *Statistics & Probability Letters* 127 (2017), pp. 111–119.

[145] W. Mou, P. Ding, M. J. Wainwright, and P. L. Bartlett. "Kernel-based off-policy estimation without overlap: instance-dependent optimality beyond semiparametric efficiency". In: *arXiv preprint arXiv:2301.06240* (2023).

[146] W. Mou, N. Flammarion, M. J. Wainwright, and P. L. Bartlett. "An efficient sampling algorithm for non-smooth composite potentials". In: *Journal of Machine Learning Research* (2022).

[147] W. Mou, N. Flammarion, M. J. Wainwright, and P. L. Bartlett. "Improved bounds for discretization of Langevin diffusions: Near-optimal rates without convexity". In: *Bernoulli* 28.3 (2022), pp. 1577–1601.

[148] W. Mou, N. Ho, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan. "A diffusion process perspective on posterior contraction rates for parameters". In: *arXiv preprint arXiv:1909.00966* (2019).

[149] W. Mou, K. Khamaru, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan. "Optimal variance-reduced stochastic approximation in Banach spaces". In: *arXiv preprint arXiv:2201.08518* (2022).

[150] W. Mou, C. J. Li, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan. "On linear stochastic approximation: Fine-grained Polyak–Ruppert and non-asymptotic concentration". In: *Proceedings of Thirty Third Conference on Learning Theory.* Vol. 125. 2020, pp. 2947–2997.

[151] W. Mou, Y.-A. Ma, M. J. Wainwright, P. L. Bartlett, and M. I. Jordan. "High-order Langevin diffusion yields an accelerated MCMC algorithm." In: *Journal of Machine Learning Research* 22 (2021), pp. 42–1.

[152] W. Mou, A. Pananjady, and M. J. Wainwright. "Optimal oracle inequalities for solving projected fixed-point equations, with applications to policy evaluation". In: *Mathematics of Operations Research* (2022). Article in Advance.

[153] W. Mou, A. Pananjady, M. J. Wainwright, and P. L. Bartlett. "Optimal and instance-dependent guarantees for Markovian linear stochastic approximation". In: *arXiv preprint arXiv:2112.12770* (2021). Extended abstract appeared at *COLT 2022*.

[154] W. Mou, A. Pananjady, M. J. Wainwright, and P. L. Bartlett. "Policy Evaluation via Projected Fixed-Points I: Statistical Oracle Inequalities". In: *preprint* (2023).

[155] W. Mou, M. J. Wainwright, and P. L. Bartlett. "Off-policy estimation of linear functionals: non-asymptotic theory for semi-parametric efficiency". In: *arXiv preprint arXiv: 2209.13075* (2022).

[156] W. Mou, Z. Wen, and X. Chen. "On the sample complexity of reinforcement learning with policy space generalization". In: *arXiv preprint arXiv:2008.07353* (2020).

[157] É. Moulines and F. R. Bach. "Non-asymptotic analysis of stochastic approximation algorithms for machine learning". In: *Advances in Neural Information Processing Systems*. 2011, pp. 451–459.

[158] R. Munos and Cs. Szepesvári. "Finite-time bounds for fitted value iteration". In: *Journal of Machine Learning Research* 9.May (2008), pp. 815–857.

[159] D. Nagaraj, X. Wu, G. Bresler, P. Jain, and P. Netrapalli. "Least squares regression with Markovian data: fundamental limits and algorithms". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 16666–16676.

[160] A. Nemirovski. "Lectures on modern convex optimization". In: *Society for Industrial and Applied Mathematics (SIAM*. Citeseer. 2001.

[161] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. "Robust stochastic approximation approach to stochastic programming". In: *SIAM Journal on Optimization* 19.4 (2009), pp. 1574–1609.

[162] R. Nickl. "On Bayesian inference for some statistical inverse problems with partial differential equations". In: *Bernoulli News* 24.2 (2017), pp. 5–9.

[163] X. Nie and S. Wager. "Quasi-oracle estimation of heterogeneous treatment effects". In: *Biometrika* 108.2 (2021), pp. 299–319.

[164] A. Pananjady and M. J. Wainwright. "Instance-dependent ℓ_∞-bounds for policy evaluation in tabular reinforcement learning". In: *IEEE Transactions on Information Theory* 67.1 (2021), pp. 566–585.

[165] S. Penev. “Efficient estimation of the stationary distribution for exponentially ergodic Markov chains”. In: *Journal of Statistical Planning and Inference* 27.1 (1991), pp. 105–123.

[166] M. Penrose. *Random Geometric Graphs*. Vol. 5. Oxford University Press, 2003.

[167] M. S. Pinsker. “Optimal filtering of square-integrable signals in Gaussian noise”. In: *Problemy Peredachi Informatsii* 16.2 (1980), pp. 52–68.

[168] B. T. Polyak. “A new method of stochastic approximation type”. In: *Automat. i Telemekh* 7.98-107 (1990), p. 2.

[169] B. T. Polyak and A. B. Juditsky. “Acceleration of stochastic approximation by averaging”. In: *SIAM Journal on Control and Optimization* 30.4 (1992), pp. 838–855.

[170] N. Polydorides, M. Wang, and D. P. Bertsekas. “A quasi Monte Carlo Method for Large-Scale Inverse Problems”. In: *Monte Carlo and Quasi-Monte Carlo Methods 2010*. Springer, 2012, pp. 623–637.

[171] N. Polydorides, M. Wang, and D. P. Bertsekas. “Approximate solution of large-scale linear inverse problems with Monte Carlo simulation”. In: *Lab. for Information and Decision Systems Report, MIT* (2009).

[172] A. Rakhlin, K. Sridharan, and A. B. Tsybakov. “Empirical entropy, minimax regret and minimax risk”. In: *Bernoulli* 23.2 (2017), pp. 789–824.

[173] P. Rigollet and J.-C. Hütter. “High Dimensional Statistics”. In: *Lecture notes for course 18S997* (2015).

[174] H. Robbins and S. Monro. “A stochastic approximation method”. In: *The Annals of Mathematical Statistics* (1951), pp. 400–407.

[175] J. M. Robins, L. Li, E. Tchetgen, and A. van der Vaart. “Higher order influence functions and minimax estimation of nonlinear functionals”. In: *Probability and statistics: essays in honor of David A. Freedman* 2 (2008), pp. 335–421.

[176] J. M. Robins and Y. Ritov. “Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models”. In: *Statistics in medicine* 16.3 (1997), pp. 285–319.

[177] J. M. Robins and A. Rotnitzky. “Semiparametric efficiency in multivariate regression models with missing data”. In: *Journal of the American Statistical Association* 90.429 (1995), pp. 122–129.

[178] J. M. Robins, A. Rotnitzky, and L. P. Zhao. “Analysis of semiparametric regression models for repeated outcomes in the presence of missing data”. In: *Journal of the American Statistical Association* 90.429 (1995), pp. 106–121.

[179] J. M. Robins, E. T. Tchetgen, L. Li, and A. van der Vaart. “Semiparametric minimax rates”. In: *Electronic journal of statistics* 3 (2009), p. 1305.

[180] J. P. Romano and M. Wolf. “Resurrecting weighted least squares”. In: *Journal of Econometrics* 197.1 (2017), pp. 1–19.

[181] P. R. Rosenbaum. “Model-based direct adjustment”. In: *Journal of the American Statistical Association* 82.398 (1987), pp. 387–394.

[182] P. R. Rosenbaum and D. B. Rubin. “The central role of the propensity score in observational studies for causal effects”. In: *Biometrika* 70.1 (1983), pp. 41–55.

[183] N. Ross. “Fundamentals of Stein’s method”. In: *Probability Surveys* 8 (2011), pp. 210–293.

[184] D. B. Rubin and N. Thomas. “Characterizing the effect of matching using linear propensity score methods with normal distributions”. In: *Biometrika* 79.4 (1992), pp. 797–809.

[185] G. A. Rummery and M. Niranjan. *On-line Q-learning using connectionist systems.* Tech. rep. Cambridge University Engineering Department, 1994.

[186] D. Ruppert. *Efficient estimations from a slowly convergent Robbins-Monro process.* Tech. rep. Cornell University Operations Research and Industrial Engineering, 1988.

[187] B. Scherrer. “Should one compute the temporal difference fix point or minimize the Bellman residual? The unified oblique projection view”. In: *International Conference on Machine Learning.* 2010, pp. 959–966.

[188] Y. Shen, C. Gao, D. Witten, and F. Han. “Optimal estimation of variance in nonparametric regression with random design”. In: *The Annals of Statistics* 48.6 (2020), pp. 3589–3618.

[189] J. Shin, A. Ramdas, and A. Rinaldo. “On conditional versus marginal bias in multi-armed bandits”. In: *International Conference on Machine Learning.* PMLR. 2020, pp. 8852–8861.

[190] A. Sidford, M. Wang, X. Wu, L. F. Yang, and Y. Ye. “Near-optimal time and sample complexities for solving Markov decision processes with a generative model”. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems.* 2018, pp. 5192–5202.

[191] R. Singh. “Kernel methods for unobserved confounding: Negative controls, proxies, and instruments”. In: *arXiv preprint arXiv:2012.10315* (2020).

[192] R. Singh, L. Xu, and A. Gretton. “Kernel Methods for Causal Functions: Dose, Heterogeneous, and Incremental Response Curves”. In: *arXiv preprint arXiv:2010.04855* (2020).

[193] P. Speckman. “Minimax estimates of linear functionals in a Hilbert space”. In: *Unpublished manuscript* (1979).

[194] R. Srikant and L. Ying. “Finite-time error bounds for linear stochastic approximation and TD learning”. In: *Conference on Learning Theory.* PMLR. 2019, pp. 2803–2830.

[195] J. Stander and B. W. Silverman. "Minimax estimation of linear functionals, particularly in nonparametric regression and positron emission tomography". In: *Computational Statistics* 10 (1995), pp. 259–259.

[196] C. Stein. "Efficient nonparametric testing and estimation". In: *Proceedings of the third Berkeley symposium on mathematical statistics and probability.* Vol. 1. 1956, pp. 187–195.

[197] F. Su, W. Mou, P. Ding, and M. J. Wainwright. "When is it better to estimate the propensity score? High-dimensional analysis and bias correction". In: *arXiv preprint arXiv:2303.17102* (2023).

[198] R. S. Sutton. "Learning to predict by the methods of temporal differences". In: *Machine Learning* 3.1 (1988), pp. 9–44.

[199] R. S. Sutton, H. R. Maei, D. Precup, S. Bhatnagar, D. Silver, Cs. Szepesvári, and E. Wiewiora. "Fast gradient-descent methods for temporal-difference learning with linear function approximation". In: *International Conference on Machine Learning.* 2009, pp. 993–1000.

[200] Cs. Szepesvári. *Algorithms for Reinforcement Learning.* Morgan & Claypool Publishers, 2010.

[201] Cs. Szepesvári. "The asymptotic convergence-rate of Q-learning". In: *Advances in Neural Information Processing Systems* (1998), pp. 1064–1070.

[202] M. Talagrand. *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes.* Springer Science and Business Media, 2006.

[203] J. N. Tsitsiklis. "Asynchronous stochastic approximation and Q-learning". In: *Machine Learning* 16 (1994), pp. 185–202.

[204] J. N. Tsitsiklis and B. Van Roy. "Analysis of temporal-diffference learning with function approximation". In: *Advances in Neural Information Processing Systems.* 1997, pp. 1075–1081.

[205] J. N. Tsitsiklis and B. Van Roy. "Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives". In: *IEEE Transactions on Automatic Control* 44.10 (1999), pp. 1840–1851.

[206] A. B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer Science & Business Media, 2008.

[207] A. B. Tsybakov. "Optimal aggregation of classifiers in statistical learning". In: *The Annals of Statistics* 32.1 (2004), pp. 135–166.

[208] A. W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes.* Springer-Verlag New York, 1996.

[209] A. W. van der Vaart. *Asymptotic Statistics.* Vol. 3. Cambridge university press, 2000.

[210] B. Van Roy. “Performance loss bounds for approximate value iteration with state aggregation”. In: *Mathematics of Operations Research* 31.2 (2006), pp. 234–244.

[211] M. J Wainwright. “Stochastic approximation with cone-contractive operators: Sharper ℓ_∞-bounds for Q-learning”. In: *arXiv preprint arXiv:1905.06265* (2019).

[212] M. J Wainwright. “Variance-reduced Q-learning is minimax optimal”. In: *arXiv preprint arXiv:1906.04697* (2019).

[213] Martin J Wainwright. *High-dimensional Statistics: A Non-asymptotic Viewpoint*. Vol. 48. Cambridge University Press, 2019.

[214] Y. Wang and R. D. Shah. “Debiased Inverse Propensity Score Weighting for Estimation of Average Treatment Effects with High-Dimensional Confounders”. In: *arXiv preprint arXiv:2011.08661* (2020).

[215] Y.-X. Wang, A. Agarwal, and M. Dudık. “Optimal and adaptive off-policy evaluation in contextual bandits”. In: *International Conference on Machine Learning*. PMLR. 2017, pp. 3589–3597.

[216] C. J. C. H. Watkins and P. Dayan. “Q-learning”. In: *Machine Learning* 8.3-4 (1992), pp. 279–292.

[217] G. N. Watson. *A Treatise on the Theory of Bessel Functions*. Vol. 3. The University Press, 1922.

[218] H. Widom. “Asymptotic behavior of the eigenvalues of certain integral equations”. In: *Transactions of the American Mathematical Society* 109.2 (1963), pp. 278–295.

[219] R. C. Williamson, A. J. Smola, and B. Scholkopf. “Generalization performance of regularization networks and support vector machines via entropy numbers of compact operators”. In: *IEEE transactions on Information Theory* 47.6 (2001), pp. 2516–2532.

[220] G. Wolfer and A. Kontorovich. “Statistical estimation of ergodic Markov chain kernel over discrete state space”. In: *Bernoulli* 27.1 (2021), pp. 532–553.

[221] J. M. Wooldridge. *Introductory Econometrics: a Modern Approach*. Nelson Education, 2016.

[222] T. Xie and N. Jiang. “Batch value-function approximation with only realizability”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 11404–11413.

[223] F. Yang, S. Balakrishnan, and M. J. Wainwright. “Statistical and computational guarantees for the Baum–Welch algorithm”. In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 4528–4580.

[224] Y. G. Yatracos. “Rates of convergence of minimum distance estimators and Kolmogorov’s entropy”. In: *The Annals of Statistics* (1985), pp. 768–774.

[225] M. Yin and Y.-X. Wang. “Asymptotically efficient off-policy evaluation for tabular reinforcement learning”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 3948–3958.

[226] H. Yu and D. P. Bertsekas. “Error bounds for approximations from projected linear equations”. In: *Mathematics of Operations Research* 35.2 (2010), pp. 306–329.

[227] A. Zanette, M. J. Wainwright, and E. Brunskill. “Provable benefits of actor-critic methods in offline reinforcement learning”. In: *Neural Information Processing Systems*. Dec. 2021.

[228] R. Zhan, V. Hadad, D. A. Hirshberg, and S. Athey. “Off-policy evaluation via adaptive weighting with data from contextual bandits”. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021, pp. 2125–2135.

[229] R. Zhan, Z. Ren, S. Athey, and Z. Zhou. “Policy learning with adaptively collected data”. In: *arXiv preprint arXiv:2105.02344* (2021).

[230] T. Zhang. “Effective dimension and generalization of kernel learning”. In: *Advances in Neural Information Processing Systems* 15 (2002).

[231] D.-X. Zhou. “The covering number in learning theory”. In: *Journal of Complexity* 18.3 (2002), pp. 739–767.

[232] Z. Zhou, S. Athey, and S. Wager. “Offline multi-action policy learning: Generalization and optimization”. In: *Operations Research* (2022).

[233] B. Zhu, J. Jiao, and D. Tse. “Deconstructing generative adversarial networks”. In: *IEEE Transactions on Information Theory* (2020).

Appendix A

Proofs and discussion deferred from Chapter 2

A.1 Proofs deferred from Section 2.5.1

In the following subsections, we prove Lemmas 2.5 and 2.6, the two technical lemmas used in the proof of the bound (2.45b).

A.1.1 Proof of Lemma 2.5

Recall that Lemma 2.5 provides a bound on the error of the non-averaged iterates $\{v_t\}_{t \geq 1}$, as defined in equation (2.24a). Using the form of the update, we expand the mean-squared error to find that

$$\begin{aligned}
\mathbb{E}\|v_{t+1} - \bar{v}\|^2 &= \mathbb{E}\|(I - \eta I + \eta \Pi_{\mathbb{S}} L)(v_t - \bar{v}) + \eta \Pi_{\mathbb{S}}(L_{t+1} - L)v_t + \eta \Pi_{\mathbb{S}}(b_{t+1} - b)\|^2 \\
&\overset{(i)}{=} \mathbb{E}\|(I - \eta I + \eta \Pi_{\mathbb{S}} L)(v_t - \bar{v})\|^2 + \eta^2 \mathbb{E}\|\Pi_{\mathbb{S}}(L_{t+1} - L)v_t + \Pi_{\phi}(b_{t+1} - b)\|^2 \\
&\overset{(ii)}{\leq} (1 - \eta(1 - \kappa))\mathbb{E}\|v_t - \bar{v}\|^2 + 2\eta^2 \mathbb{E}\|\Pi_{\mathbb{S}}(L_{t+1} - L)(v_t - \bar{v})\|^2 \\
&\qquad + 2\eta^2 \mathbb{E}\|\Pi_{\mathbb{S}}(L_{t+1} - L)\bar{v} + \Pi_{\mathbb{S}}(b_{t+1} - b)\|^2.
\end{aligned} \tag{A.1}$$

In step (i), we have made use of the fact that the noise is unbiased, and in step (ii), we have used the fact that for any Δ in the subspace $\mathbb{S}$ and any stepsize $\eta \in \big(0, \frac{1-\kappa}{1+|||L|||_{\mathbb{X}}^2}\big)$, we have

$$\begin{aligned}
\|(I - \eta I + \eta \Pi_{\mathbb{S}} L)\Delta\|^2 &= (1 - \eta)^2 \|\Delta\|^2 + \eta^2 \|\Pi_{\mathbb{S}} L \Delta\|^2 + 2(1 - \eta)\eta \langle \Delta, \, \Pi_{\mathbb{S}} L \Delta \rangle \\
&\leq \Big\{1 - 2\eta + \eta^2 + \eta^2 |||L|||_{\mathbb{X}}^2 + 2(1 - \eta)\eta\kappa\Big\} \|\Delta\|^2 \\
&\leq \big(1 - \eta(1 - \kappa)\big)\|\Delta\|^2.
\end{aligned}$$

Turning to the second term of equation (A.1), the moment bounds in Assumption 2.1(W) imply that

$$\mathbb{E}\|\Pi_{\mathbb{S}}(L_{t+1} - L)(v_t - \bar{v})\|^2 = \sum_{j=1}^{d} \mathbb{E}\langle \phi_j,\, (L_{t+1} - L)(v_t - \bar{v})\rangle^2 \leq \mathbb{E}\|v_t - \bar{v}\|^2 \sigma_L^2 d.$$

Finally, the last term of equation (A.1) is also handled by Assumption 2.1(W), whence we obtain

$$\begin{aligned}\mathbb{E}\|\Pi_{\mathbb{S}}(L_{t+1} - L)\bar{v} + \Pi_{\mathbb{S}}(b_{t+1} - b)\|^2 &\leq 2\sum_{j=1}^{d} \mathbb{E}\langle \phi_j,\, (L_{t+1} - L)\bar{v}\rangle^2 \\ &\quad + 2\sum_{j=1}^{d} \mathbb{E}\langle \phi_j,\, b_{t+1} - b\rangle^2 \leq 2\|\bar{v}\|^2\sigma_L^2 d + 2\sigma_b^2 d.\end{aligned}$$

Putting together the pieces, we see that for any stepsize $\eta \in \big(0, \frac{1-\kappa}{4\sigma_L^2 d + 1 + \|L\|_{\mathbb{X}}^2}\big)$, we have

$$\begin{aligned}\mathbb{E}\|v_{t+1} - \bar{v}\|^2 &\leq (1 - \eta(1-\kappa) + 2\eta^2\sigma_L^2 d)\mathbb{E}\|v_t - \bar{v}\|^2 + 4\eta^2(\|\bar{v}\|^2\sigma_L^2 d + \sigma_b^2 d) \\ &\leq \Big(1 - \frac{\eta(1-\kappa)}{2}\Big)\mathbb{E}\|v_t - \bar{v}\|^2 + 4\eta^2(\|\bar{v}\|^2\sigma_L^2 d + \sigma_b^2 d).\end{aligned}$$

Finally, rolling out the recursion yields the bound

$$\mathbb{E}\|v_n - \bar{v}\|^2 \leq e^{-(1-\kappa)\eta n/2}\mathbb{E}\|v_0 - \bar{v}\|^2 + \frac{8\eta}{1-\kappa}(\|\bar{v}\|^2\sigma_L^2 d + \sigma_b^2 d),$$

which completes the proof.

A.1.2 Proof of Lemma 2.6

Recall that $\bar{v}$ satisfies the fixed point equation $\bar{v} = \Pi_{\mathbb{S}} L\bar{v} + \Pi_{\mathbb{S}} b$. Using this fact, we can derive the following elementary identity:

$$\begin{aligned}\frac{v_{n_0} - v_n}{\eta(n - n_0)} &= \frac{1}{n - n_0}\sum_{t=n_0}^{n-1}\Big(v_t - \Pi_{\mathbb{S}} L_{t+1} v_t - \Pi_{\mathbb{S}} b_{t+1}\Big) \\ &= (I - \Pi_{\mathbb{S}} L)(\widehat{v}_n - \bar{v}) + \underbrace{\frac{1}{n - n_0}\sum_{t=n_0}^{n-1}\Pi_{\mathbb{S}}(L_{t+1} - L)(v_t - \bar{v})}_{=:\Psi_n^{(1)}} \\ &\quad + \underbrace{\frac{1}{n - n_0}\sum_{t=n_0}^{n-1}\Pi_{\mathbb{S}}\big((L_{t+1} - L)\bar{v} + b_{t+1} - b\big)}_{=:\Psi_n^{(2)}}. \qquad \text{(A.2)}\end{aligned}$$

Re-arranging terms and applying the Cauchy–Schwarz inequality, we have

$$\|\widehat{v}_n - \bar{v}\|^2 \leq \frac{3}{(n-n_0)^2\eta^2}\|(I-\Pi_{\mathbb{S}}L)^{-1}(v_n - v_{n_0})\|^2 \\ + \frac{3}{(n-n_0)^2}\Big(\|(I-\Pi_{\mathbb{S}}L)^{-1}\Psi_n^{(1)}\|^2 + \|(I-\Pi_{\mathbb{S}}L)^{-1}\Psi_n^{(2)}\|^2\Big).$$

Note that the quantities $\Psi_n^{(1)}$ and $\Psi_n^{(2)}$ are martingales adapted to the filtration $\mathcal{F}_n := \sigma(\{L_i, b_i\}_{i=1}^n)$, so that

$$\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \leq \frac{3}{(n-n_0)^2}\sum_{t=n_0}^{n-1}\mathbb{E}\|(I-\Pi_{\mathbb{S}}L)^{-1}\Pi_{\mathbb{S}}(L_{t+1}-L)(v_t - \bar{v})\|^2 \\ + \frac{3}{(n-n_0)^2}\sum_{t=n_0}^{n-1}\mathbb{E}\|(I-\Pi_{\mathbb{S}}L)^{-1}\Pi_{\mathbb{S}}\big((L_{t+1}-L)\bar{v} + b_{t+1} - b\big)\|^2 \\ + \frac{3}{(n-n_0)^2\eta^2}\mathbb{E}\|(I-\Pi_{\mathbb{S}}L)^{-1}(v_n - v_{n_0})\|^2.$$

We claim that for any vector $v \in \mathbb{X}$, we have

$$(I-\Pi_{\mathbb{S}}L)^{-1}\Pi_{\mathbb{S}}v = \Phi_d^*\Big((I-M)^{-1}\Phi_d v\Big). \tag{A.3}$$

Taking this claim as given for the moment, by applying equation (A.3) with $v = (L_{t+1}-L)(v_t - \bar{v})$ and $v = (L_{t+1}-L)\bar{v} + b_{t+1} - b$, respectively, we find that

$$\mathbb{E}\|(I-\Pi_{\mathbb{S}}L)^{-1}\Pi_{\mathbb{S}}(L_{t+1}-L)(v_t-\bar{v})\|^2 = \mathbb{E}\|(I-M)^{-1}\Phi_d(L_{t+1}-L)(v_t-\bar{v})\|_2^2,$$

and

$$\mathbb{E}\|(I-L)^{-1}\Pi_{\mathbb{S}}\big((L_{t+1}-L)\bar{v}+b_{t+1}-b\big)\|^2 = \mathbb{E}\|(I-M)^{-1}\Phi_d\big((L_{t+1}-L)\bar{v}+b_{t+1}-b\big)\|_2^2.$$

Putting together the pieces, we obtain

$$\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \leq \frac{3}{n-n_0}\,\text{trace}\,\Big((I-M)^{-1}\cdot\text{cov}\,\big(\Phi_d(b_1-b)+\Phi_d(L_1-L)\bar{v}\big)\cdot(I-M)^{-\top}\Big) \\ + \frac{3}{(n-n_0)^2}\sum_{t=n_0}^{n}\mathbb{E}\|(I-M)^{-1}\Phi_d(L_{t+1}-L)(v_t-\bar{v})\|_2^2 + \frac{3\mathbb{E}\|v_n - v_{n_0}\|^2}{\eta^2(n-n_0)^2(1-\kappa)^2},$$

as claimed.

It remains to prove the identity (A.3).

Proof of claim (A.3): Note that for any vector $v \in \mathbb{X}$, the vector $z := (I-\Pi_{\mathbb{S}}L)^{-1}\Pi_{\mathbb{S}}v$ is a member of $\mathbb{S}$, since $z = \Pi_{\mathbb{S}}Lz + \Pi_{\mathbb{S}}v$. Furthermore, since $\{\phi_j\}_{j=1}^d$ is a standard basis for $\mathbb{S}$, we have $z = \Pi_{\mathbb{S}}z = \Phi_d^*\Phi_d z$, and consequently,

$$\Phi_d z = \Phi_d L z + \Phi_d v = (\Phi_d L \Phi_d^*)\Phi_d z + \Phi_d v = M\Phi_d z + \Phi_d v.$$

Since the matrix M is invertible, we have $\Phi_d z = (I_d - M)^{-1}\Phi_d v$. Consequently, we have the identity $z = \Phi_d^*\Phi_d z = \Phi_d^*(I_d - M)^{-1}\Phi_d v$, which proves the claim.

A.2 Proofs deferred from Section 2.5.2

In the following subsections, we prove Lemma 2.7 and 2.8, the two technical lemmas used in the proof of Theorem 2.2.

A.2.1 Proof of Lemma 2.7

We prove the two parts of the lemma separately. Once again, recall our definition of the pair (w, y) from equation (2.55).

Proof of part (a): In order to study the operator norm of the matrix $L^{(\varepsilon,z)}$ in the Hilbert space $\mathbb{X}$, we consider a vector $p = \begin{bmatrix} p^{(1)} \\ p^{(2)} \end{bmatrix} \in \mathbb{R}^D$, with $p^{(1)} \in \mathbb{R}^d$ and $p^{(2)} \in \mathbb{R}^{D-d}$. Assuming $\|p\| = 1$, we have

$$\|L^{(\varepsilon,z)}p\|^2 = \frac{1}{2d}\|M_0 p^{(1)} + w \cdot \frac{\sqrt{d}}{D-d}\sum_{j=d+1}^{D} \varepsilon_j p_j^{(2)}\|_2^2.$$

By the Cauchy–Schwarz inequality, we have

$$\left|\frac{\sqrt{d}}{D-d}\sum_{j=d+1}^{D}\varepsilon_j p_j^{(2)}\right|^2 \leq \frac{d}{(D-d)^2}\Big(\sum_{j=d+1}^{D}\varepsilon_j^2\Big)\Big(\sum_{j=d+1}^{D}(p_j^{(2)})^2\Big) = \frac{d}{D-d}\|p^{(2)}\|_2^2.$$

Define the vector $a_1 := \frac{1}{\sqrt{2d}}p^{(1)} \in \mathbb{R}^d$ and $a_2 := \frac{1}{\sqrt{2(D-d)}}\|p^{(2)}\|_2$. Clearly, we have $1 = \|p\|^2 = \|a_1\|_2^2 + a_2^2$, and so

$$\begin{aligned}
\|L^{(\varepsilon,z)}p\|^2 &\leq \frac{1}{2d}\cdot \sup_{t\in[-1,1]} \|M_0 p^{(1)} + \sqrt{2d}a_2 t w\|_2^2 \\
&= \frac{1}{2d}\cdot \max\Big(\|M_0 p^{(1)} + \sqrt{2d}a_2 w\|_2^2, \|M_0 p^{(1)} - \sqrt{2d}a_2 w\|_2^2\Big) \\
&= \max\left(\|M_0 a_1 + a_2 w\|_2^2, \|M_0 a_1 - a_2 w\|_2^2\right) \\
&\leq \| \begin{bmatrix} M_0 & w \end{bmatrix} \|_{\mathrm{op}}^2.
\end{aligned}$$

Equation (2.56) implies that $\| \begin{bmatrix} M_0 & w \end{bmatrix} \|_{\mathrm{op}}^2 = \lambda_{\max}\Big(M_0 M_0^\top + ww^\top\Big) \leq \gamma_{\max}^2$, and therefore, for all $\varepsilon \in \{-1,1\}^{D-d}$ and $z \in \{-1,1\}$, we have

$$\|L^{(\varepsilon,z)}\|_{\mathbb{X}} = \sup_{\|p\|=1}\|L^{(\varepsilon,z)}p\| \leq \gamma_{\max},$$

as desired.

Proof of part (b): Consider any pair of vectors $p, q \in \mathbb{R}^D$ such that $\|p\| = \|q\| = 1$. Using the decompositions $p = \begin{bmatrix} p^{(1)} \\ p^{(2)} \end{bmatrix}$ and $q = \begin{bmatrix} q^{(1)} \\ q^{(2)} \end{bmatrix}$, with $p^{(1)}, q^{(1)} \in \mathbb{R}^d$, $p^{(2)}, q^{(2)} \in \mathbb{R}^{D-d}$, we have

$$\begin{aligned}
\mathbb{E}\langle p,\, (L_i^{(\varepsilon,z)} - L^{(\varepsilon,z)})q\rangle^2 &\leq \frac{1}{(2d)^2}\mathbb{E}\Big(\sqrt{d}\varepsilon_{\tau_L} q^{(2)}_{\tau_L^{(i)}} w^\top p^{(1)}\Big)^2 \\
&= \frac{1}{4d} \cdot \Big(w^\top p^{(1)}\Big)^2 \cdot \mathbb{E}\Big(q^{(2)}_{\tau_L^{(i)}}\Big)^2 \\
&\leq \frac{1}{4d}\|p^{(1)}\|_2^2 \cdot \|w\|_2^2 \cdot \frac{1}{D-d}\|q^{(2)}\|_2^2 \\
&\leq \|w\|_2^2 \cdot \|p\|^2 \cdot \|q\|^2 \leq \|w\|_2^2.
\end{aligned}$$

Recall that $M_0 M_0^\top + ww^\top \preceq \gamma_{\max}^2 I_d$ by equation (2.56). Consequently, we have $\|w\|_2^4 \leq \|M_0^\top w\|_2^2 + \|w\|_2^4 \leq \gamma_{\max}^2\|w\|_2^2$, which implies that $\|w\|_2 \leq \gamma_{\max}$. Therefore, the noise assumption in equation (2.12a) is satisfied with parameter $\sigma_L = \gamma_{\max}$.

For the noise on the vector b, we note that

$$\begin{aligned}
\mathbb{E}\langle b_i^{(\varepsilon,z)} - b,\, p\rangle^2 \;&\leq\; \frac{1}{4(D-d)^2}\mathbb{E}\Big(\sqrt{2}(D-d)z\delta\varepsilon_{\tau_b^{(i)}} p^{(2)}_{\tau_b^{(i)}}\Big)^2 \\
&\leq \frac{\delta^2}{2}\mathbb{E}\Big(p^{(2)}_{\tau_b^{(i)}}\Big)^2 \leq \frac{\delta^2}{2} \cdot \frac{1}{D-d}\|p^{(2)}\|_2^2 \leq \delta^2\|p\|^2 = \delta^2,
\end{aligned}$$

showing the the the noise assumption (2.12b) is satisfied with $\sigma_b = \delta$.

A.2.2 Proof of Lemma 2.8

Recall that $\tau_L^{(i)}, \tau_b^{(i)}$ are the random indices in the i-th sample. We define $\mathscr{E}$ to be the event that the indices $\big(\tau_L^{(i)}\big)_{i=1}^n, \big(\tau_b^{(i)}\big)_{i=1}^n$ are not all distinct, i.e.,

$$\begin{aligned}
\mathscr{E} := \Big\{\exists i_1, i_2 \in [n],\ \text{s.t. } \tau_L^{(i_1)} = \tau_b^{(i_2)}\Big\} \cup \Big\{\exists i_1 \neq i_2,\ \text{s.t. } \tau_L^{(i_1)} = \tau_L^{(i_2)}\Big\} \\
\cup \Big\{\exists i_1 \neq i_2,\ \text{s.t. } \tau_b^{(i_1)} = \tau_b^{(i_2)}\Big\}.
\end{aligned}$$

We claim that

$$\mathbb{P}_1^{(n)}|\mathscr{E}^C = \mathbb{P}_{-1}^{(n)}|\mathscr{E}^C. \tag{A.4}$$

Assuming equation (A.4), we now give a proof of the upper bound on the total variation distance δ. We use the following lemma:

Lemma A.1. *Given two probability measures $\mathbb{P}_1, \mathbb{P}_2$ and an event $\mathscr{E}$ with $\mathbb{P}_1(\mathscr{E}), \mathbb{P}_2(\mathscr{E}) < 1/2$, we have*

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) \leq d_{\mathrm{TV}}(\mathbb{P}_1|\mathscr{E}^C, \mathbb{P}_2|\mathscr{E}^C) + 3\mathbb{P}_1(\mathscr{E}) + 3\mathbb{P}_2(\mathscr{E}).$$

In order to bound the probability of $\mathscr{E}$, we apply a union bound. Under either of the probability measures $\mathbb{P}_1^{(n)}$ and $\mathbb{P}_{-1}^{(n)}$, we have the following bound:

$$\mathbb{P}(\mathscr{E}) \leq \sum_{i,j\in[n]} \mathbb{P}\Big(\tau_L^{(i)} = \tau_b^{(j)}\Big) + \sum_{i_1<i_2} \mathbb{P}\Big(\tau_L^{(i_1)} = \tau_L^{(i_2)}\Big) + \sum_{i_1<i_2} \mathbb{P}\Big(\tau_b^{(i_1)} = \tau_b^{(i_2)}\Big) \leq \frac{2n^2}{D-d}.$$

Applying Lemma A.1 in conjunction with equation (A.4) yields

$$d_{\mathrm{TV}}(\mathbb{P}_1^{(n)}, \mathbb{P}_{-1}^{(n)}) \leq \frac{12n^2}{D-d}. \tag{A.5}$$

It remains to prove claim (A.4) and Lemma A.1.

Proof of equation (A.4): For $D \geq 2n+d$, we define a probability measure $\mathbb{Q}$ through the following sampling procedure:

- Sample a subset $S \subseteq \{d+1, \cdots, D\}$ of size $2n$ uniformly at random over all possible $\binom{D-d}{2n}$ possible subsets.
- Partition the set S into two disjoint subsets $S = S_L \cup S_b$, each of size n. The partition is chosen uniformly at random over all $\binom{2n}{n}$ possible partitions. Let
$$S_L := \Big\{\widetilde{\tau}_L^{(1)}, \widetilde{\tau}_L^{(2)}, \cdots, \widetilde{\tau}_L^{(n)}\Big\} \quad \text{and } S_b := \Big\{\widetilde{\tau}_b^{(1)}, \widetilde{\tau}_b^{(2)}, \cdots, \widetilde{\tau}_b^{(n)}\Big\}.$$
- For each $i \in [n]$, sample two random bits $\zeta_L^{(i)}, \zeta_b^{(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(\{-1, 1\})$.
- Let $\mathbb{Q}$ be the probability distribution of the observations $(L_i, b_i)_{i=1}^n$, that are constructed from the tuple $(\widetilde{\tau}_L^{(i)}, \widetilde{\tau}_b^{(i)}, \zeta_L^{(i)}, \zeta_b^{(i)})$ defined above. Specifically, we let
$$L_i := \begin{bmatrix} M_0 & 0 & \cdots & 0 & \sqrt{d}\zeta_L^{(i)} w & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & & & \cdots & & & 0 \\ & & \vdots & & & & \vdots & \\ 0 & 0 & & & \cdots & & & 0 \end{bmatrix}, \quad b_i := \begin{bmatrix} \sqrt{2d}h_0 \\ 0 \\ \vdots \\ 0 \\ \sqrt{2}(D-d)\zeta_b^{(i)}\delta \\ 0 \\ \cdots \\ 0 \end{bmatrix},$$
where the vector $\sqrt{d}\zeta_L^{(i)} w$ appears at the $\widetilde{\tau}_L^{(i)}$-th column of the matrix L_i, and the scalar $\sqrt{2}(D-d)\zeta_b^{(i)}\delta$ appears at the $\widetilde{\tau}_b^{(i)}$-th row of the vector b_i.

For either choice of the bit $z \in \{\pm 1\}$, we claim that the probability measure $\mathbb{P}_z^{(n)}|\mathscr{E}^C$ is identical to the distribution $\mathbb{Q}$. To prove this claim, we first note that conditioned on the event $\mathscr{E}^C$, the indices $(\tau_L^{(i)}, \tau_b^{(i)})_{i=1}^n$ actually form a uniform random subset of $\{d+1, \cdots, D\}$ with cardinality $2n$, and the partition into $(\tau_L^{(i)})_{i=1}^n$ and $(\tau_b^{(i)})_{i=1}^n$ is a uniform random partition, i.e.,

$$\left(\tau_L^{(i)}, \tau_b^{(i)}\right)_{i=1}^n \Big| \mathscr{E}^C \overset{d}{=} \left(\widetilde{\tau}_L^{(i)}, \widetilde{\tau}_b^{(i)}\right)_{i=1}^n, \quad \text{under both } \mathbb{P}_1^{(n)} \text{ and } \mathbb{P}_{-1}^{(n)}. \tag{A.6}$$

Given an index subset $\left(t_L^{(i)}, t_b^{(i)}\right)_{i=1}^n \subseteq \{d+1, \cdots, D\}$ that are mutually distinct, conditioned on the value of $\left(\tau_L^{(i)}, \tau_b^{(i)}\right)_{i=1}^n = \left(t_L^{(i)}, t_b^{(i)}\right)_{i=1}^n$, the observed random bits under the probability distribution $\mathbb{P}_z^{(n)}$ are given by

$$\varepsilon_{\tau_L^{(1)}}, \varepsilon_{\tau_L^{(2)}}, \cdots, \varepsilon_{\tau_L^{(n)}}, z\varepsilon_{\tau_b^{(1)}}, \cdots, z\varepsilon_{\tau_b^{(n)}},$$

which are $2n$ independent Rademacher random variables.

On the other hand, the random bits $\zeta_L^{(1)}, \zeta_L^{(2)}, \cdots, \zeta_L^{(n)}, \zeta_b^{(1)}, \zeta_b^{(2)}, \cdots, \zeta_b^{(n)}$ are also $2n$ independent Rademacher random variables. Consequently, for any index subset $\left(t_L^{(i)}, t_b^{(i)}\right)_{i=1}^n \subseteq \{d+1, \cdots, D\}$ that are mutually distinct, we have the following equality-in-distribution:

$$\left(\varepsilon_{\tau_L^{(i)}}, z\varepsilon_{\tau_b^{(i)}}\right)_{i=1}^n \Big| (\tau_L^{(i)} = t_L^{(i)}, \tau_b^{(i)} = t_b^{(i)})_{i=1}^n \overset{d}{=} \left(\zeta_L^{(i)}, \zeta_b^{(i)}\right)_{i=1}^n \Big| (\widetilde{\tau}_L^{(i)} = t_L^{(i)}, \widetilde{\tau}_b^{(i)} = t_b^{(i)})_{i=1}^n. \tag{A.7}$$

Putting equations (A.6) and (A.7) together completes the proof.

Proof of Lemma A.1: Given a function f with range contained in $[0, 1]$, we have

$$\begin{aligned}
&\left| \int f(x)\mathbb{P}_1(dx) - \int f(x)\mathbb{P}_2(dx) \right| \\
&\leq \left| \int_{\mathscr{E}} f(x)\mathbb{P}_1(dx) \right| + \left| \int_{\mathscr{E}} f(x)\mathbb{P}_2(dx) \right| + \left| \int_{\mathscr{E}^C} f(x)\mathbb{P}_1(dx) - \int_{\mathscr{E}^C} f(x)\mathbb{P}_2(dx) \right| \\
&\leq \mathbb{P}_1(\mathscr{E}) + \mathbb{P}_2(\mathscr{E}) + \mathbb{P}_1(\mathscr{E}^C) \cdot \left| \frac{\int_{\mathscr{E}^C} f(x)\mathbb{P}_1(dx)}{\mathbb{P}_1(\mathscr{E}^C)} - \frac{\int_{\mathscr{E}^C} f(x)\mathbb{P}_2(dx)}{\mathbb{P}_2(\mathscr{E}^C)} \right| + \tfrac{|\mathbb{P}_1(\mathscr{E}^C) - \mathbb{P}_2(\mathscr{E}^C)|}{\mathbb{P}_2(\mathscr{E}^C)} \\
&\leq \mathbb{P}_1(\mathscr{E}) + \mathbb{P}_2(\mathscr{E}) + d_{\mathrm{TV}}(\mathbb{P}_1|\mathscr{E}^C, \mathbb{P}_2|\mathscr{E}^C) + 2|\mathbb{P}_1(\mathscr{E}) - \mathbb{P}_2(\mathscr{E})| \\
&\leq 3(\mathbb{P}_1(\mathscr{E}) + \mathbb{P}_2(\mathscr{E})) + d_{\mathrm{TV}}(\mathbb{P}_1|\mathscr{E}^C, \mathbb{P}_2|\mathscr{E}^C),
\end{aligned}$$

which completes the proof.

A.3 Proof of Corollary 2.1

We begin by applying Theorem 2.1 with $\omega = 1$. Applying Lemmas 2.1 and 2.2 yield the desired bounds on the approximation error in parts (a) and (b). We also claim that

$$\mathcal{E}_n(M, \Sigma^*) \leq \frac{(\sigma_L^2 \|\bar{v}\|^2 + \sigma_b^2)d}{(1-\kappa)^2 n}, \tag{A.8a}$$

and that given a sample size such that $n \geq \frac{c\sigma_L^2 d}{(1-\kappa)^2} \log^2\left(\frac{\|v_0 - \bar{v}\|^2 d}{1-\kappa}\right) > \frac{c\sigma_L^2 d}{(1-\kappa)^2}$, we have

$$\mathcal{H}_n(\sigma_L, \sigma_b, \bar{v}) \leq \frac{\sigma_L}{1-\kappa}\sqrt{\frac{d}{n}} \cdot \frac{(\sigma_L^2\|\bar{v}\|^2 + \sigma_b^2)d}{(1-\kappa)^2 n} \leq \frac{(\sigma_L^2\|\bar{v}\|^2 + \sigma_b^2)d}{(1-\kappa)^2 n}, \tag{A.8b}$$

Combining these two auxiliary claims establishes the corollary. It remains to establish the bounds (A.8).

Proof of claim (A.8): Let us first handle the contribution to this error from the noise variables b_i. We begin with the following sequence of bounds:

$$\begin{aligned}
&\text{trace}\Big((I-M)^{-1}\operatorname{cov}(\Phi_d(b_1 - b))(I-M)^{-\top}\Big) \\
&\qquad = \text{trace}\Big((I-M)^{-\top}(I-M)^{-1} \cdot \operatorname{cov}(\Phi_d(b_1 - b))\Big) \\
&\qquad \leq |\!|\!|(I-M)^{-\top}(I-M)^{-1}|\!|\!|_{\text{op}} \cdot |\!|\!|\operatorname{cov}(\Phi_d(b_1 - b))|\!|\!|_{nuc} \\
&\qquad \leq |\!|\!|(I-M)^{-1}|\!|\!|_{\text{op}}^2 \,\text{trace}\big(\operatorname{cov}(\Phi_d(b_1 - b))\big).
\end{aligned}$$

By the assumption $\kappa(M) < 1$, for any vector $u \in \mathbb{R}^d$, we have that

$$(1-\kappa)\|u\|_2^2 \leq \langle (I-M)u,\, u\rangle \leq \|(I-M)u\|_2 \cdot \|u\|_2.$$

Consequently, we have the bound $|\!|\!|(I-M)^{-1}|\!|\!|_{\text{op}} \leq \frac{1}{1-\kappa(M)}$. For the trace of the covariance, we note by Assumption 2.1(W) that

$$\text{trace}\big(\operatorname{cov}(\Phi_d(b_1 - b))\big) = \sum_{j=1}^{d}\langle \phi_j,\, b_1 - b\rangle^2 \leq \sigma_b^2 d.$$

Putting together the pieces yields trace $\big((I-M)^{-1}\operatorname{cov}(\Phi_d(b_1 - b))(I-M)^{-\top}\big) \leq \frac{\sigma_b^2 d}{(1-\kappa)^2}$.

Turning now to the contribution to the error from the random observation L_i, we have

$$\begin{aligned}
&\text{trace}\Big((I-M)^{-1}\operatorname{cov}(\Phi_d(L_1 - L)\bar{v})(I-M)^{-\top}\Big) \\
&\qquad\qquad \leq |\!|\!|(I-M)^{-1}|\!|\!|_{\text{op}}^2 \,\text{trace}\big(\operatorname{cov}(\Phi_d(L_1 - L)\bar{v})\big).
\end{aligned}$$

Once again, Assumption 2.1(W) yields the bound

$$\text{trace}\big(\operatorname{cov}(\Phi_d(L_1 - L)\bar{v})\big) = \sum_{j=1}^{d}\langle \phi_j,\, (L_1 - L)\bar{v}\rangle^2 \leq \sigma_L^2\|\bar{v}\|^2 d,$$

and combining the pieces proves the claim.

Proof of claim (A.8b)**:** The proof of this claim is immediate. Simply note that for $n \geq \frac{c\sigma_L^2 d}{(1-\kappa)^2} \log^2\left(\frac{\|v_0 - \bar{v}\|^2 d}{1-\kappa}\right) > \frac{c\sigma_L^2 d}{(1-\kappa)^2}$, we have

$$\mathcal{H}_n(\sigma_L, \sigma_b, \bar{v}) \leq \frac{\sigma_L}{1-\kappa}\sqrt{\frac{d}{n}} \cdot \frac{(\sigma_L^2 \|\bar{v}\|^2 + \sigma_b^2)d}{(1-\kappa)^2 n} \leq \frac{(\sigma_L^2 \|\bar{v}\|^2 + \sigma_b^2)d}{(1-\kappa)^2 n}.$$

A.4 Proof of Theorem 2.3

In order to prove our local minimax lower bound, we make use of the Bayesian Cramér–Rao bound, also known as the van Trees inequality. In particular, we use a functional version of this inequality. Before stating the result, it is useful to introduce the general setup and basic notation for parametric models. Given a family $\mathcal{P}_\Theta = \big(\mathbb{P}_\eta : \eta \in \Theta\big)$ of probability distributions of sample $X \in \mathbb{X}$, parameterized by $\eta \in \Theta$, where Θ is an open subset of $\mathbb{R}^d$. Assume that each element in this family is absolute continuous with respect to a base measure λ over $\mathbb{X}$, and denote the Radon–Nikodym derivative by $p_\eta := \frac{d\mathbb{P}_\eta}{d\lambda}$. Assuming differentiability and integrability of relevant quantities, for any $\eta \in \Theta$, we define the Fisher information matrix $I_0(\eta)$ for a single sample as

$$I_0(\eta) := n \cdot \mathbb{E}_{X \sim \mathbb{P}_\eta}\left[\nabla_\eta \log p_\eta(X) \nabla_\eta \log p_\eta(X)^\top\right] \in \mathbb{R}^{d \times d}.$$

For i.i.d. samples of size n, the Fisher information is given by $I_n(\eta) := n \cdot I_0(\eta)$.

Now we are ready to state the Bayesian Cramér–Rao lower bound.

Proposition A.1 (Theorem 1 of [68], special case)**.** *Given a prior distribution ρ with bounded support contained within Θ, let $T : \mathrm{supp}(\rho) \mapsto \mathbb{R}^p$ denote a locally smooth functional. Then for any estimator $\widehat{T}$ based on sample X and for any smooth matrix-valued function $C : \mathbb{R}^d \to \mathbb{R}^{p \times d}$, we have*

$$\mathop{\mathbb{E}}_{\eta \sim \rho} \mathop{\mathbb{E}}_{X \sim p_\eta}\left[\|\widehat{T}(X_1^n) - T(\eta)\|_2^2\right] \geq \frac{\left(\int \mathrm{trace}\left(C(\eta)\frac{\partial T}{\partial \eta}(\eta)\right)\rho(\eta)d\eta\right)^2}{\int \mathrm{trace}\left(C(\eta)I(\eta)C(\eta)^\top\right)\rho(\eta)d\eta + \int \|\nabla \cdot C(\eta) + C(\eta) \cdot \nabla \log \rho(\eta)\|_2^2 \rho(\eta)d\eta}.$$

Recall that our lower bound is local, and holds for problem instances (L, b) such that the pair $(\Phi_d L \Phi_d^*, \Phi_d b)$ is within a small ℓ_2 neighborhood of a fixed pair (M_0, h_0). We proceed by constructing a careful prior on such instances that will allow us to apply Proposition A.1; note that it suffices to construct our prior over the $d \times d$ matrix $\Phi_d L \Phi_d^*$ and d-dimensional vector $\Phi_d b$. For the rest of the proof, we work under the orthonormal basis $\{\phi_j\}_{j=1}^d$. We also use the convenient shorthand $\bar{x}_0 := \Phi_d \bar{v}_0$.

As a building block for our construction, we consider the one-dimensional density function $\mu(t) := \cos^2\left(\frac{\pi t}{2}\right) \cdot \mathbf{1}_{t \in [-1,1]}$, borrowed from Section 2.7 of Tsybakov [206]. It can be verified that μ defines a probability measure supported on the interval $[-1, 1]$. We denote by $\mu^{\otimes d}$ the d-fold product measure of μ. Let Z and Z' denote two random vectors drawn i.i.d. from the distribution $\mu^{\otimes d}$.

We use an auxiliary pair of $\mathbb{R}^d$-valued random variables (ψ, λ) given by

$$\psi := \frac{1}{\sqrt{n}} \Sigma_b^{\frac{1}{2}} Z \quad \text{and} \quad \lambda := \frac{1}{\sqrt{n}} \Sigma_L^{\frac{1}{2}} Z'. \tag{A.9}$$

Our choice of this pair is motivated by the fact that the Fisher information matrix of this distribution takes a desirable form. In particular, we have the following lemma.

Lemma A.2. *Let $\rho : \mathbb{R}^{2d} \to \mathbb{R}_+$ denote the density of (ψ, λ) defined in equation (A.9). Then*

$$I(\rho) = n\pi \begin{bmatrix} \Sigma_b^{-1} & 0 \\ 0 & \Sigma_L^{-1} \end{bmatrix}.$$

We now use the pair (ψ, λ) in order to define the ensemble of population-level problem instances

$$M^{(\psi,\lambda)} := M_0 + \|\bar{x}_0\|_2^{-2} \lambda (\bar{x}_0)^\top \quad \text{and} \quad h^{(\psi,\lambda)} := h_0 + \psi. \tag{A.10}$$

In order to define the problem instance in the Hilbert space $\mathbb{X}$, we simply let $L^{(\psi,\lambda)} := \Phi_d^* M^{(\psi,\lambda)} \Phi_d$ and $b^{(\psi,\lambda)} := \Phi_d^* h^{(\psi,\lambda)}$, for a given basis $(\phi_i)_{i=1}^d$ in the space $\mathbb{S}$.

The matrix-vector pair $(L^{(\psi,\lambda)}, b^{(\psi,\lambda)})$ induces the fixed point equation $\bar{x}_{\psi,\lambda} = M^{(\psi,\lambda)} \bar{x}_{\psi,\lambda} + h^{(\psi,\lambda)}$, and its solution is given by

$$\bar{x}_{\psi,\lambda} = (I - M^{(\psi,\lambda)})^{-1} h^{(\psi,\lambda)} = (I - M_0 - \|\bar{x}_0\|_2^{-2} \lambda (\bar{x}_0)^\top)^{-1} (h_0 + \psi).$$

Note that by construction, the Jacobian matrix formed by taking the partial derivative of $\bar{x}_{\psi,\lambda}$ with respect to ψ and λ is given by

$$\nabla_{\psi,\lambda} \bar{x}_{\psi,\lambda} = \begin{bmatrix} (I - M^{(\psi,\lambda)})^{-1} & \|\bar{x}_0\|_2^{-2} (\bar{x}_0)^\top (I - M^{(\psi,\lambda)})^{-1} (h_0 + \psi) \cdot (I - M^{(\psi,\lambda)})^{-1} \end{bmatrix}.$$

Now define the observation model via $L_i^{(\psi,\lambda)} := \Phi_d^* M_i^{(\psi,\lambda)} \Phi_d$ and $b_i^{(\psi,\lambda)} := \Phi_d^* h_i^{(\psi,\lambda)}$, where

$$M_i^{(\psi,\lambda)} := M^{(\psi,\lambda)} + \|\bar{x}_0\|_2^{-2} w_i (\bar{x}_0)^\top \quad \text{and} \quad h_i^{(\psi,\lambda)} := h^{(\psi,\lambda)} + w_i', \tag{A.11}$$

where $w_i \sim N(0, \Sigma_L)$ and $w_i' \sim N(0, \Sigma_b)$ are independent.

The following lemma certifies some basic properties of observation model constructed above.

Lemma A.3. *Consider the ensemble of problem instances defined in equations (A.10) and (A.11). For each pair (ψ, λ) in the support of ρ, each index $j \in [d]$ and each unit vector $u \in \mathbb{S}^{d-1}$, we have*

$$\|M^{(\psi,\lambda)} - M_0\|_F \leq \sigma_L \sqrt{\frac{d}{n}}, \quad \textit{and} \quad \|h^{(\psi,\lambda)} - h_0\|_2 \leq \sigma_b \sqrt{\frac{d}{n}}, \tag{A.12a}$$

$$\operatorname{cov}\Big(\big(M_1^{(\psi,\lambda)} - M^{(\psi,\lambda)}\big) \bar{x}_0 \Big) = \Sigma_L, \quad \textit{and} \quad \operatorname{cov}\Big(h_1^{(\psi,\lambda)} - h^{(\psi,\lambda)} \Big) = \Sigma_b, \tag{A.12b}$$

$$\mathbb{E}\Big(e_j^\top \big(M_1^{(\psi,\lambda)} - M^{(\psi,\lambda)}\big) u \Big)^2 \leq \sigma_L^2, \quad \textit{and} \quad \mathbb{E}\Big(e_j^\top h_1^{(\psi,\lambda)} - h^{(\psi,\lambda)} \Big)^2 \leq \sigma_b^2. \tag{A.12c}$$

Lemma A.3 ensures that our problem instance lies in the desired class In particular, equation (A.12a) guarantees that the population-level problem instance $\big(L^{(\psi,\lambda)}, b^{(\psi,\lambda)}\big)$ lies in the class $\mathbb{C}_{\mathsf{est}}$; on the other hand, equation (A.12b) and (A.12c) guarantees that the probability distribution $\mathbb{P}_{L,b}$ we constructed lies in the class $\mathbf{G}_{\mathsf{cov}}$.

Some calculations yield that the Fisher information matrix for this observation model is given by

$$I_n(\psi, \lambda) = I_0(\psi, \lambda) = n \begin{bmatrix} \Sigma_b^{-1} & 0 \\ 0 & \Sigma_L^{-1} \end{bmatrix},$$

for any $\psi, \lambda \in \mathbb{R}^d$.

We will apply Proposition A.1 shortly, for which we use the following matrix C:

$$C(\psi, \lambda) := \nabla_{\psi,\lambda} \bar{x}_{\psi,\lambda}\big|_{(0,0)} \cdot I_0(\psi, \lambda)^{-1} = \begin{bmatrix} (I - M_0)^{-1}\Sigma_b & (I - M_0)^{-1}\Sigma_L \end{bmatrix}.$$

Note that by construction, the matrix C does not depend on the pair (ψ, λ).

We also claim that if $n \geq 16\sigma_L^2 \|\!|(I - M_0)^{-1}|\!\|_{\mathrm{op}}^2 d$, then the following inequalities hold for our construction:

$$\begin{aligned} T_b &:= \mathbb{E}_\rho\Big[\operatorname{trace}\Big((I - M_0)^{-1}\Sigma_b (I - M^{(\psi,\lambda)})^{-\top}\Big)\Big] \\ &\geq \frac{1}{2}\operatorname{trace}\Big((I - M_0)^{-1}\Sigma_b (I - M_0)^{-\top}\Big) \text{ and} \qquad \text{(A.13a)} \\ T_L &:= \mathbb{E}_\rho\Big[\frac{(\bar{x}_0)^\top \bar{x}_{\psi,\lambda}}{\|\bar{x}_0\|_2^2}\operatorname{trace}\Big((I - M_0)^{-1}\Sigma_L (I - M^{(\psi,\lambda)})^{-\top}\Big)\Big] \\ &\geq \frac{1}{3}\operatorname{trace}\Big((I - M_0)^{-1}\Sigma_L (I - M_0)^{-\top}\Big). \qquad \text{(A.13b)} \end{aligned}$$

Taking these two claims as given for the moment, let us complete the proof of the theorem. First, note that

$$\begin{aligned} &\mathbb{E}_\rho\Big[\operatorname{trace}\Big(C(\psi, \lambda) \cdot \nabla_{\psi,\lambda}\bar{x}_{\psi,\lambda}^\top\Big)\Big] \\ &= \mathbb{E}_\rho\Big[\operatorname{trace}\Big((I - M_0)^{-1}\Sigma_b (I - M^{(\psi,\lambda)})^{-\top}\Big)\Big] \\ &\qquad + \mathbb{E}_\rho\Big[\frac{(\bar{x}_0)^\top \bar{x}_{\psi,\lambda}}{\|\bar{x}_0\|_2^2}\operatorname{trace}\Big((I - M_0)^{-1}\Sigma_L (I - M^{(\psi,\lambda)})^{-\top}\Big)\Big] \\ &\geq \frac{5}{6}\operatorname{trace}\Big((I - M_0)^{-1}\Sigma_b (I - M_0)^{-\top}\Big). \qquad \text{(A.14)} \end{aligned}$$

Second, since $I_n(\cdot, \cdot)$ and $C(\cdot, \cdot)$ are both constant functionals, we have that

$$\mathbb{E}\Big[\operatorname{trace}\Big(C(\psi, \lambda) I_n(\psi, \lambda) C(\psi, \lambda)\Big)\Big] = \operatorname{trace}\Big((I - M_0)^{-1}(\Sigma_L + \Sigma_b)(I - M_0)^{-\top}\Big). \qquad \text{(A.15)}$$

Additionally, Lemma A.2 yields

$$\begin{aligned}
&\mathbb{E}\|\nabla \cdot C(\psi, \lambda) + C(\psi, \lambda) \cdot \nabla \log \rho(\psi, \lambda)\|_2^2 \\
&= \mathrm{trace}\Big(C(0,0) \cdot \mathbb{E}\Big[\nabla \log \rho(\psi, \lambda) \nabla \log \rho(\psi, \lambda)^\top\Big] C(0,0)^\top\Big) \\
&= n\pi \cdot \mathrm{trace}\Big((I - M_0)^{-1}(\Sigma_L + \Sigma_b)(I - M_0)^{-\top}\Big). \qquad \text{(A.16)}
\end{aligned}$$

We are finally in a position to put together the pieces. Applying Proposition A.1 and combining equations (A.14), (A.15), and (A.16), we obtain the lower bound

$$\int \mathbb{E}\|\widehat{x}_n(L_1^n, b_1^n) - \bar{x}_{\psi,\lambda}\|_2^2 \rho(d\psi, d\lambda) \geq \frac{\mathrm{trace}\Big((I - M_0)^{-1}(\Sigma_L + \Sigma_b)(I - M_0)^{-\top}\Big)}{9(1+\pi)n}, \qquad \text{(A.17)}$$

for any estimator $\widehat{x}_n$ that takes values in $\mathbb{R}^d$.

For the problem instances we construct, note that

$$\bar{v}^{(\psi,\lambda)} = (I - \Pi_{\mathbb{S}} L^{(\psi,\lambda)})^{-1} \Pi_{\mathbb{S}} b^{(\psi,\lambda)} = \Phi_d^* (I - M^{(\psi,\lambda)})^{-1} h^{(\psi,\lambda)} = \Phi_d^* \bar{x}_{\psi,\lambda}.$$

For any estimator $\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}$, we note that

$$\|\bar{v}^{(\psi,\lambda)} - \widehat{v}_n\|^2 \geq \|\Pi_{\mathbb{S}}(\bar{v}^{(\psi,\lambda)} - \widehat{v}_n)\|^2 = \|\Phi_d \widehat{v}_n - \bar{x}_{\psi,\lambda}\|_2^2.$$

Recall by Lemma A.3 that on the support of the prior distribution ρ, the population-level problem instance $\big(L^{(\psi,\lambda)}, b^{(\psi,\lambda)}\big)$ lies in the class $\mathbb{C}_{\mathsf{est}}$, and that the probability distribution $\mathbb{P}_{L,b}$ we constructed lies in the class $\mathbf{G}_{\mathsf{cov}}$. We thus have the minimax lower bound

$$\begin{aligned}
\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_n} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{est}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n(L_1^n, b_1^n) - \bar{v}\|^2 &\geq \inf_{\widehat{x}_n} \sup_{(\psi,\lambda) \in \mathrm{supp}(\rho)} \mathbb{E}\|\widehat{x}_n(L_1^n, b_1^n) - \bar{x}_{\psi,\lambda}\|_2^2 \\
&\geq \int \mathbb{E}\|\widehat{x}_n(L_1^n, b_1^n) - \bar{x}_{\psi,\lambda}\|_2^2 \rho(d\psi, d\lambda) \geq c \cdot \frac{\mathrm{trace}\Big((I - M_0)^{-1}\Sigma^*(I - M_0)^{-\top}\Big)}{n}
\end{aligned}$$

for $c = \frac{1}{9(1+\pi)} > 0$, which completes the proof of the theorem.

A.4.1 Proof of Lemma A.2

We first note that λ is independent of ψ, and consequently $\rho = \rho_b \otimes \rho_a$, where ρ_b, ρ_L are the marginal densities for ψ and λ respectively. Since the Fisher information tensorizes over product measures, it suffices to compute the Fisher information of ρ_L and ρ_b separately.

By a change of variables, we have

$$\rho_L(\lambda) = n^{\frac{d}{2}} \det(\Sigma_L)^{-\frac{1}{2}} \cdot \mu^{\otimes d}\Big(\sqrt{n}\Sigma_L^{-1/2}\lambda\Big).$$

Substituting this into the expression for Fisher information, we obtain

$$\begin{aligned} I(\rho_a) &= \int (\nabla \log \rho_L(\lambda))(\nabla \log \rho_a(\lambda))^\top \rho_L(\lambda) d\lambda \\ &= \int \Big(\sqrt{n}\Sigma_L^{-1/2}\nabla \log \mu^{\otimes d}(y)\Big) \cdot \Big(\sqrt{n}\Sigma_L^{-1/2}\nabla \log \mu^{\otimes d}(z)\Big)^\top \mu^{\otimes d}(z) dz \\ &= n\Sigma_L^{-1/2} \cdot \underbrace{\mathbb{E}_{Z\sim\mu^{\otimes d}}\Big[(\nabla \log \mu^{\otimes d}(Z))(\nabla \log \mu^{\otimes d}(Z))^\top\Big]}_{I\left(\mu^{\otimes d}\right)} \cdot \Sigma_L^{-1/2}. \end{aligned}$$

Finally, since $\mu^{\otimes d}$ is a product measure, we have $I\Big(\mu^{\otimes d}\Big) = I(\mu) \cdot I_d = \pi I_d$, and hence $I(\rho_L) = \pi n \Sigma_L^{-1}$. Reasoning similarly for ρ_b, we have that $I(\rho_b) = \pi n \Sigma_b^{-1}$. This completes the proof.

A.4.2 Proof of Lemma A.3

We prove the three facts in sequence.

Proof of equation (A.12a)**:** Note that the scalars σ_L and σ_b satisfies the compatibility condition (2.36), we therefore have the bounds

$$\begin{aligned} \|M^{(\psi,\lambda)} - M_0\|_F = \|\bar{x}_0\|_2^{-1} \cdot \|\lambda\|_2 \leq n^{-1/2}\|\bar{x}_0\|_2^{-1} \cdot \sqrt{\text{trace}(\Sigma_L)} \leq \sigma_L\sqrt{\frac{d}{n}}, \\ \|h^{(\psi,\lambda)} - h_0\|_2 = \|\psi\|_2 \leq \sqrt{n^{-1}\,\text{trace}(\Sigma_b)} \leq \sigma_b\sqrt{\frac{d}{n}}, \end{aligned}$$

which completes the proof of the first bound.

Proof of equation (A.12b)**:** Straightforward calculation leads to the following identities

$$\begin{aligned} \text{cov}\left(\big(M_1^{(\psi,\lambda)} - M^{(\psi,\lambda)}\big)\bar{x}_0\right) = \text{cov}\left(\|\bar{x}_0\|_2^{-2} w_1 \bar{x}_0^\top \bar{x}_0\right) = \text{cov}(w_1) = \Sigma_L, \\ \text{cov}\left(h_1^{(\psi,\lambda)} - h^{(\psi,\lambda)}\right) = \text{cov}(w_1') = \Sigma_b. \end{aligned}$$

Proof of equation (A.12c): Given any index $j \in [d]$ and vector $u \in \mathbb{S}^{d-1}$, we note that:

$$\begin{aligned}\mathbb{E}\Big(e_j^\top \big(M_1^{(\psi,\lambda)} - M^{(\psi,\lambda)}\big)u\Big)^2 &= \frac{1}{\|\bar{x}_0\|_2^4}\mathbb{E}\Big(e_j^\top w_1 \cdot \bar{x}_0^\top u\Big)^2 \\ &\le \frac{1}{\|\bar{x}_0\|_2^2}\mathbb{E}\Big(e_j^\top w_1\Big)^2 = \frac{1}{\|\bar{x}_0\|_2^2} e_j^\top \Sigma_L e_j \le \sigma_L^2,\end{aligned}$$

where the last inequality is due to the compatibility condition (2.36).

Similarly, for the noise $h_1^{(\psi,\lambda)} - h^{(\psi,\lambda)}$, we have:

$$\mathbb{E}\Big(e_j^\top \big(h_1^{(\psi,\lambda)} - h^{(\psi,\lambda)}\big)\Big)^2 = \mathbb{E}\Big(e_j^\top w_1'\Big)^2 = e_j^\top \Sigma_b e_j \le \sigma_b^2,$$

which completes the proof of the last condition.

A.4.3 Proof of claims (A.13)

We prove the two bounds separately, using the convenient shorthand I_b for the LHS of claim (A.13a) and I_a for the LHS of claim (A.13b).

Proof of claim (A.13a): We begin by applying the matrix inversion formula, which yields

$$(I - M^{(\psi,\lambda)})^{-1} = (I - M_0)^{-1} - \underbrace{\frac{1}{\|\bar{x}_0\|_2^2 - (\bar{x}_0)^\top (I - M_0)\lambda}(I - M_0)^{-1}\lambda(\bar{x}_0)^\top (I - M_0)^{-1}}_{=:H}.$$

For $n \ge 16\sigma_L^2 d$, we have

$$\big|(\bar{x}_0)^\top (I - M_0)\lambda\big| \le 2\|\bar{x}_0\|_2 \cdot \|\lambda\|_2 \le 2\|\bar{x}_0\|_2 \sqrt{n^{-1}\,\mathrm{trace}(\Sigma_L)} \le 2\sigma_L \|\bar{x}_0\|_2^2 \sqrt{\frac{d}{n}} \le \frac{1}{2}\|\bar{x}_0\|_2^2,$$

To bound T_b from below, we note that

$$\begin{aligned}T_b &= \mathrm{trace}\Big((I - M_0)^{-1}\Sigma_b (I - M_0)^{-\top}\Big) - \mathrm{trace}\Big((I - M_0)^{-1}\Sigma_b \mathbb{E}[H^\top]\Big) \\ &\ge \mathrm{trace}\Big((I - M_0)^{-1}\Sigma_b (I - M_0)^{-\top}\Big) \\ &\qquad - |\!|\!|(I - M_0)^{-1}\Sigma_b (I - M_0)^{-\top}|\!|\!|_{nuc} \cdot |\!|\!|(I - M_0)^\top \mathbb{E}[H]^\top|\!|\!|_{\mathrm{op}} \\ &= \mathrm{trace}\Big((I - M_0)^{-1}\Sigma_b (I - M_0)^{-\top}\Big) \cdot \Big(1 - |\!|\!|\mathbb{E}[H](I - M_0)|\!|\!|_{\mathrm{op}}\Big).\end{aligned}$$

When $n \geq 4\sigma_L^2 d$, we have

$$\begin{aligned}
\vert\!\vert\!\vert \mathbb{E}[H](I - M_0) \vert\!\vert\!\vert_{\mathrm{op}} &\leq \sum_{k=0}^{\infty} \frac{1}{\|\bar{x}_0\|_2^2} \vert\!\vert\!\vert \mathbb{E}\Big[\Big(\frac{(\bar{x}_0)^\top (I - M_0)\lambda}{\|\bar{x}_0\|_2^2}\Big)^k (I - M_0)^{-1}\lambda(\bar{x}_0)^\top\Big] \vert\!\vert\!\vert_{\mathrm{op}} \\
&\leq \frac{1}{\|\bar{x}_0\|_2^2} \sum_{k=1}^{\infty} \mathbb{E}\Big(\Big| \frac{(\bar{x}_0)^\top (I - M_0)\lambda}{\|\bar{x}_0\|_2^2} \Big|^k \cdot \vert\!\vert\!\vert (I - M_0)^{-1}\lambda(\bar{x}_0)^\top \vert\!\vert\!\vert_F \Big) \\
&\leq 2\sigma_L^2 \vert\!\vert\!\vert (I - M_0)^{-1} \vert\!\vert\!\vert_{\mathrm{op}} \frac{d}{n}.
\end{aligned}$$

Therefore, for $n \geq 16\sigma_L^2 \vert\!\vert\!\vert (I - M_0)^{-1} \vert\!\vert\!\vert_{\mathrm{op}} d$, we obtain the lower bound

$$T_b \geq \frac{1}{2}\,\mathrm{trace}\,\Big((I - M_0)^{-1}\Sigma_b(I - M_0)^{-\top}\Big),$$

as desired.

Proof of claim (A.13b): We note that

$$\bar{x}_{\psi,\lambda} = \bar{x}_0 - Hh_0 + (I - M_0)\psi - H\psi.$$

Consequently, we have

$$T_L = \mathbb{E}\Big[\frac{\bar{x}_0^\top(\bar{x}_0 - Hh_0 + (I - M_0)\psi - H\psi)}{\|\bar{x}_0\|_2^2}\,\mathrm{trace}\,\big((I - M_0)^{-1}\Sigma_L((I - M_0)^{-\top} - H^\top)\big)\Big].$$

Since λ is independent of ψ and H is dependent only upon λ, by taking expectation with respect to ψ, we have that

$$T_L = \mathbb{E}\Big[\frac{(\bar{x}_0)^\top(\bar{x}_0 - Hh_0)}{\|\bar{x}_0\|_2^2} \cdot \mathrm{trace}\,\Big((I - M_0)^{-1}\Sigma_L((I - M_0)^{-\top} - H^\top)\Big)\Big].$$

We note that

$$\frac{|(\bar{x}_0)^\top Hh_0|}{\|\bar{x}_0\|_2^2} = \frac{|(\bar{x}_0)^\top (I - M_0)^{-1}\lambda|}{\|\bar{x}_0\|_2^2 - (\bar{x}_0)^\top(I - M_0)} \leq 2\vert\!\vert\!\vert (I - M_0)^{-1} \vert\!\vert\!\vert_{\mathrm{op}} \sigma_L \sqrt{\frac{d}{n}}.$$

Therefore, for $n \geq 16\sigma_L^2 \vert\!\vert\!\vert (I - M_0)^{-1} \vert\!\vert\!\vert_{\mathrm{op}}^2 d$, we have the bound $\frac{1}{2} \leq \frac{(\bar{x}_0)^\top(\bar{x}_0 - Hh_0)}{\|\bar{x}_0\|_2^2} \leq \frac{3}{2}$, and consequently, we have

$$\begin{aligned}
T_L &\geq \frac{1}{2}\,\mathrm{trace}\,\Big((I - M_0)^{-1}\Sigma_L(I - M_0)^{-\top}\Big) - \frac{3}{2}\mathbb{E}\,\Big|\mathrm{trace}\,\Big((I - M_0)^{-1}\Sigma_L H^\top\Big)\Big| \\
&\geq \frac{1}{2}\,\mathrm{trace}\,\Big((I - M_0)^{-1}\Sigma_L(I - M_0)^{-\top}\Big) \\
&\qquad - \frac{3}{2}\vert\!\vert\!\vert (I - M_0)^{-1}\Sigma_L(I - M_0)^{-\top} \vert\!\vert\!\vert_{\mathrm{nuc}} \cdot \mathbb{E}\vert\!\vert\!\vert H(I - M_0) \vert\!\vert\!\vert_{\mathrm{op}} \\
&\geq \frac{1}{2}\,\mathrm{trace}\,\Big((I - M_0)^{-1}\Sigma_L(I - M_0)^{-\top}\Big) \cdot \Big(1 - 3\mathbb{E}\vert\!\vert\!\vert H(I - M_0) \vert\!\vert\!\vert_{\mathrm{op}}\Big).
\end{aligned}$$

For the matrix $H(I - M_0)$, we have the almost-sure upper bound

$$\begin{aligned}\|H(I - M_0)\|_{\mathrm{op}} &\leq \frac{2}{\|\bar{x}_0\|_2^2} \|(I - M_0)^{-1} \lambda(\bar{x}_0)^\top\|_{\mathrm{op}} \\ &\leq \frac{2}{\|\bar{x}_0\|_2} \cdot \|(I - M_0)^{-1}\lambda\|_2 \leq 2\|(I - M_0)^{-1}\|_{\mathrm{op}} \sigma_L \sqrt{\frac{d}{n}}.\end{aligned}$$

Thus, provided $n \geq 18\sigma_L^2 \|(I - M_0)^{-1}\|_{\mathrm{op}}^2 d$, putting together the pieces yields the lower bound

$$T_L \geq \frac{1}{3} \operatorname{trace}\Big((I - M_0)^{-1} \Sigma_L (I - M_0)^{-\top} \Big).$$

A.5 Proof of Corollary 2.2

Define the terms $\Delta_1 := \frac{2}{3} \alpha(M_0, \gamma_{\max})\delta^2$ and $\Delta_2 := c \cdot \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b)$. We split our proof into two cases.

Case I: if $\Delta_1 \leq \Delta_2$. Consider the function class:

$$\widetilde{\mathbb{C}} := \bigcup_{(M', h') \in \mathfrak{N}(M_0, h_0)} \mathbb{C}_{\mathsf{approx}}(M', h', D, 0, \gamma_{\max}).$$

Clearly, we have the inclusion $\widetilde{\mathbb{C}} \subseteq \mathbb{C}_{\mathsf{final}}$. Moreover, for a problem instance in $\widetilde{\mathbb{C}}$, we have $\mathcal{A}(\mathbb{S}, v^*) = 0$, and consequently $v^* = \bar{v}$. Note that the construction of problem instances in Theorem 2.3 can be embedded in $\mathbb{X}$ of any dimension D, and the linear operator L constructed in the proof of Theorem 2.3 satisfies the bound

$$\|L^{(\psi,\lambda)}\|_{\mathbb{X}} \leq \|M^{(\psi,\lambda)}\|_{\mathrm{op}} \leq \|M_0\|_{\mathrm{op}} + \sigma_L \sqrt{\frac{d}{n}} \leq \gamma_{\max}.$$

Consequently, the class $\widetilde{\mathbb{C}}$ contains the population-level problem instances $(M^{(\psi,\lambda)}, h^{(\psi,\lambda)})$ constructed in the proof of Theorem 2.3, for any choice of $\psi, \lambda \in \mathbb{R}^d$. Invoking Theorem 2.3, we thus obtain the sequence of bounds

$$\begin{aligned}\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{final}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq \inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \widetilde{\mathbb{C}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 &= \inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \widetilde{\mathbb{C}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \\ &= \inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{est}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \\ &\geq \Delta_2.\end{aligned}$$

Case II: if $\Delta_1 > \Delta_2$. In this case, we consider the class of noise distributions

$$\widetilde{\mathbf{G}} := \mathbf{G}_{\mathsf{cov}}(0, 0, \sigma_L, \sigma_b).$$

Clearly, $\widetilde{\mathbf{G}}$ is a sub-class of $\mathbf{G}_{\mathsf{cov}}$. Note that the observation model $\big(L_i^{(\varepsilon,y)}, b_i^{(\varepsilon,y)}\big)_{i=1}^n$ constructed in the proof of Theorem 2.2 satisfies the following identities almost surely:

$$\Phi_d L_i^{(\varepsilon,y)} \Phi_d^* = \Phi_d L^{(\varepsilon,y)} \Phi_d^*, \quad \Phi_d b_i^{(\varepsilon,y)} = \Phi_d b^{(\varepsilon,y)}.$$

So the problem instances constructed in the proof of Theorem 2.2 belongs to class $\widetilde{\mathbf{G}}$. Invoking Theorem 2.2, we obtain the bound

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b)\in\mathbb{C}_{\mathsf{final}} \\ \mathbb{P}_{L,b}\in\mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq \inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b)\in\mathbb{C}_{\mathsf{final}} \\ \mathbb{P}_{L,b}\in\widetilde{\mathbf{G}}}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq \frac{2}{3}(\alpha(M_0, \gamma_{\max}) - 1)\delta^2 = \Delta_1.$$

Combining the results in two cases, we arrive at the lower bound:

$$\begin{aligned}\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b)\in\mathbb{C}_{\mathsf{final}} \\ \mathbb{P}_{L,b}\in\mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 &\geq \max(\Delta_1, \Delta_2) \\ &\geq \frac{1}{2}\Delta_1 + \frac{1}{2}\Delta_2 \geq \frac{1}{3}(\alpha(M_0, \gamma_{\max}) - 1)\delta^2 + \frac{c}{2}\mathcal{E}_n(M, \Sigma_L + \Sigma_b),\end{aligned}$$

which completes the proof of the corollary.

A.6 Proof of Theorem 2.4

We begin by defining some additional notation needed in the proof. For a non-negative integer k, we use $H_k \in \mathbb{R}^{2^k \times 2^k}$ to denote the Hadamard matrix of order k, recursively defined as:

$$H_0 := 1, \quad H_k := \begin{bmatrix} H_{k-1} & H_{k-1} \\ H_{k-1} & -H_{k-1} \end{bmatrix}, \text{ for } k \geq 1$$

For any integer $q \geq 2$, we define

$$J_q := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & \vdots & & \vdots & \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix},$$

which is a $q \times q$ Jordan block with zeros in the diagonal.

Now we turn to the proof of the theorem. We assume that D is an integer multiple of q, and that $m := \frac{D-d}{q}$ is an integer power of 2; the complementary case can be handled by adjusting the constant factors in our bounds. Similarly to the proof of Theorem 2.2, we let $u \in \mathbb{S}^{d-1}$ be an eigenvector associated to the largest eigenvalue of the matrix $(I - M_0)^{-1}\big(\gamma_{\max}^2 I - M_0 M_0^\top\big)(I - M_0)^{-\top}$, and define the d-dimensional vectors:

$$w := \sqrt{\alpha(M_0, \gamma_{\max}) - 1} \cdot (I - M_0)u, \quad \text{and} \quad y := \sqrt{\alpha(M_0, \gamma_{\max}) - 1} \cdot \delta u.$$

We first construct the following $(D-d) \times (D-d)$ block matrix, indexed by the bits $(\varepsilon_{ij})_{1 \leq i \leq q, 1 \leq j \leq m}$:

$$J^{(\varepsilon)} := \begin{bmatrix} I_m & \mathrm{diag}(\varepsilon_{1j} \cdot \varepsilon_{2j})_{j=1}^m & 0 & \cdots & 0 & 0 \\ 0 & I_m & \mathrm{diag}(\varepsilon_{2j} \cdot \varepsilon_{3j})_{j=1}^m & \cdots & 0 & 0 \\ & \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & & I_m & \mathrm{diag}(\varepsilon_{(q-1)j} \cdot \varepsilon_{qj})_{j=1}^m \\ 0 & 0 & \cdots & & 0 & I_m \end{bmatrix}. \tag{A.18a}$$

Each submatrix depicted above is an $m \times m$ matrix, and the diagonal blocks are given by identity matrices. We use this construction to define the population-level instance $(L^{(\varepsilon,z)}, b^{(\varepsilon,z)})$ as follows:

$$L^{(\varepsilon,z)} := \begin{bmatrix} M_0 & \frac{\sqrt{d/2}}{D-d} z\varepsilon_{11} w & \frac{\sqrt{d/2}}{D-d} z\varepsilon_{12} w & \cdots & \frac{\sqrt{d/2}}{D-d} z\varepsilon_{1m} w & 0 & \cdots & 0 \\ 0 & & & & & & & \\ 0 & & & & & & & \\ \vdots & & & \frac{1}{2} J^{(\varepsilon)} & & & & \\ 0 & & & & & & & \end{bmatrix}, \quad \text{and}$$

$$b^{(\varepsilon,z)} := \begin{bmatrix} \sqrt{2d} h_0 \\ 0 \\ \vdots \\ 0 \\ \frac{\delta}{\sqrt{2}} \varepsilon_{q1} \\ \vdots \\ \frac{\delta}{\sqrt{2}} \varepsilon_{qm} \end{bmatrix}. \tag{A.18b}$$

It can then be verified that the solution to the fixed point equation $v^*_{\varepsilon,z} = (I - L^{(\varepsilon,z)})^{-1} b^{(\varepsilon,z)}$

is given by

$$v^*_{\varepsilon,z} = \begin{bmatrix} \frac{\sqrt{d}}{q} zy + \sqrt{2d}(I - M_0)^{-1} h_0 \\ \sqrt{2}\delta\varepsilon_{11} \\ \sqrt{2}\delta\varepsilon_{12} \\ \vdots \\ \sqrt{2}\delta\varepsilon_{q1} \\ \vdots \\ \sqrt{2}\delta\varepsilon_{qm} \end{bmatrix}. \tag{A.18c}$$

Similarly to the proof of Theorem 2.2, we take the subspace $\mathbb{S}$ to be the one spanned by first d coordinates, and take the weight vector be

$$\xi = \Big[\underbrace{\tfrac{1}{2d} \quad \cdots \quad \tfrac{1}{2d}}_{d} \quad \underbrace{\tfrac{1}{2(D-d)} \quad \cdots \quad \tfrac{1}{2(D-d)}}_{(D-d)}\Big].$$

The inner product is then given by $\langle p,\, p' \rangle := \sum_{j=1}^{D} p_j \xi_j p'_j$ for vectors $p, p' \in \mathbb{R}^D$.

It remains to define our basis vectors. For $j \in [d]$, we let $\phi_j := \sqrt{2d} e_j$. For the orthogonal complement $\mathbb{S}^\perp$, we use the basis vectors

$$\begin{bmatrix} \phi_{d+1} & \phi_{d+2} & \cdots & \phi_{d+qm} \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2q}\; H_{\log_2 m} \otimes I_q \end{bmatrix}.$$

Recall that $H_k \in \mathbb{R}^{2^k \times 2^k}$ denotes the Hadamard matrix of order k, and $\otimes$ denotes Kronecker product. The first d rows of the matrix are zeros, while the following $D - d = mq$ columns are given by the Kronecker product. By the definition of Hadamard matrix, we have $\|\phi_i\| = 1$ and $\langle \phi_i,\, \phi_j \rangle = 0$ for $i \neq j$.

As before, the construction of equations (A.18a)-(A.18c) ensures that for any choice of the binary string $\varepsilon \in \{-1, 1\}^{mq}$ and bit $y \in \{-1, 1\}$, the oracle approximation error is equal to

$$\mathcal{A}(\mathbb{S}, v^*_{\varepsilon,z}) = \inf_{v \in \mathbb{S}} \|v^*_{\varepsilon,z} - v\|^2 = \frac{1}{2(D-d)} \sum_{i=1}^{q} \sum_{j=1}^{m} (\sqrt{2}\delta\varepsilon_{ij})^2 = \delta^2. \tag{A.19a}$$

Furthermore, straightforward calculation yields that the projected matrix-vector pair takes the form

$$\Phi_d L^{(\varepsilon,z)} \Phi_d^* = M_0, \quad \text{and} \quad \Phi_d b^{(\varepsilon,z)} = h_0. \tag{A.19b}$$

Now, we construct our observation model from which samples $\big(L_i^{(\varepsilon,z)}, b_i^{(\varepsilon,z)}\big)_{i=1}^n$ are generated. For each $i \in [n]$ and $j \in [m]$, we sample independently Bernoulli random variables $\chi_{0j}^{(i)}, \chi_{1j}^{(i)}, \cdots, \chi_{qj}^{(i)} \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(1/m)$. For $k = 1, 2, \cdots, q - 1$, define the random matrix

$$Z_k^{(\varepsilon)} := \mathrm{diag}\Big(m \chi_{(k-1)j}^{(i)} \varepsilon_{(k-1)j} \varepsilon_{kj}\Big)_{j=1}^m. \tag{A.20a}$$

The random observations are then generated by the random matrix

$$J_i^{(\varepsilon)} := \begin{bmatrix} I_m & Z_1^{(\varepsilon)} & 0 & \cdots & 0 & 0 \\ 0 & I_m & Z_2^{(\varepsilon)} & \cdots & 0 & 0 \\ & \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & & I_m & Z_{q-1}^{(\varepsilon)} \\ 0 & 0 & \cdots & & 0 & I_m \end{bmatrix}, \tag{A.20b}$$

where once again, the diagonal blocks correspond to $m \times m$ identity matrices. We use this random matrix to generate the observations

$$L_i^{(\varepsilon,z)} = \left[\begin{array}{cccccccc} M_0 & \chi_{01}\frac{\sqrt{d/2}}{q} z\varepsilon_{11} w & \chi_{02}\frac{\sqrt{d/2}}{q} z\varepsilon_{12} w & \cdots & \chi_{0m}\frac{\sqrt{d/2}}{q} z\varepsilon_{1m} w & 0 & \cdots & 0 \\ 0 & & & & & & & \\ 0 & & & & & & & \\ \vdots & & & & \frac{1}{2}J_i^{(\varepsilon)} & & & \\ 0 & & & & & & & \end{array}\right] \tag{A.20c}$$

and

$$b_i^{(\varepsilon,z)} = \left[\sqrt{2d}h_0^\top \quad \underbrace{0 \quad \cdots \quad 0}_{D-d-m} \quad m\chi_{q1}\tfrac{\delta}{\sqrt{2}}\varepsilon_{q1} \quad \cdots \quad m\chi_{qm}\tfrac{\delta}{\sqrt{2}}\varepsilon_{qm}\right]^\top. \tag{A.20d}$$

This concludes our description of the problem instances themselves. As before, our proof proceeds via Le Cam's lemma, and we use similar notation for product distributions and mixtures under this observation model. Let $\mathbb{P}_{\varepsilon,z}^{(n)}$ denote the n-fold product of the probability laws of $(L_i^{(\varepsilon,z)}, b_i^{(\varepsilon,z)})$. We also define the following mixture of product measures for each $z \in \{-1, 1\}$:

$$\mathbb{P}_z^{(n)} := \frac{1}{2^{D-d}} \sum_{\varepsilon \in \{\pm 1\}^{m \times q}} \mathbb{P}_{\varepsilon,z}^{(n)}.$$

We seek bounds on the total variation distance

$$\Delta := d_{\mathrm{TV}}\Big(\mathbb{P}_1^{(n)}, \mathbb{P}_{-1}^{(n)}\Big).$$

With this setup, the following lemmas assert that (a) Our construction satisfies the operator norm condition and the noise conditions in Assumption 2.1(S) with the associated parameters bounded by dimension-independent constants, and (b) The total variation distance Δ is small provided $n \ll m^{1+1/q}$.

Lemma A.4. *For $q \in \Big[2, \frac{1}{\sqrt{2(1-1\wedge\gamma_{\max})}}\Big]$, and any $\varepsilon \in \{-1,1\}^{m\times q}$ and $z \in \{-1, 1\}$,*

(a) The construction in equation (A.18b) satisfies the bound $|\!|\!| L^{(\varepsilon,z)} |\!|\!|_{\mathbb{X}} \leq \gamma_{\max}$.

(b) The observation model defined in equation (A.20a)-(A.20d) satisfies Assumption 2.1(S) with $\sigma_L = \gamma_{\max} + 1$ and $\sigma_b = \delta/q$.

Lemma A.5. *Under the setup above, we have*

$$\Delta \leq \frac{12 n^{q+1}}{m^q}.$$

Part (a) of Lemma A.4 and equation (A.19a)-(A.19b) together ensure that population-level problem instance (L, b) we constructed belongs to the class $\mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max})$. Part (b) of Lemma A.4 further ensures the probability distribution $\mathbb{P}_{L,b}$ belongs to the class $\mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)$. Lemma A.5 ensures that the two mixture distributions corresponding to different choices of the bit z are close provided n is not too large. The final step in applying Le Cam's mixture-vs-mixture result is to show that the approximation error is large for at least one of the choices of the bit z.

Given any pair $\varepsilon, \varepsilon' \in \{-1, 1\}^{q \times m}$, we note that

$$\|v^*_{\varepsilon,1} - v^*_{\varepsilon',-1}\| \geq \left\| \begin{bmatrix} 2\sqrt{\frac{d}{q}} y^\top & 0 & \cdots & 0 \end{bmatrix}^\top \right\| = \frac{\sqrt{2}}{q} \|y\|_2 = \frac{\sqrt{2}}{q} \sqrt{\alpha(M_0, \gamma_{\max}) - 1} \cdot \delta.$$

Applying triangle inequality and Young's inequality, we have the bound

$$\begin{aligned} \frac{1}{2}(\|\widehat{v} - v^*_{\varepsilon,1}\|^2 + \|\widehat{v} - v^*_{\varepsilon',1}\|^2) &\geq \frac{1}{4}(\|\widehat{v} - v^*_{\varepsilon,1}\| + \|\widehat{v} - v^*_{\varepsilon',1}\|)^2 \\ &\geq \frac{1}{4}\|v^*_{\varepsilon,1} - v^*_{\varepsilon',-1}\|^2 \geq \frac{\alpha(M_0, \gamma_{\max}) - 1}{2q^2} \delta^2. \end{aligned}$$

Finally, applying Le Cam's lemma yields

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq \frac{\alpha(M_0, \gamma_{\max}) - 1}{2q^2} \delta^2 \cdot \Big(1 - d_{\mathrm{TV}}(\mathbb{P}^{(n)}_{-1}, \mathbb{P}^{(n)}_{1})\Big)$$

and using Lemma A.5 in conjunction with the condition $D \geq d + 3qn^{1+1/q}$, we arrive at the final bound

$$\inf_{\widehat{v}_n \in \widehat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\widehat{v}_n - v^*\|^2 \geq \frac{\alpha(M_0, \gamma_{\max}) - 1}{2q^2} \delta^2$$

as desired.

A.6.1 Proof of Lemma A.4

We prove the two parts of the lemma separately.

Proof of part (a)**:** We first show the upper bound on the operator norm. For any vector $p \in \mathbb{R}^D$, we employ the decomposition $p = \begin{bmatrix} p^{(1)} \\ p^{(2)} \end{bmatrix}$ with $p^{(1)} \in \mathbb{R}^d$ and $p^{(2)} \in \mathbb{R}^{qm}$. Assuming that $\|p\|^2 = \frac{1}{2d}\|p^{(1)}\|_2^2 + \frac{1}{2(D-d)}\|p^{(2)}\|_2^2 = 1$, we have that

$$\|L^{(\varepsilon,z)}p\|^2 = \frac{1}{2d}\|M_0 p^{(1)} + w \cdot \frac{z\sqrt{d}}{D-d}\sum_{j=1}^m \varepsilon_{1j} p_j^{(2)}\|_2^2 + \frac{1}{2(D-d)}\|\frac{1}{2}J^{(\varepsilon)}p^{(2)}\|_2^2.$$

Define the vector $a_1 := \frac{1}{\sqrt{2d}}p^{(1)}$ and scalar $a_2 := \frac{1}{\sqrt{2(D-d)}}\|p^{(2)}\|_2$ for convenience; we have the identity $\|a_1\|_2^2 + a_2^2 = 1$. Following the same arguments as in the proof of Lemma 2.7, we then have

$$\begin{aligned}
&\frac{1}{2d}\|Mp^{(1)} + w \cdot \frac{z\sqrt{d}}{D-d}\sum_{j=1}^m \varepsilon_{1j} p_j^{(2)}\|_2^2 \\
&\leq \frac{1}{2d}\sup_{t\in[-1,1]}\|M_0 p^{(1)} + \frac{\sqrt{d}}{q}a_2 t w\|_2^2 \\
&\leq \max\Big(\|M_0 a_1 + \frac{1}{q\sqrt{2}}a_2 w\|_2^2, \|M_0 a_1 - \frac{1}{q\sqrt{2}}a_2 w\|_2^2\Big) \\
&\leq |\!|\!| \begin{bmatrix} M_0 & w \end{bmatrix} |\!|\!|_{\text{op}}^2 \Big(\|a_1\|_2^2 + \frac{1}{2q^2}a_2^2\Big).
\end{aligned}$$

By the definition of the vector w, we have the bound

$$|\!|\!| \begin{bmatrix} M_0 & w \end{bmatrix} |\!|\!|_{\text{op}}^2 = \lambda_{\max}\Big(M_0 M_0^\top + ww^\top\Big) \leq \gamma_{\max}^2.$$

On the other hand, note that

$$\frac{1}{2(D-d)}\|\frac{1}{2}J^{(\varepsilon)}p^{(2)}\|_2^2 \leq \frac{1}{8(D-d)}|\!|\!| J^{(\varepsilon)} |\!|\!|_{\text{op}}^2 \cdot \|p^{(2)}\|_2^2 = \frac{1}{4}|\!|\!| J^{(\varepsilon)} |\!|\!|_{\text{op}}^2 a_2^2,$$

and consequently, that

$$\|L^{(\varepsilon,z)}p\|^2 \leq \gamma_{\max}^2\Big(\|a_1\|_2^2 + \frac{1}{2q^2}a_2^2\Big) + \frac{1}{4}|\!|\!| J^{(\varepsilon)} |\!|\!|_{\text{op}}^2 \cdot a_2^2. \tag{A.21}$$

In order to bound the operator norm of the matrix $J^{(\varepsilon)}$, we use the following fact about operator norm, proved at the end of this section for convenience. For any block matrix $T = [T_{ij}]_{1\leq i,j\leq q}$, with each block $T_{ij} \in \mathbb{R}^{m\times m}$, we have

$$|\!|\!| [T_{ij}]_{1\leq i,j\leq q} |\!|\!|_{\text{op}} \leq |\!|\!| \big[|\!|\!| T_{ij} |\!|\!|_{\text{op}}\big]_{1\leq i,j\leq q} |\!|\!|_{\text{op}}. \tag{A.22}$$

Applying equation (A.6.1) to matrix $J^{(\varepsilon)}$ yields the bound

$$|\!|\!| J^{(\varepsilon)} |\!|\!|_{\text{op}} \leq |\!|\!| I_q + J_q |\!|\!|_{\text{op}}.$$

Recall that J_q is the Jordan block of size q with zeros in the diagonal. Straightforward calculation yields the bound

$$(I_q + J_q)(I_q + J_q)^\top \preceq \begin{bmatrix} 2 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 2 & 1 & \cdots & 0 & 0 \\ & \vdots & & \vdots & & \\ 0 & 0 & \cdots & 1 & 2 & 1 \\ 0 & 0 & \cdots & 0 & 1 & 2 \end{bmatrix} =: C_q.$$

Note that C_q is a tridiagonal Toeplitz matrix, whose norm admits the closed-form expression

$$|\!|\!| C_q |\!|\!|_{\mathrm{op}} = 2 + 2\cos\left(\frac{\pi}{q+1}\right) \le 4 - \frac{4}{q^2}.$$

Therefore, we have $|\!|\!| J^{(\varepsilon)} |\!|\!|_{\mathrm{op}} \le \sqrt{|\!|\!| C_q |\!|\!|_{\mathrm{op}}} \le \sqrt{4 - 4/q^2}$. Substituting into equation (A.21), we obtain

$$\|L^{(\varepsilon,z)} p\|^2 \le \gamma_{\max}^2 \Big(\|a_1\|_2^2 + \frac{1}{2q^2} a_2^2 \Big) + \Big(1 - \frac{1}{q^2}\Big) \cdot a_2^2.$$

Invoking the condition $q \le \frac{1}{\sqrt{2(1 - 1 \wedge \gamma_{\max})}}$, we have the bound

$$\|L^{(\varepsilon,z)} p\|^2 \le \gamma_{\max}^2 \Big(\|a_1\|_2^2 + \frac{a_2^2}{2q} \Big) + \Big(1 - \frac{1}{q^2}\Big) a_2^2 \le \gamma_{\max}^2 (\|a_1\|_2^2 + a_2^2) = \gamma_{\max}^2,$$

Since the choice of the vector p is arbitrary, this yields

$$|\!|\!| L^{(\varepsilon,z)} |\!|\!|_{\mathbb{X}} \le \gamma_{\max},$$

which completes the proof.

Proof of part (b)**:** Next, we verify the noise conditions in Assumption 2.1(S). For a vector $p \in \mathbb{X}$ such that $\|p\| = 1$, denote it with $p = \begin{bmatrix} p^{(1)} \\ p^{(2)} \end{bmatrix}$, where $p^{(1)} \in \mathbb{R}^d$ and $p^{(2)} \in \mathbb{R}^{D-d}$. We have the identity $\frac{1}{2d} \|p^{(1)}\|_2^2 + \frac{1}{2(D-d)} \|p^{(2)}\|_2^2 = 1$.

For the noise $b_i^{(\varepsilon,z)} - b^{(\varepsilon,z)}$, we note that

$$\mathbb{E}\langle p,\, b_i^{(\varepsilon,z)} - b^{(\varepsilon,z)} \rangle^2 \le \frac{1}{4(D-d)^2} \sum_{j=1}^{m} \mathbb{E}\Big((m\chi_{qj} - 1) \frac{\delta}{\sqrt{2}} \varepsilon_{qj} \Big)^2 \big(p^{(2)}_{(q-1)m+j} \big)^2$$
$$\le \frac{1}{4(D-d)^2} \cdot \frac{m\delta^2}{2} \|p^{(2)}\|_2^2 \le \frac{\delta^2}{q^2}.$$

Consequently, equation (2.12b) is satisfied for $\sigma_b = \delta/q$.

In order to bound the noise in the L component, we consider the first d basis vectors and the last $(D-d)$ basis vectors separately. First, note that for each $k \in [d]$, we have

$$\begin{aligned}\mathbb{E}\langle \phi_k,\, \big(L_i^{(\varepsilon,z)} - L^{(\varepsilon,z)}\big)p\rangle^2 &= \frac{1}{4d^2}\mathbb{E}\Big(\frac{\sqrt{d/2}}{D-d}\sum_{j=1}^m z(m\chi_{1j}^{(i)}-1)\varepsilon_{1j}p_j^{(2)} \cdot \sqrt{2d}w^\top e_k\Big)^2\\ &\le \frac{1}{4}\|w\|_2^2 \cdot \frac{m}{(D-d)^2}\sum_{j=1}^m \big(p_j^{(2)}\big)^2\\ &\le \frac{\|w\|_2^2}{4q}.\end{aligned}$$

Following the derivation in Lemma 2.7, we have $\|w\|_2 \le \gamma_{\max}$, and consequently, we have the bound

$$\mathbb{E}\langle \phi_k,\, \big(L_i^{(\varepsilon,z)} - L^{(\varepsilon,z)}\big)p\rangle^2 \le \gamma_{\max}^2.$$

On the other hand, for $k \ge d+1$, the basis vector ϕ_k is constructed through the Hadamard matrix. Let $\psi^{(k)} := (2q)^{-1/2} \cdot \phi_k$, and note that the entries of $\psi^{(k)}$ are uniformly bounded by 1. Letting $k - d = (k_0-1)m + k_1$ for some choice of integers $k_0 \in [q]$ and $k_1 \in [m]$, we have

$$\begin{aligned}&\mathbb{E}\langle \phi_k,\, (L_i^{(\varepsilon,z)} - L^{(\varepsilon,z)})p\rangle^2\\ &\le \frac{1}{4(D-d)^2}\mathbb{E}\Big(\sum_{j=1}^m (m\chi_{k_0j}^{(i)}-1)\varepsilon_{k_0j}\varepsilon_{(k_0+1)j} \cdot \sqrt{2q}\psi_{d+(k_0-1)m+j}^{(k)} \cdot p_{k_0m+j}^{(2)}\Big)^2\\ &\le \frac{2mq}{4(D-d)^2}\sum_{j=1}^m \big(\psi_{d+(k_0-1)m+j}^{(k)}\big)^2 \cdot \big(p_{k_0m+j}^{(2)}\big)^2\\ &\le \frac{1}{2(D-d)}\|p^{(2)}\|_2^2 = \|p^{(2)}\|^2 \le 1.\end{aligned}$$

This verifies that equation (2.12a) is satisfied with parameter $\sigma_L = \gamma_{\max} + 1$.

Proof of equation (A.6.1): For any vector $x \in \mathbb{R}^{mq}$, consider the decomposition $x^\top = \begin{bmatrix} x_1^\top & \cdots & x_q^\top\end{bmatrix}$ with $x_j \in \mathbb{R}^m$ for each $j \in [q]$. We have the bound

$$\begin{aligned}\|Tx\|_2^2 = \sum_{i=1}^q \|\sum_{j=1}^q T_{ij}x_j\|_2^2 &\le \sum_{i=1}^q \Big(\sum_{j=1}^q |\!|\!| T_{ij} |\!|\!|_{\mathrm{op}} \|x_j\|_2\Big)^2\\ &= \|\big[|\!|\!| T_{ij}|\!|\!|_{\mathrm{op}}\big]_{1\le i,j\le q} \cdot \big[\|x_j\|_2\big]_{1\le j\le q}\|_2^2 \le |\!|\!| \big[|\!|\!| T_{ij}|\!|\!|_{\mathrm{op}}\big]_{1\le i,j\le q}|\!|\!|_{\mathrm{op}}^2 \cdot \|x\|_2^2,\end{aligned}$$

which proves this inequality.

A.6.2 Proof of Lemma A.5

For each $j \in [m]$, define the event

$$\mathscr{E}_j := \Big\{ \text{ for all } i \in \{0, 1, \cdots, q\}, \text{ there exists } \ell_i \in [n] \text{ such that } \chi_{ij}^{(\ell_i)} = 1 \Big\}. \tag{A.23}$$

Let $\mathscr{E} := \bigcup_{j=1}^m \mathscr{E}_j$. We note that under both $\mathbb{P}_1^{(n)}$ and $\mathbb{P}_{-1}^{(n)}$, we have the inequality

$$\begin{aligned}\mathbb{P}(\mathscr{E}) \le \sum_{j=1}^m \mathbb{P}(\mathscr{E}_j) &= \sum_{j=1}^m \prod_{i=0}^q \Big\{ 1 - \mathbb{P}\Big(\chi_{ij}^{(\ell)} = 1 \text{ for all } \ell \in [n]\Big)\Big\} \\ &= m \cdot \Big(1 - \Big(\frac{m-1}{m}\Big)^n\Big)^{q+1} \le \frac{n^{q+1}}{m^q}.\end{aligned}$$

As before, our choice of the event $\mathscr{E}$ was guided by the fact that the two mixture distributions are identical on its complement. In particular, we claim that

$$\mathbb{P}_1^{(n)} | \mathscr{E}^C = \mathbb{P}_{-1}^{(n)} | \mathscr{E}^C. \tag{A.24}$$

Taking this claim as given for the moment and applying Lemma A.1, we arrive at the bound

$$d_{\mathrm{TV}}(\mathbb{P}_1^{(n)}, \mathbb{P}_{-1}^{(n)}) \le \frac{12 n^{q+1}}{m^q}.$$

This completes the proof of this lemma. It remains to prove equation (A.24).

Proof of equation (A.24): We first note that both $\mathbb{P}_1^{(n)}$ are $\mathbb{P}_{-1}^{(n)}$ are m-fold i.i.d. product distributions: given $z = 1$ or $z = -1$, the random objects $\Big((\varepsilon_{ij})_{1 \le i \le q}, (\chi_{ij}^{(\ell)})_{0 \le i \le q, 1 \le \ell \le n}\Big)_{j=1}^m$ are independent and identically distributed. Now for each $j \in [m]$, let $\mathbb{Q}_{z,j}^{(n)}$ be the joint law of the random object $(\rho^{(\ell)})_{\ell=1}^n$, where $\rho^{(\ell)} = \Big(z\chi_{0j}^{(\ell)}\varepsilon_{1j}, \chi_{1j}^{(\ell)}\varepsilon_{1j}\varepsilon_{2j}, \cdots, \chi_{(q-1)j}^{(\ell)}\varepsilon_{(q-1)j}\varepsilon_{qj}, \chi_{qj}^{(\ell)}\varepsilon_{qj}\Big)$. It suffices to show that

$$\forall j \in [m], \quad \mathbb{Q}_{1,j}^{(n)} | \mathscr{E}_j^C = \mathbb{Q}_{-1,j}^{(n)} | \mathscr{E}_j^C. \tag{A.25}$$

To prove equation (A.25), we construct a distribution $\mathbb{Q}_{*,j}^{(n)}$, and show that it is equal to both of the conditional laws above. In analogy to the proof of Theorem 2.2, we construct $\mathbb{Q}_{*,j}^{(n)}$ according to the following sampling procedure:

- Sample the indicators $(\chi_{ij}^{(\ell)})_{0 \le i \le q, 1 \le \ell \le n}$, each from the Bernoulli distribution $\mathrm{Ber}(1/m)$, conditioned[1] on the event $\mathscr{E}_j^C$.

[1] Note that $\mathbb{P}(\mathscr{E}_j^C) > 0$ in this sampling procedure. So the conditional distribution is well-defined.

- For each $i \in \{0, 1, \cdots, q\}$, sample a random bit $\zeta^{(i)} \sim \mathcal{U}(\{-1, 1\})$ independently.
- For each $\ell \in [n]$, generate the random object

$$\rho^{(\ell)} = \left(\chi_{0j}^{(\ell)}\zeta^{(0)}, \chi_{1j}^{(\ell)}\zeta^{(1)}, \cdots, \chi_{(q-1)j}^{(\ell)}\zeta^{(q-1)}, \chi_{qj}^{(\ell)}\zeta^{(q)}\right).$$

In the following, we construct a coupling between $\mathbb{Q}_{z,j}^{(n)}|\mathscr{E}_j^C$ and $\mathbb{Q}_{*,j}^{(n)}$, for any $z \in \{-1, 1\}$, and show that they are actually the same.

First, we couple the random indicators $(\chi_{ij}^{(\ell)})_{0\le i\le q, 1\le \ell\le n}$ under $\mathbb{Q}_{z,j}^{(n)}|\mathscr{E}_j^C$ and $\mathbb{Q}_{*,j}^{(n)}$ directly so that they are equal almost surely. By the first step in the sampling procedure, we know that the conditional law of these indicators are the same under both probability distributions.

By definition (A.23), we note that

$$\mathscr{E}_j^C = \left\{\text{there exists } i \in \{0, 1, \cdots, q\}, \text{ such that } \chi_{ij}^{(\ell)} = 0 \text{ for all } \ell \in [n]\right\}.$$

Let the random variable $\iota \in \{0, 1, \cdots, q\}$ be the smallest such index[2] i. We construct the joint distribution by conditioning on different values of ι. First, note that the value of the random variable $\zeta^{(\iota)}$ is never observed, so we may set it to be an independent Rademacher random variable without loss of generality, and this does not affect the law of the random object $(\rho^{(\ell)})_{\ell=1}^n$ under consideration. Now consider the following three cases:

Case I: $\iota = 0$: In this case, we have $\chi_{0j}^{(\ell)} = 0$ for all $\ell \in [n]$. So the first coordinate of each $\rho^{(\ell)}$ is always zero. We define the following random variables in the probability space of $(\varepsilon_{ij})_{i=1}^q$:

$$\zeta^{(q)'} := \varepsilon_{qj}, \quad \text{and} \quad \zeta^{(i)'} := \varepsilon_{ij}\varepsilon_{(i+1)j} \text{ for } i \in \{1, 2, \cdots, q-1\}.$$

Since $(\varepsilon_{ij})_{i=1}^q$ are i.i.d. Rademacher random variables, it is easy to show by induction that the random sequence $\left(\zeta^{(i)'}\right)_{i=1}^q$ is also i.i.d. Rademacher, which has the same law as $(\zeta^{(i)})_{i=1}^q$. Consequently, we can construct the coupling such that $\left(\zeta^{(i)}\right)_{i=1}^q = \left(\zeta^{(i)'}\right)_{i=1}^q$ almost surely. Under this coupling, when $\iota = 0$, the random objects $(\rho^{(\ell)})_{\ell=1}^n$ generated under both probability distributions are almost-surely the same.

Case II: $\iota \in \{1, 2, \cdots, q-1\}$: In this case, we have $\chi_{\iota j}^{(\ell)} = 0$ for any $\ell \in [n]$. We define the following random variables in the probability space of $(\varepsilon_{ij})_{i=1}^q$:

$$\zeta^{(0)'} := z\varepsilon_{1j}, \quad \text{and} \quad \zeta^{(i)'} := \varepsilon_{ij}\varepsilon_{(i+1)j} \text{ for } i \in \{1, 2, \cdots, q-1\} \setminus \{\iota\}, \quad \text{and} \quad \zeta^{(q)'} := \varepsilon_{qj}.$$

Note that the tuple of random variables $(\zeta^{(i)'})_{i=0}^{\iota-1}$ lives in the sigma-field $\sigma(\varepsilon_{1j}, \cdots, \varepsilon_{\iota j})$, and $(\zeta^{(i)'})_{i=\iota+1}^q$, on the other hand, lives in the sigma-field $\sigma(\varepsilon_{(\iota+1)j}, \cdots, \varepsilon_{qj})$, so these

[2] Note that under the joint distribution we construct, ι is well-defined almost surely.

two tuples are independent. For any fixed bit $z \in \{-1, 1\}$, it is easy to show by induction that $\left(\zeta^{(i)'}\right)_{i=0}^{\iota-1}$ is an i.i.d. Rademacher random sequence. Similarly, applying the induction backwards from q to $\iota + 1$, we can also show that $\left(\zeta^{(i)'}\right)_{i=\iota+1}^{q}$ is also an i.i.d. Rademacher random sequence. Putting together the pieces, we see that the random variables $\left(\zeta^{(i)'}\right)_{0 \leq i \leq q, i \neq \iota}$ are i.i.d. Rademacher, and have the same law as the tuple $\left(\zeta^{(i)}\right)_{0 \leq i \leq q, i \neq \iota}$. We can therefore construct the coupling such that they are equal almost surely. Under this coupling, the random objects $(\rho^{(\ell)})_{\ell=1}^{n}$ generated under both probability distributions are almost-surely the same.

Case III: $\iota = q$**:** In this case, we have $\chi_{\iota j}^{(\ell)} = 0$ for any $\ell \in [n]$. We define the following random variables in the probability space of $(\varepsilon_{ij})_{i=1}^{q}$:

$$\zeta^{(0)'} := z\varepsilon_{1j}, \quad \text{and} \quad \zeta^{(i)'} := \varepsilon_{ij}\varepsilon_{(i+1)j} \text{ for } i \in \{1, 2, \cdots, q-1\}.$$

Note that $(\varepsilon_{ij})_{i=1}^{q}$ are i.i.d. Rademacher random variables. For each choice of the bit $z \in \{-1, 1\}$, we can show by induction that the random sequence $\left(\varepsilon^{(i)'}\right)_{i=1}^{q}$ is also i.i.d. Rademacher, which has the same law as the tuple $(\varepsilon^{(i)})_{i=1}^{q}$. Making them equal almost surely in the coupling leads to the corresponding random object $(\rho^{(\ell)})_{\ell=1}^{n}$ being almost-surely equal.

Therefore, we have constructed a coupling between $\mathbb{Q}_{z,j}^{(n)} | \mathscr{E}_j^C$ and $\mathbb{Q}_{*,j}^{(n)}$ so that the generated random objects are always the equal, for any $z \in \{-1, +1\}$. This shows that

$$\mathbb{Q}_{1,j}^{(n)} | \mathscr{E}_j^C = \mathbb{Q}_{*,j}^{(n)} = \mathbb{Q}_{-1,j}^{(n)} | \mathscr{E}_j^C,$$

which completes the proof of equation (A.25), and hence, the lemma.

A.7 Proof of the bounds on the approximation factor

In this section, we prove the claims on the quantity $\alpha(M, \gamma_{\max})$ that defines the optimal approximation factor.

A.7.1 Proof of Lemma 2.1

Recall that

$$\alpha(M, s) = 1 + \lambda_{\max}\Big((I - M)^{-1}(s^2 I_d - MM^\top)(I - M)^{-\top}\Big). \tag{A.26}$$

In the following, we prove upper bounds for the two different cases separately.

Bounds in the general case: By assumption, we have $\|M\|_{\text{op}} \leq s$, and consequently,

$$0 \preceq s^2 I_d - MM^\top \preceq s^2.$$

Thus, we have the sequence of implications

$$\begin{aligned}
\alpha(M, s) - 1 &= \lambda_{\max}\Big((I - M)^{-1}(s^2 I - MM^\top)(I - M)^{-\top}\Big) \\
&= \|(I - M)^{-1}(s^2 I - MM^\top)(I - M)^{-\top}\|_{\text{op}} \\
&\leq \|(I - M)^{-1}\|_{\text{op}} \cdot \|s^2 I_d - MM^\top\|_{\text{op}} \cdot \|(I - M)^{-1}\|_{\text{op}} \\
&\leq \|(I - M)^{-1}\|_{\text{op}}^2 \cdot s^2,
\end{aligned}$$

which proves the bound.

Bounds under non-expansive condition: When $s \leq 1$, we have

$$s^2 I - MM^\top \preceq I - MM^\top = \frac{1}{2}(I - M)(I + M^\top) + \frac{1}{2}(I + M)(I - M^\top).$$

Consequently, we have the chain of bounds

$$\begin{aligned}
\alpha(M, s) - 1 &\leq \lambda_{\max}\Big((I - M)^{-1}(I - MM^\top)(I - M)^{-\top}\Big) \\
&= \frac{1}{2}\lambda_{\max}\Big((I + M)^\top (I - M^\top)^{-1} + (I - M)^{-1}(I + M)\Big) \\
&\leq \frac{1}{2}\|(I + M)^\top (I - M^\top)^{-1}\|_{\text{op}} + \frac{1}{2}\|(I - M)^{-1}(I + M)\|_{\text{op}} \\
&\leq \|(I - M)^{-1}\|_{\text{op}} + \|(I - M^\top)^{-1}\|_{\text{op}} \\
&= 2\|(I - M)^{-1}\|_{\text{op}}.
\end{aligned}$$

Finally, we note that if $\kappa(M) < 1$, then for any $u \in \mathbb{R}^d$, we have

$$(1 - \kappa(M))\|u\|_2^2 \leq \langle (I - M)u,\, u\rangle \leq \|(I - M)u\|_2 \cdot \|u\|_2.$$

Consequently, we have $\|(I - M)^{-1}\|_{\text{op}} \leq \frac{1}{1-\kappa(M)}$, which completes the proof of this lemma.

A.7.2 Proof of Lemma 2.2

Once again, recall the definition

$$\alpha(M, s) = 1 + \lambda_{\max}\Big((I - M)^{-1}(s^2 I_d - MM^\top)(I - M)^{-\top}\Big).$$

Since M is symmetric, let $M = P\Lambda P^\top$ be its eigen-decomposition, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_d)$, and note that

$$\begin{aligned}\alpha(M, s) &= 1 + \lambda_{\max}\Big(P(I-\Lambda)^{-1}(s^2 - \Lambda^2)(I-\Lambda)^{-1}P^\top\Big)\\ &= 1 + \lambda_{\max}\Big((I-\Lambda)^{-2}(s^2-\Lambda^2)\Big)\\ &= 1 + \max_{1\leq i\leq d}\Big(\frac{\gamma_{\max}^2 - \lambda_i^2}{(1-\lambda_i)^2}\Big),\end{aligned}$$

which completes the proof.

A.8 Proofs for the examples

In this section, we provide proofs for the results related to three examples discussed in Section 2.4. Note that Corollary 2.3 follows directly from Theorem 2.1. Moreover, the proof of Corollary 2.4 builds on some technical results, and so we postpone the proof of all results related to elliptic equations to Appendix A.8.3. We begin this section with proofs of results related to temporal difference methods, i.e., Corollary 2.5 and Proposition 2.1.

A.8.1 Proof of Corollary 2.5

Recall our definition of the positive definite matrix B, with $B_{ij} = \langle \psi_i,\, \psi_j\rangle$. Letting $\theta_t := B^{1/2}\vartheta_t$, the iterates (2.41a) can be equivalently written as

$$\theta_{t+1} = \theta_t - \eta B\cdot\Big(\phi(s_{t+1})\phi(s_{t+1})^\top\theta_t - \gamma\phi(s_{t+1})\phi(s_{t+1}^+)^\top\theta_t + R_{t+1}(s_{t+1})\phi(s_{t+1})\Big), \tag{A.27}$$

and the Polyak–Ruppert averaged iterate is given by $\widehat{\theta}_n := \frac{2}{n}\sum_{t=n/2}^{n-1}\theta_t$. We also define $\bar{\theta} := \Phi_d\bar{v}$, which is the solution to projected linear equations under the orthogonal basis. Clearly, we have $\bar{\theta} = B^{1/2}\bar{\vartheta}$.

We now claim that if $n \geq \frac{c_0\varsigma^4\beta^2}{\mu^2(1-\kappa(M))^2}d\log^2\Big(\frac{\|\vartheta_0-\bar{\vartheta}\|_2 d\beta}{\mu(1-\kappa(M))}\Big)$, then

$$\|\Phi_d^*\bar{\theta} - v^*\|^2 \leq \alpha(M,\gamma)\mathcal{A}(\mathbb{S}, v^*), \text{ and} \tag{A.28a}$$

$$\mathbb{E}\|\widehat{\theta}_n - \bar{\theta}\|_2^2 \leq c\mathcal{E}_n(M,\Sigma^*) + c\big(1+\|\bar{v}\|^2\big)\Big(\frac{\varsigma^2\beta}{(1-\kappa(M))\mu}\sqrt{\frac{d}{n}}\Big)^3. \tag{A.28b}$$

Taking both inequalities as given for now, we proceed with the proof of this corollary. Combining equation (A.28a) and equation (A.28b) via Young's inequality, we arrive at the bound

$$\begin{aligned}
&\mathbb{E}\|\widehat{v}_n - v^*\|^2 \\
&\leq (1+\omega)\|\Phi_d^*\bar{\theta} - v^*\|^2 + \Big(1 + \frac{1}{\omega}\Big)\mathbb{E}\|\widehat{\theta}_n - \bar{\theta}\|_2^2 \\
&\leq (1+\omega)\mathcal{A}(\mathbb{S}, v^*) + c\Big(1+\frac{1}{\omega}\Big)\Big[\mathcal{E}_n(M, \Sigma^*) + \big(1 + \|\bar{v}\|^2\big)\Big(\frac{\varsigma^2\beta}{(1-\kappa(M))\mu}\sqrt{\frac{d}{n}}\Big)^3\Big],
\end{aligned}$$

which completes the proof of this corollary.

Proof of equation (A.28a)**:** By equation (2.23) and the definition of $\bar{\theta}$, we have

$$\bar{\theta} = \gamma M \bar{\theta} + \mathbb{E}_\xi[R(s)\phi(s)].$$

It is easy to see that $\Phi_d^*\bar{\theta}$ solves the projected Bellman equation (2.22). Note furthermore that the projected linear operator is given by

$$\Phi_d L \Phi_d^* = \gamma \Phi_d P \Phi_d^* = M.$$

Invoking the bound in equation (2.45a), we complete the proof of this inequality.

Proof of equation (A.28b)**:** Following the proof strategy for the bound (2.45b), we first show an upper bound on the iterates $\mathbb{E}\|\vartheta_t - \bar{\vartheta}\|^2$ under the non-orthogonal basis $(\psi_j)_{j\in[d]}$, and then use this bound to establish the final estimation error guarantee under $\|\cdot\|$-norm.

Recall the stochastic approximation procedure under the non-orthogonal basis:

$$\vartheta_{t+1} = \vartheta_t - \eta\Big(\psi(s_{t+1})\psi(s_{t+1})^\top\vartheta_t - \gamma\psi(s_{t+1})\psi(s_{t+1}^+)^\top\vartheta_t - R_{t+1}(s_{t+1})\psi(s_{t+1})\Big).$$

Let $\widetilde{M} := I_d - \frac{1}{\beta}B^{1/2}\Big(I_d - M\Big)B^{1/2}$ and $\widetilde{h} := \frac{1}{\beta}\mathbb{E}[R(s)\psi(s)]$. We can view equation (2.41a) as a stochastic approximation procedure for solving the linear fixed-point equation $\bar{\vartheta} = \widetilde{M}\bar{\vartheta} + \widetilde{h}$, with stochastic observations

$$\widetilde{M}_t := I_d - \beta^{-1}\Big(\psi(s_t)\psi(s_t)^\top - \gamma\psi(s_t)\psi(s_t^+)^\top\Big), \quad \text{and} \quad \widetilde{h}_t := \beta^{-1}R(s_t)\psi(s_t).$$

To verify Assumption 2.1(W), we note that for $p, q \in \mathbb{S}^{d-1}$, the following bounds directly follows from the condition (2.42):

$$\begin{aligned}
&\mathbb{E}\Big(p^\top\big(\widetilde{M}_t - \widetilde{M}\big)q\Big)^2 \\
&\leq 2\beta^{-2}\mathbb{E}\Big((p^\top\psi(s_t))\cdot(\psi(s_t)^\top q)\Big)^2 + 2\beta^{-2}\mathbb{E}\Big((p^\top\psi(s_t))\cdot(\psi(s_t^+)^\top q)\Big)^2 \\
&\leq 2\beta^{-2}\sqrt{\mathbb{E}\Big(p^\top\psi(s_t)\Big)^4\cdot\mathbb{E}\Big(\psi(s_t)^\top q\Big)^4} + 2\beta^{-2}\sqrt{\mathbb{E}\Big(p^\top\psi(s_t)\Big)^4\cdot\mathbb{E}\Big(\psi(s_t^+)^\top q\Big)^4} \\
&\leq \frac{4\varsigma^4}{\beta^2}\|B^{1/2}p\|_2^2\cdot\|B^{1/2}q\|_2^2 \leq 4\varsigma^4,
\end{aligned}$$

and

$$\mathbb{E}\Big(p^\top\big(\widetilde{h}_t - \widetilde{h}\big)\Big)^2 \leq \beta^{-2}\mathbb{E}\Big(R(s_t)\cdot p^\top \psi(s_t)\Big)^2$$
$$\leq \beta^{-2}\sqrt{\mathbb{E}\Big[R(s_t)^4\Big]\cdot \mathbb{E}\Big(p^\top B^{1/2}\phi(s_t)\Big)^4} \leq \varsigma^4/\beta.$$

Consequently, for the stochastic approximation procedure in equation (2.41a), Assumption 2.1(W) is satisfied with $\sigma_L = 2\varsigma^2$ and $\sigma_b = \varsigma^2/\sqrt{\beta}$.

To establish an upper bound on $\kappa(\widetilde{M})$, we note that

$$\begin{aligned}1-\kappa(\widetilde{M}) &= \frac{1}{\beta}\lambda_{\min}\Big(B - B^{1/2}\frac{M+M^\top}{2}B^{1/2}\Big)\\ &= \frac{1}{\beta}\inf_{u\in\mathbb{S}^{d-1}}(B^{1/2}u)^\top\Big(I_d - \frac{M+M^\top}{2}\Big)(B^{1/2}u)\\ &\geq \frac{\mu}{\beta}\inf_{u\in\mathbb{S}^{d-1}}u^\top\Big(I_d - \frac{M+M^\top}{2}\Big)u \geq \frac{\mu}{\beta}\Big(1-\kappa(M)\Big).\end{aligned}$$

Invoking Lemma 2.5, for $\eta < \frac{c_0(1-\kappa(M))\mu}{(\varsigma^4 d+1)\beta^2}$, we have

$$\mathbb{E}\|\vartheta_t - \bar{\vartheta}\|_2^2 \leq e^{-\frac{\mu}{2}(1-\kappa(M))\eta t}\mathbb{E}\|\vartheta_0 - \bar{\vartheta}\|_2^2 + \frac{8\eta\beta}{(1-\kappa(M))\mu}\Big(\|\vartheta\|_2^2\varsigma^4 d + \varsigma^4 d/\beta\Big). \tag{A.29}$$

On the other hand, applying Lemma 2.6 to the stochastic approximation procedure (A.27) under the orthogonal coordinates, we have the bound

$$\begin{aligned}\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 &\leq \frac{6}{n-n_0}\,\mathrm{trace}\,\Big((I-M)^{-1}\Sigma^*(I-M)^{-\top}\Big)\\ &\quad + \frac{6}{(n-n_0)^2}\sum_{t=n_0}^{n}\mathbb{E}\|(I - B^{1/2}\widetilde{M}B^{-1/2})^{-1}B^{1/2}(\widetilde{M}_{t+1}-\widetilde{M})B^{-1/2}(\theta_t - \bar{\theta})\|_2^2\\ &\quad + \frac{3\mathbb{E}\|\big(I_d - B^{1/2}\widetilde{M}B^{-1/2}\big)^{-1}(\theta_n - \theta_{n_0})\|_2^2}{\eta^2\beta^2(n-n_0)^2}.\end{aligned} \tag{A.30}$$

Straightforward calculation yields

$$\begin{aligned}&\mathbb{E}\|(I - B^{1/2}\widetilde{M}B^{-1/2})^{-1}B^{1/2}(\widetilde{M}_{t+1}-\widetilde{M})B^{-1/2}(\theta_t - \bar{\theta})\|_2^2\\ &= \beta^2\mathbb{E}\|(I-M)^{-1}B^{-1/2}(\widetilde{M}_{t+1}-\widetilde{M})(\vartheta_t - \bar{\vartheta})\|_2^2.\end{aligned}$$

For any vector $p\in\mathbb{R}^d$, using condition (2.42), we note that

$$\begin{aligned}&\mathbb{E}\|B^{-1/2}(\widetilde{M}_t - \widetilde{M})p\|_2^2\\ &\leq 2\beta^{-2}\mathbb{E}\|\phi(s_t)\phi(s_t)^\top B^{1/2}p\|_2^2 + 2\beta^{-2}\mathbb{E}\|\phi(s_t)\phi(s_t^+)^\top B^{1/2}p\|_2^2\\ &\leq 2\beta^{-2}\sqrt{\mathbb{E}\|\phi(s_t)\|_2^4}\cdot\sqrt{\mathbb{E}\big(\phi(s_t)^\top B^{1/2}p\big)^4} + 2\beta^{-2}\sqrt{\mathbb{E}\|\phi(s_t)\|_2^4}\cdot\sqrt{\mathbb{E}\big(\phi(s_t^+)^\top B^{1/2}p\big)^4}\\ &\leq 4\beta^{-1}\varsigma^4 d.\end{aligned}$$

Substituting into the identity above, we obtain

$$\mathbb{E}\|(I - B^{1/2}\widetilde{M}B^{-1/2})^{-1}B^{1/2}(\widetilde{M}_{t+1} - \widetilde{M})B^{-1/2}(\theta_t - \bar{\theta})\|_2^2 \leq \frac{4\beta\varsigma^4 d}{\big(1 - \kappa(M)\big)^2}\mathbb{E}\|\vartheta_t - \bar{\vartheta}\|_2^2.$$

For the third term in equation (A.30), we note that

$$\begin{aligned}\mathbb{E}\|\big(I_d - B^{1/2}\widetilde{M}B^{-1/2}\big)^{-1}(\theta_n - \theta_{n_0})\|_2^2 &= \beta^2\mathbb{E}\|(I - M)^{-1}B^{-1/2}(\vartheta_n - \vartheta_{n_0})\|_2^2 \\ &\leq \frac{2\beta^2}{\mu\big(1 - \kappa(M)\big)^2}\Big(\mathbb{E}\|\vartheta_n - \bar{\vartheta}\|_2^2 + \mathbb{E}\|\vartheta_{n_0} - \bar{\vartheta}\|_2^2\Big).\end{aligned}$$

Putting together the pieces and invoking the bound (A.29), we see that if $n_0 \geq c_0\frac{1}{\mu\eta(1-\kappa)}\log\big(\frac{d\beta}{\mu(1-\kappa)}\big)$, then

$$\begin{aligned}\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 &\leq 6\mathcal{E}_n(M, \Sigma^*) + \Bigg[\frac{24\beta\varsigma^4 d}{\big(1 - \kappa(M)\big)^2 n} + \frac{48}{\mu\big(1 - \kappa(M)\big)^2\eta^2 n^2}\Bigg] \cdot \sup_{n_0 \leq t \leq n}\mathbb{E}\|\vartheta_t - \bar{\vartheta}\|_2^2 \\ &\leq 6\mathcal{E}_n(M, \Sigma^*) + c\frac{\beta^3}{\mu^2\big(1 - \kappa(M)\big)^3}\Big[\frac{\varsigma^4\eta d}{n} + \frac{1}{\eta\beta^2 n^2}\Big]\Big(\|\bar{\vartheta}\|_2^2\varsigma^4 d + \varsigma^4 d/\beta\Big).\end{aligned}$$

Now note that $\|\bar{\vartheta}\|_2^2 = \|B^{-1/2}\bar{\theta}\|_2^2 \leq \mu^{-1}\|\bar{v}\|^2$, and so choosing the step size $\eta := \frac{1}{c_0\varsigma^2\beta\sqrt{dn}}$ yields

$$\mathbb{E}\|\widehat{v}_n - \bar{v}\|^2 \leq 6\mathcal{E}_n(M, \Sigma^*) + c\frac{\beta^3\varsigma^6}{\mu^3\big(1 - \kappa(M)\big)^3}\Big(\frac{d}{n}\Big)^{3/2}.$$

This completes the proof of equation (A.28b), and thus the corollary.

A.8.2 Proof of Proposition 2.1

Our construction and proof is inspired by the proof of Theorem 2.2, with some crucial differences in the analysis that result from the specific noise model in the MRP setting. Letting D and d be integer multiples of four without loss of generality, we denote the state space by $\mathcal{S} = \{1, 2, \cdots, D\}$. We decompose the state space into $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \mathcal{S}_2$, with $\mathcal{S}_0 := \{1, 2, \cdots, 2d\}$, $\mathcal{S}_1 := \{2d + 1, \cdots, d + \frac{D}{2}\}$, and $\mathcal{S}_2 := \{d + \frac{D}{2} + 1, \cdots, D\}$. Define the scalars $\rho = \min(\gamma, \nu) \in (0, 1)$ and $\tau := \frac{\delta}{\sqrt{2(1-\rho)}} \wedge 1$. In the following, we assume that $\delta < \sqrt{2(1 - \rho)}$. When $\delta \geq \sqrt{2(1 - \rho)}$, the lower bound for a smaller class $\mathbb{C}_{\mathsf{MRP}}(\nu, \gamma, D, \sqrt{2(1 - \rho)})$ directly applies to the original class $\mathbb{C}_{\mathsf{MRP}}(\nu, \gamma, D, \delta)$.

Given a sign $z \in \{-1, 1\}$ and subsets $\Gamma_1 \subseteq \mathcal{S}_1$ and $\Gamma_2 \subseteq \mathcal{S}_2$ such that $|\Gamma_i| = \frac{1}{2}|\mathcal{S}_i|$ for each $i \in \{1, 2\}$, we let $\bar{\Gamma}_i := \mathcal{S}_i \setminus \Gamma_i$ for $i \in \{1, 2\}$. We then construct Markov reward processes $(P^{(\Gamma_1, \Gamma_2, z)}, r^{(\Gamma_1, \Gamma_2, z)})$ and feature vectors $(\psi^{(\Gamma_1, \Gamma_2, z)}(s_i))_{i=1}^D$, indexed by the tuple

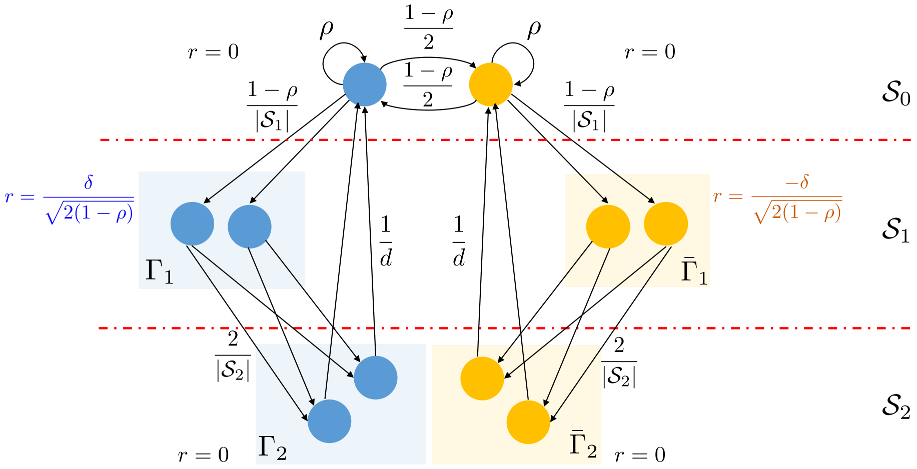


Figure A.1: A graphical illustration of the MRP instance constructed above. For this instance, we let $d=1$, $|\mathcal{S}_1|=4$ and $|\mathcal{S}_2|=4$, so that the total number of states is $D=10$. In the graph, solid rounds stand for states, and arrows stand for the possible transitions. The numbers associated to the arrows stand for the probability of the transitions, and the equations $r=\cdots$ standard for the reward at a state. The sets $\mathcal{S}_0$, $\mathcal{S}_1$ and $\mathcal{S}_2$ are separated by red dotted lines, and the sets Γ_1, $\bar{\Gamma}_1$, Γ_2, and $\bar{\Gamma}_2$ are marked by transparent rectangles. A blue round stands for a state with positive value function, and an orange round stands for a state with negative value function.

(Γ_1, Γ_2, z). Entry (i,j) of the transition matrix is given by

$$P^{(\Gamma_1,\Gamma_2,z)}(i,j) := \begin{cases} \rho & i=j\in\mathcal{S}_0, \\ \frac{1-\rho}{2} & i,j\in\mathcal{S}_0,\ |i-j|=d, \\ \frac{1-\rho}{|\mathcal{S}_1|} & (i,j)\in\Big(\{1,\cdots,d\}\times\Gamma_1\Big)\cup\Big(\{d+1,\cdots,2d\}\times\bar{\Gamma}_1\Big), \\ \frac{2}{|\mathcal{S}_2|} & (i,j)\in\Big(\Gamma_1\times\Gamma_2\Big)\cup\Big(\bar{\Gamma}_1\times\bar{\Gamma}_2\Big), \\ \frac{1}{d} & (i,j)\in\Big(\Gamma_2\times\{1,2,\cdots,d\}\Big)\cup\Big(\bar{\Gamma}_2\times\{d+1,\cdots,2d\}\Big) \\ 0 & \text{otherwise.} \end{cases} \tag{A.31a}$$

The reward function at state i is given by

$$r^{(\Gamma_1,\Gamma_2,z)}(i) := \begin{cases} z\tau & i \in \Gamma_1, \\ -z\tau & i \in \bar{\Gamma}_1, \\ 0 & \text{otherwise.} \end{cases} \tag{A.31b}$$

This MRP is illustrated in Figure A.1 for convenience. It remains to specify the feature vectors, and we use the same set of features for each tuple (Γ_1, Γ_2, z). The i-th such feature vector is given by

$$\psi(i) := \begin{cases} \sqrt{\frac{3-\rho}{2}d}e_i & i \in \{1, 2, \cdots, d\}, \\ -\sqrt{\frac{3-\rho}{2}d}e_{i-d} & i \in \{d+1, \cdots, 2d\}, \\ 0 & \text{otherwise.} \end{cases} \tag{A.31c}$$

It can be verified that for any tuple (z, Γ_1, Γ_2), the Markov chain is irreducible and aperiodic, and furthermore, that the stationary distribution of the transition kernel $P^{(\Gamma_1,\Gamma_2,z)}$ is independent of the tuple (Γ_1, Γ_2, z), and given by

$$\xi = \Big[\underbrace{\tfrac{1}{(3-\rho)d} \quad \cdots \quad \tfrac{1}{(3-\rho)d}}_{2d} \quad \underbrace{\tfrac{1-\rho}{(3-\rho)(D-2d)} \quad \cdots \quad \tfrac{1-\rho}{(3-\rho)(D-2d)}}_{D-2d}\Big].$$

By construction, we have $\mathbb{E}_\xi[\psi(s)\psi(s)^\top] = I_d$ under the stationary distribution. For the projected transition kernel, we have

$$\mathbb{E}[\psi(s)\psi(s^+)^\top] = \frac{3-\rho}{2} \cdot \Big(\rho - \frac{1-\rho}{2}\Big) I_d \preceq \rho I_d \preceq \nu I_d. \tag{A.32}$$

Given the discount factor $\gamma \in (0, 1)$, let $c_0 := \frac{(1-\rho)/2}{1-\gamma(\rho-(1-\rho)(1-\gamma^2)/2)}$ for convenience. Straightforward calculation then yields that the value function for the problem instance $\Big(P^{(\Gamma_1,\Gamma_2,z)}, r^{(\Gamma_1,\Gamma_2,z)}\Big)$ at state i is given by

$$v^*_{\Gamma_1,\Gamma_2,z}(i) = \begin{cases} c_0 z\tau & i \in \{1, 2, \cdots, d\}, \\ -c_0 z\tau & i \in \{d+1, \cdots, 2d\}, \\ (1+\gamma^2 c_0)z\tau & i \in \Gamma_1, \\ -(1+\gamma^2 c_0)z\tau & i \in \bar{\Gamma}_1, \\ \gamma c_0 z\tau & i \in \Gamma_2, \\ -\gamma c_0 z\tau & i \in \bar{\Gamma}_2. \end{cases}$$

For $\rho > 1/2$, we have the bounds

$$c_0 \geq \frac{1}{4} \cdot \frac{1-\rho}{1-\gamma\rho} \geq \frac{1-\rho}{4(1-\rho^2)} \geq \frac{1}{8}, \quad \text{and} \quad c_0 \leq \frac{1-\rho}{1-\gamma\rho} \leq 1.$$

Consequently, we have $|v^*_{\Gamma_1,\Gamma_2,z}(i)| \asymp |v^*_{\Gamma_1,\Gamma_2,z}(j)|$ for each pair (i,j), where $\asymp$ denotes an equivalence up to a universal constant factor.

Note that by our construction, the subspace $\mathbb{S}$ spanned by the basis functions $\psi(1), \psi(2), \cdots, \psi(2d)$ is given by

$$\mathbb{S} = \Big\{ v \in \mathbb{L}^2(\mathcal{S}, \xi) : v(s) = 0 \text{ for } s \notin \mathcal{S}_0, \text{ and } v(i+d) = -v(i) \text{ for all } i \in [d] \Big\}.$$

Consequently, we have

$$\inf_{v \in \mathbb{S}} \|v - v^*_{\Gamma_1,\Gamma_2,z}\|^2 = \frac{1-\rho}{3-\rho} \cdot \Big(\frac{1}{2}(1+\gamma^2 c_0)^2 \tau^2 + \frac{1}{2} \gamma^2 c_0^2 \tau^2 \Big) \leq 2(1-\rho)\tau^2 = \delta^2. \quad \text{(A.33)}$$

Putting the equations (A.32) and (A.33) together, for any tuple (Γ_1, Γ_2, z), we conclude that the problem instance $(P^{(\Gamma_1,\Gamma_2,z)}, r^{(\Gamma_1,\Gamma_2,z)}, \gamma, \psi^{((\Gamma_1,\Gamma_2,z))})$ belongs to the class $\mathbb{C}_{\mathsf{MRP}}(\nu, \gamma, D, \delta)$.

Having verified particular properties of our MRP construction, we are now ready to carry out the lower bound argument via Le Cam's lemma, as above. We define the following mixture distributions for each $z \in \{-1, 1\}$:

$$\mathbb{P}^{(n)}_z := \binom{|\mathcal{S}_1|}{|\mathcal{S}_1|/2}^{-2} \sum_{\substack{\Gamma_1 \subseteq \mathcal{S}_1, \Gamma_2 \subseteq \mathcal{S}_2 \\ |\Gamma_1| = |\Gamma_2| = \frac{1}{2}|\mathcal{S}_1|}} \mathbb{P}^{\otimes n}_{\Gamma_1,\Gamma_2,z},$$

where $\mathbb{P}_{\Gamma_1,\Gamma_2,z}$ is the law of an observed tuple $(s_i, s_i^+, r(s_i))$ under the MRP $\big(P^{(\Gamma_1,\Gamma_2,z)}, \gamma, r^{(\Gamma_1,\Gamma_2,z)}\big)$, and $\mathbb{P}^{\otimes n}_{\Gamma_1,\Gamma_2,z}$ denotes its n-fold product. Our next result gives a bound on the total variation distance between $\mathbb{P}^{(n)}_1$ and $\mathbb{P}^{(n)}_{-1}$.

Lemma A.6. *Under the set-up above, we have* $d_{\mathrm{TV}}\Big(\mathbb{P}^{(n)}_1, \mathbb{P}^{(n)}_{-1}\Big) \leq \frac{Cn^2}{D-2d}$.

Taking this lemma as given, we now turn to the proof of the proposition. Consider any estimator $\widehat{v}$ for the value function. For any pair Γ_1, Γ_2 and Γ_1', Γ_2', we have

$$\|\widehat{v} - v^*_{\Gamma_1,\Gamma_2,1}\|^2 + \|\widehat{v} - v^*_{\Gamma_1',\Gamma_2',-1}\|^2 \geq \frac{1}{2}\|v^*_{\Gamma_1,\Gamma_2,1} - v^*_{\Gamma_1',\Gamma_2',-1}\|^2 \geq \frac{1}{2} c_0^2 \tau^2 \geq \frac{\delta^2}{64(1-\rho)}.$$

Let C denote the constant appearing in Lemma A.6, and suppose that $D > 2C(n^2 + d)$. Invoking Le Cam's lemma yields the lower bound

$$\inf_{\widehat{v}_n} \sup_{(P,\gamma,r,\psi) \in \mathbb{C}_{\mathsf{MRP}}} \geq \frac{c}{1-\rho} \delta^2 \Big(1 - d_{\mathrm{TV}}(\mathbb{P}^{(n)}_1, \mathbb{P}^{(n)}_{-1}) \Big) \geq \frac{c'}{1-\nu\gamma} \delta^2,$$

which completes the proof.

A.8.2.1 Proof of Lemma A.6

Our argument is spiritually similar to the proof of Theorem 2.2, but involves some delicate technical work owing to the nature of the sampling model. In contrast to the proof of Theorem 2.2, the underlying mixing components here are indexed by random subsets of a given size, instead of random bit strings. This sampling procedure introduces additional dependency, so that the arguments in the proof of Theorem 2.2 do not directly apply. Instead, we use an induction-type argument by constructing the coupling directly.

Similarly to before, we construct a probability distribution $\mathbb{Q}^{(n)}$ and bound the total variation distance between $\mathbb{Q}^{(n)}$ and $\mathbb{P}_z^{(n)}$ for each $z \in \{-1, 1\}$. In particular, for $k \in [n]$, we let $\mathbb{Q}^{(k)}$ be the law of k independent samples drawn from the following observation model:

- (Initial state:) Generate the state $s_i \sim \xi$.
- (Next state:) If $s_i \in \mathcal{S}_1$, then generate $s_i^+ \sim \mathcal{U}(\mathcal{S}_2)$. If $s_i \in \mathcal{S}_2$, then generate $s_i^+ \sim \mathcal{U}(\mathcal{S}_0)$. On the other hand, if $s_i \in \mathcal{S}_0$, then generate $S \sim \mathcal{U}(\mathcal{S}_1)$ and let[3]

$$s_i^+ = \begin{cases} s_i & \text{w.p. } \rho, \\ (s_i + d) \mod 2d & \text{w.p. } \frac{1-\rho}{2}, \\ S & \text{w.p. } \frac{1-\rho}{2}. \end{cases} \tag{A.34}$$

- (Reward:) If $s_i \in \mathcal{S}_1$, randomly draw $R_i = \zeta^{(i)} \sim \mathcal{U}(\{-1, 1\})$, and output $\zeta^{(i)}\tau$ as the reward. Otherwise, output the reward $R_i = 0$.

To bound the total variation distance $d_{\mathrm{TV}}(\mathbb{Q}^{(n)}, \mathbb{P}_z^{(n)})$, we use the following recursive relation, which holds for each $k = 0, 1, \cdots, n-1$:

$$\begin{aligned} d_{\mathrm{TV}}\Big(\mathbb{Q}^{(k+1)}, \mathbb{P}_z^{(k+1)}\Big) &\leq d_{\mathrm{TV}}\Big(\mathbb{Q}^{(k)}, \mathbb{P}_z^{(k)}\Big) \\ &\quad + \sup_{(s_i, s_i^+, R_i)_{i=1}^k} d_{\mathrm{TV}}\Big(\mathbb{Q}^{(k+1)}|(s_i, s_i^+, R_i)_{i=1}^k, \mathbb{P}_z^{(k+1)}|(s_i, s_i^+, R_i)_{i=1}^k\Big). \end{aligned} \tag{A.35}$$

Owing to the i.i.d. nature of the sampling model for $\mathbb{Q}^{(k+1)}$, note that we have the equivalence $(s_{k+1}, s_{k+1}^+, R_{k+1})|(s_i, s_i^+, R_i)_{i=1}^k \overset{d}{=} (s_{k+1}, s_{k+1}^+, R_{k+1})$.

At this juncture, it is helpful to view the probability distributions $\mathbb{P}_1^{(k)}$ and $\mathbb{P}_{-1}^{(k)}$ via the following two-step sampling procedure: First, for $j \in \{1, 2\}$, sample the subsets $\Gamma_j \subseteq \mathcal{S}_j$ uniformly at random from the collection of all subsets of size $|\mathcal{S}_j|/2$. Then, generate k i.i.d. samples $(s_i, s_i^+, R_i)_{i=1}^k$ according to the observation model (A.31a)-(A.31b). Consequently, for the rest of this proof, we view Γ_1 and Γ_2 as *random sets*.

With this equivalence in hand, the following technical lemma shows that the posterior distribution of the subsets (Γ_1, Γ_2) conditioned on sampling the tuple $(s_i, s_i^+, R_i)_{i=1}^k$ is very close to the distribution of subsets chosen uniformly at random.

[3] The expression $a \mod b$ denotes the remainder of a divided by b, when a and b are integers.

Lemma A.7. *There is a universal positive constant c such that for each bit $z \in \{\pm 1\}$ and indices $j \in \{1, 2\}$ and $k \in [n]$, the following statement is true almost surely. For each tuple $(s_i, s_i^+, R_i)_{i=1}^k$ in the support of $\mathbb{P}_z^{(k)}$, the posterior distribution of Γ_j conditioned on $(s_i, s_i^+, R_i)_{i=1}^k \sim \mathbb{P}_z^{(k)}$ satisfies*

$$\max_{s \in \mathcal{S}_j \setminus \cup_{i=1}^k \{s_i, s_i^+\}} \left| \mathbb{P}\Big(\Gamma_j \ni s \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) - \frac{1}{2} \right| \leq \frac{ck}{D-d}.$$

In words, for any "observable" tuple $(s_i, s_i^+, R_i)_{i=1}^k$ and each state $s \in \mathcal{S}_j \setminus \cup_{i=1}^k \{s_i, s_i^+\}$, the posterior probability of the event $\{\Gamma_j \ni s\}$ conditioned on observing the tuple $(s_i, s_i^+, R_i)_{i=1}^k$ is close to $1/2$ provided $D - d$ is large relative to k. In addition to the sets $\Gamma_j, j = 1, 2$ being close to uniformly random, we also require the following analog of a "birthday-paradox" argument in this setting. For convenience, we let $\mathcal{T}_k := \bigcup_{i=1}^k \{s_i, s_i^+\}$ denote the subset of states seen up until sample k.

Lemma A.8. *There is a universal positive constant c such that for each $k \in [n]$ and each distribution $\mathbb{M}^{(k+1)} \in \left\{\mathbb{P}_{-1}^{(k+1)}, \mathbb{P}_{-1}^{(k+1)}, \mathbb{Q}^{(k+1)}\right\}$, the following statement holds almost surely. For each tuple $(s_i, s_i^+, R_i)_{i=1}^{k+1}$ in the support of $\mathbb{M}^{(k+1)}$, the probability the tuple of states $\left\{s_{k+1}, s_{k+1}^+\right\}$ conditioned on $(s_i, s_i^+, R_i)_{i=1}^k \sim \mathbb{M}^{(k)}$ satisfies*

$$\mathbb{P}\Big(\underbrace{\left\{s_{k+1}, s_{k+1}^+\right\} \cap \mathcal{T}_k \cap (\mathcal{S}_1 \cup \mathcal{S}_2) \neq \varnothing}_{:= \mathscr{E}_{k+1}^{(1)}} \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) \leq \frac{ck}{D-d}. \tag{A.36}$$

In words, Lemma A.8 ensures that if $D - d$ is large relative to k, then the states seen in sample $k+1$ are different from those seen up until that point (provided we only count states in the set $\mathcal{S}_1 \cup \mathcal{S}_2$). Lemmas A.7 and A.8 are both proved at the end of this section; we take them as given for the rest of this proof.

Now consider tuples $(s_{k+1}, s_{k+1}^+, R_{k+1}) \sim \mathbb{P}_z^{(k+1)} | (s_i, s_i^+, R_i)_{i=1}^k$ and $(\widetilde{s}_{k+1}, \widetilde{s}_{k+1}^+, \widetilde{R}_{k+1}) \sim \mathbb{Q}^{(k+1)} | (s_i, s_i^+, R_i)_{i=1}^k$; we will now construct a coupling between these two tuples in order to show that the total variation between between the respective laws is small. First, note that under both $\mathbb{P}_z^{(k+1)}$ and $\mathbb{Q}^{(k+1)}$, the initial state is drawn from the stationary distribution, i.e., $s_{k+1}, \widetilde{s}_{k+1} \sim \xi$, regardless of the sequence $(s_i, s_i^+, R_i)_{i=1}^k$. We can therefore couple the two conditional laws together so that $s_{k+1} = \widetilde{s}_{k+1}$ almost surely. To construct the coupling for the rest, we consider the following three cases:

Coupling on the event $s_{k+1} \in \mathcal{S}_0$: We begin by coupling the reward random variables; we have $R_{k+1} = \widetilde{R}_{k+1} = 0$ under both conditional distributions, so this component of the distribution can be coupled trivially. Next, we couple the next state: By construction of the observation models (A.31a) and (A.34), we have

$$\mathbb{P}\big(s_{k+1}^+ = s_{k+1} | s_{k+1}\big) = \mathbb{P}\big(\widetilde{s}_{k+1}^+ = \widetilde{s}_{k+1} | \widetilde{s}_{k+1}\big) = \rho, \quad \text{and}$$
$$\mathbb{P}\big(s_{k+1}^+ = s_{k+1} + d \mod 2d \mid s_{k+1}\big) = \mathbb{P}\big(\widetilde{s}_{k+1}^+ = \widetilde{s}_{k+1} + d \mod 2d \mid \widetilde{s}_{k+1}\big) = \frac{1-\rho}{2},$$

and so these two components of the distribution can be coupled trivially. It remains to handle the case where $s_{k+1} \in \mathcal{S}_0$ and $s_{k+1}^+ \in \mathcal{S}_1$. By the symmetry of elements within set $\mathcal{S}_1$, we note that on the event $\big(\mathscr{E}_{k+1}^{(1)}\big)^C$, both random variables $\widetilde{s}_{k+1}^+$ and s_{k+1}^+ are uniformly distributed on the set $\mathcal{S}_1 \setminus \mathcal{T}_k$. Consequently, on the event $\big(\mathscr{E}_{k+1}^{(1)}\big)^C$, we can couple the conditional laws so that $s_{k+1}^+ = \widetilde{s}_{k+1}^+$ almost surely.

Coupling on the event $s_{k+1} \in \mathcal{S}_1$: As before, we begin by coupling the rewards, but first, note that on the event $\big(\mathscr{E}_{k+1}^{(1)}\big)^C$, we have $s_{k+1} \in \mathcal{S}_1 \setminus \mathcal{T}_k$. Invoking Lemma A.7, under $\mathbb{P}_z^{(k)}$ and conditionally on the value of s_{k+1}, we have the bound

$$\left|\mathbb{P}\Big(s_{k+1} \in \Gamma_1 \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) - \frac{1}{2}\right| \leq \frac{ck}{D-d}.$$

Now the reward function (A.31b) satisfies $r(s) = z\tau$ for $s \in \Gamma_1$ and $r(s) = -z\tau$ for $s \in \bar{\Gamma}_1$. On the other hand, under $\mathbb{Q}^{(k+1)}$, the reward $\widetilde{R}_{k+1}$ takes value of τ and $-\tau$, each with probability half. Consequently, there exists a coupling between R_{k+1} and $\widetilde{R}_{k+1}$ such that

$$\mathbb{P}\Big(\underbrace{R_{k+1} \neq \widetilde{R}_{k+1},\ s_{k+1} \in \mathcal{S}_1}_{:=\mathscr{E}_{k+1}^{(2)}} \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) \leq \frac{ck}{D-d}.$$

Next, we construct the coupling for next-step transition conditionally on the current step. By the symmetry of elements within set $\mathcal{S}_2$, we note that under $\big(\mathscr{E}_{k+1}^{(1)}\big)^C$, both random variables $\widetilde{s}_{k+1}^+$ and s_{k+1}^+ are uniformly distributed on the set $\mathcal{S}_2 \setminus \mathcal{T}_k$. Consequently, on the event $\big(\mathscr{E}_{k+1}^{(1)}\big)^C$, we can couple the conditional laws so that $s_{k+1}^+ = \widetilde{s}_{k+1}^+$ almost surely.

Coupling on the event $s_{k+1} \in \mathcal{S}_2$: In this case, we have $R_{k+1} = \widetilde{R}_{k+1} = 0$ under both conditional distributions, so this coupling is once again trivial. It remains to construct a coupling between next-step transitions s_{k+1}^+ and $\widetilde{s}_{k+1}^+$. On the event $\big(\mathscr{E}_{k+1}^{(1)}\big)^C$, we have $s_{k+1} \in \mathcal{S}_2 \setminus \mathcal{T}_k$. Under $\mathbb{P}_z^{(k)}$ and conditionally on the value of s_{k+1}, Lemma A.7 leads to the bound

$$\left|\mathbb{P}\Big(s_{k+1} \in \Gamma_2 \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) - \frac{1}{2}\right| \leq \frac{ck}{D-d}.$$

By definition, under $\mathbb{P}_z^{(n)}$, we have that $s_{k+1}^+ \sim \mathcal{U}(\{1, 2, \cdots, d\})$ when $s_{k+1} \in \Gamma_2$, and $s_{k+1}^+ \sim \mathcal{U}(\{d+1, \cdots, 2d\})$ when $s_{k+1} \in \bar{\Gamma}_2$. Under $\mathbb{Q}^{(n)}$, we have $\widetilde{s}_{k+1}^+ \sim \mathcal{U}(\{1, 2, \cdots, 2d\})$. Consequently, there exists a coupling such that

$$\mathbb{P}\Big(\underbrace{s_{k+1}^+ \neq \widetilde{s}_{k+1}^+, \;\, s_{k+1} \in \mathcal{S}_2}_{:= \mathscr{E}_{k+1}^{(3)}} \mid (s_i, s_i^+, R_i)_{i=1}^k \Big) \leq \frac{ck}{D-d}.$$

Putting together our bounds from the three cases, note that for any sequence $(s_i, s_i^+, R_i)_{i=1}^k$ on the support of $\mathbb{Q}^{(k)}$ and $\mathbb{P}_z^{(k)}$, we almost surely have

$$\begin{aligned} & d_{\mathrm{TV}}\Big(\mathcal{L}\Big[(s_{k+1}, s_{k+1}^+, R_{k+1}) \mid (s_i, s_i^+, R_i)_{i=1}^k \Big], \mathcal{L}\Big[(\widetilde{s}_{k+1}, \widetilde{s}_{k+1}^+, \widetilde{R}_{k+1}) \mid (s_i, s_i^+, R_i)_{i=1}^k \Big] \Big) \\ & \leq \sum_{j=1}^{3} \mathbb{P}\Big(\mathscr{E}_{k+1}^{(j)} \mid (s_i, s_i^+, R_i)_{i=1}^k \Big) \leq \frac{c'k}{D-d}, \end{aligned}$$

where the final inequality follows from applying Lemma A.8. Substituting into the recursion (A.35), we conclude that for any $z \in \{-1, 1\}$, we have

$$d_{\mathrm{TV}}\Big(\mathbb{Q}^{(n)}, \mathbb{P}_z^{(n)} \Big) \leq \sum_{k=0}^{n-1} \sum_{j=1}^{3} \sup_{(s_i, s_i^+, R_i)_{i=1}^k} \mathbb{P}\Big(\mathscr{E}_{k+1}^{(j)} \mid (s_i, s_i^+, R_i)_{i=1}^k \Big) \leq \frac{c'n^2}{D-d},$$

which completes the proof of this lemma.
It remains to prove the two helper lemmas.

Proof of Lemma A.7: Given $z \in \{\pm 1\}$, we define the sets

$$\begin{aligned} Z_1 &:= \Big\{ s_i : i \in [k], s_i \in \mathcal{S}_1, R_i = z\tau \Big\}, \quad \bar{Z}_1 := \big(\{s_i\}_{i \in [k]} \cap \mathcal{S}_1 \big) \setminus Z_1, \quad \text{and} \\ Z_2 &:= \Big\{ s_i : i \in [k], s_i \in \mathcal{S}_2, s_i^+ \in [d] \Big\}, \quad \bar{Z}_2 := \big(\{s_i\}_{i \in [k]} \cap \mathcal{S}_2 \big) \setminus Z_2 \end{aligned}$$

By the reward model (A.31b) in our construction, for any valid pair of subsets (Γ_1, Γ_2), under the law $\mathbb{P}_{\Gamma_1, \Gamma_2, z}^{\otimes k}$, the observations $(s_i, s_i^+, R_i)_{i=1}^k$ have positive probability if and only if $Z_1 \subseteq \Gamma_1$ and $\Gamma_1 \cap \bar{Z}_1 = \varnothing$. Furthermore, by the symmetry between the elements in Γ_1, for any Γ_1 such that $Z_1 \subseteq \Gamma_1$ and $\Gamma_1 \cap \bar{Z}_1 = \varnothing$, the probability of observing $(s_i, s_i^+, R_i)_{i=1}^k$ under $\mathbb{P}_{\Gamma_1, \Gamma_2, z}^{\otimes k}$ is independent of the choice of Γ_1. Consequently, the probability under the mixture distribution $\mathbb{P}_z^{(k)}$ can be calculated as

$$\begin{aligned}
&\mathbb{P}\Big(\Gamma_1 \ni s \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) \\
&= \sum_{\substack{s \in \Gamma' \\ |\Gamma'| = |\mathcal{S}_1|/2}} \frac{\mathbb{P}\Big((s_i, s_i^+, R_i)_{i=1}^k \mid \Gamma_1 = \Gamma'\Big) \cdot \mathbb{P}(\Gamma_1 = \Gamma')}{\mathbb{P}\Big((s_i, s_i^+, R_i)_{i=1}^k\Big)} \\
&= \frac{\Big|\Big\{\Gamma' \subseteq \mathcal{S}_1 : \ |\Gamma'| = \frac{1}{2}|\mathcal{S}_1|, \ Z_1 \subseteq \Gamma', \ \bar{Z}_1 \cap \Gamma' = \varnothing, \ s \in \Gamma'\Big\}\Big|}{\Big|\Big\{\Gamma' \subseteq \mathcal{S}_1 : \ |\Gamma'| = \frac{1}{2}|\mathcal{S}'|, \ Z_1 \subseteq \Gamma' \ \bar{Z}_1 \cap \Gamma' = \varnothing\Big\}\Big|} \\
&= \binom{|\mathcal{S}_1| - |Z_1| - |\bar{Z}_1|}{|\mathcal{S}_1|/2 - |Z_1|}^{-1} \binom{|\mathcal{S}_1| - |Z_1| - |\bar{Z}_1| - 1}{|\mathcal{S}_1|/2 - |Z_1| - 1} \\
&= \frac{|\mathcal{S}_1|/2 - |Z_1|}{|\mathcal{S}_1| - |Z_1| - |\bar{Z}_1|}.
\end{aligned}$$

By definition, we have $|Z_1| + |\bar{Z}_1| \leq k$, and $|\mathcal{S}_1| = \frac{D-2d}{2}$. For $D \geq d + 8k$, this yields

$$\left|\mathbb{P}\Big(\Gamma_1 \ni s \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) - \frac{1}{2}\right| \leq \frac{4k}{D - 2d} \leq \frac{8k}{D - d}.$$

Similarly, by the transition model (A.31a) in our construction, for any $\Gamma_2 \subseteq \mathcal{S}_2$ with $|\Gamma_2| = \frac{1}{2}|\mathcal{S}_2|$, under the law $\mathbb{P}_{\Gamma_1, \Gamma_2, z}^{\otimes k}$, the observations $(s_i, s_i^+, R_i)_{i=1}^k$ have positive probability if and only if $Z_2 \subseteq \Gamma_2$ and $\Gamma_2 \cap \bar{Z}_2 = \varnothing$. Following exactly the same calculation as above, we arrive at the bound

$$\left|\mathbb{P}\Big(\Gamma_2 \ni s \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) - \frac{1}{2}\right| \leq \frac{8k}{D - d},$$

as desired.

Proof of Lemma A.8: Under the conditional distribution $\mathbb{M}^{(k+1)} | (s_i, s_i^+, R_i)_{i=1}^k$, for each $s \in \mathcal{S}_1 \cup \mathcal{S}_2$, we have

$$\mathbb{P}\Big(s_{k+1} = s\Big) \leq \frac{2}{|\mathcal{S}_1|}, \quad \text{and} \quad \mathbb{P}\Big(s_{k+1}^+ = s\Big) \leq \frac{2}{|\mathcal{S}_1|}.$$

Applying the union bound yields

$$\begin{aligned}
&\mathbb{P}\Big(\mathscr{E}_{k+1}^{(1)} \mid (s_i, s_i^+, R_i)_{i=1}^k\Big) \\
&\leq \sum_{\substack{i \in [k] \\ s_i \in \mathcal{S}_1 \cup \mathcal{S}_2}} \Big(\mathbb{P}\Big(s_{k+1} = s_i\Big) + \mathbb{P}\Big(s_{k+1}^+ = s_i\Big)\Big) + \sum_{\substack{i \in [k] \\ s_i \in \mathcal{S}_1 \cup \mathcal{S}_2}} \Big(\mathbb{P}\Big(s_{k+1} = s_i^+\Big) + \mathbb{P}\Big(s_{k+1}^+ = s_i^+\Big)\Big) \\
&\leq \frac{8k}{|\mathcal{S}_1|} \leq \frac{32k}{D - d},
\end{aligned}$$

which completes the proof.

A.8.3 Proofs of results for elliptic equations

In this section, we prove our results for the elliptic equation example in Section 2.2.2.2, splitting the section into proofs of technical results and Corollary 2.4.

A.8.3.1 Technical results from Section 2.2.2.2

The main technical result that was assumed in Section 2.2.2.2 is collected as a lemma below.

Lemma A.9. *There exists a bounded, self-adjoint, linear operator $\widetilde{A} \in \mathfrak{L}$ and a function $g \in \dot{\mathbb{H}}^1$ such that, for all $u, v \in \dot{\mathbb{H}}^1$,*

$$\langle u,\, \widetilde{A}v\rangle_{\dot{\mathbb{H}}^1} = \langle u,\, Av\rangle_{\mathbb{L}^2}, \tag{A.37a}$$

$$\langle u,\, g\rangle_{\dot{\mathbb{H}}^1} = \langle u,\, f\rangle_{\mathbb{L}^2}, \tag{A.37b}$$

and such that

$$\mu\|u\|^2_{\dot{\mathbb{H}}^1} \le \langle u,\, \widetilde{A}u\rangle_{\dot{\mathbb{H}}^1} \le \beta\|u\|^2_{\dot{\mathbb{H}}^1}. \tag{A.37c}$$

We prove the three claims in turn.

Proof of equation (A.37a): For any pair of test functions $u, v \in \dot{\mathbb{H}}^1$, integration by parts and the uniform ellipticity condition yield

$$\langle u,\, Av\rangle_{\mathbb{L}^2} = -\int_\Omega u(x)\nabla\cdot\big(a(x)\nabla v(x)\big)dx = \int_\Omega \nabla u^\top a\nabla v dx \le \beta\|u\|_{\dot{\mathbb{H}}^1}\cdot\|v\|_{\dot{\mathbb{H}}^1}.$$

Now given a fixed function $v \in \dot{\mathbb{H}}^1$, the above equation ensures that $\langle\cdot,\, Av\rangle_{\mathbb{L}^2}$ is a bounded linear functional. By the Riesz representation theorem, there exists a unique function $v' \in \dot{\mathbb{H}}^1$ with $\|v'\|_{\dot{\mathbb{H}}^1} \le \beta\|v\|_{\dot{\mathbb{H}}^1}$ such that

$$\forall u \in \dot{\mathbb{H}}^1, \quad \langle u,\, Av\rangle_{\mathbb{L}^2} = \langle u,\, v'\rangle_{\dot{\mathbb{H}}^1}.$$

Clearly, the mapping from v to v' is linear, and we have $\|v'\|_{\dot{\mathbb{H}}^1} \le \beta\|v\|_{\dot{\mathbb{H}}^1}$ for any $v \in \dot{\mathbb{H}}^1$. Thus, the mapping $v \mapsto v'$ is a bounded linear operator. Using $\widetilde{A}$ to denote this operator, equation (A.37a) then directly follows. It remains to verify that $\widetilde{A}$ is self-adjoint. Indeed, for $u, v \in \dot{\mathbb{H}}^1$, we have the identity

$$\langle u,\, \widetilde{A}v\rangle_{\dot{\mathbb{H}}^1} = \langle u,\, Av\rangle_{\mathbb{L}^2} = \int_\Omega \nabla u^\top a\nabla v dx = -\int_\Omega v\nabla\cdot(a\nabla u)dx = \langle v,\, Au\rangle_{\mathbb{L}^2} = \langle v,\, \widetilde{A}u\rangle_{\dot{\mathbb{H}}^1},$$

which proves the self-adjoint property.

Proof of equation (A.37b)**:** Since the domain Ω is bounded and connected, there exists a constant ρ_P, depending only on Ω, such that the following Poincaré equation holds:

$$\|v\|_{\mathbb{L}^2}^2 \leq \frac{1}{\rho_P}\|v\|_{\dot{\mathbb{H}}^1}^2 \quad \text{for all } v \in \dot{\mathbb{H}}^1. \tag{A.38}$$

For any test function $u \in \dot{\mathbb{H}}^1$, equation (A.38) leads to the bound

$$\langle u,\, f\rangle_{\mathbb{L}^2} \leq \|f\|_{\mathbb{L}^2} \cdot \|u\|_{\mathbb{L}^2} \leq \frac{1}{\rho_P}\|f\|_{\mathbb{L}^2} \cdot \|u\|_{\dot{\mathbb{H}}^1}.$$

So $\langle \cdot,\, f\rangle_{\mathbb{L}^2}$ is a bounded linear functional on $\dot{\mathbb{H}}^1$. Again, by the Riesz representation theorem, there exists a unique $g \in \dot{\mathbb{H}}^1$ such that $\langle u,\, f\rangle_{\mathbb{L}^2} = \langle u,\, g\rangle_{\dot{\mathbb{H}}^1}$ for all $u \in \dot{\mathbb{H}}^1$, which completes the proof.

Proof of equation (A.37c)**:** For any test function $u \in \dot{\mathbb{H}}^1$, we note that

$$\langle u,\, \widetilde{A}u\rangle_{\dot{\mathbb{H}}^1} = \langle u,\, Au\rangle_{\mathbb{L}^2} = \int_\Omega \big(\nabla u(x)\big)^\top a(x)\big(\nabla u(x)\big)dx.$$

From our uniform ellipticity condition, we know that $\mu I_m \preceq a(x) \preceq \beta I_m$ for any $x \in \Omega$. Substituting this relation yields $\mu\|u\|_{\dot{\mathbb{H}}^1}^2 \leq \langle u,\, \widetilde{A}u\rangle_{\dot{\mathbb{H}}^1} \leq \beta\|u\|_{\dot{\mathbb{H}}^1}^2$, as claimed.

A.8.3.2 Proof of Corollary 2.4

The matrices M, Σ_L, Σ_b for the projected problem instances can be obtained by straightforward calculation. In order to apply Theorem 2.1, it remains to verify the assumptions.

By Lemma A.9, the operator L is self-adjoint in $\mathbb{X}$, and is sandwiched as $0 \leq \langle u,\, Lu\rangle_{\dot{\mathbb{H}}^1} \leq \big(1 - \frac{\mu}{\beta}\big)\|u\|_{\dot{\mathbb{H}}^1}^2$ for all $u \in \mathbb{X}$. This yields the operator norm bound $\|\!|L|\!\|_{\mathbb{X}} \leq 1 - \frac{\mu}{\beta}$.

Now we verify the conditions in Assumption 2.1(W). For any basis function ϕ_j with $j \in [d]$ and vector $v \in \dot{\mathbb{H}}^1$, we have

$$\begin{aligned}
&\mathbb{E}\langle \phi_j,\, (L_i - L)u\rangle_{\dot{\mathbb{H}}^1}^2\\
&\leq \frac{1}{\beta^2}\mathbb{E}\Big(\int_\Omega \delta_{x_i}\nabla\phi_j(x)^\top a(x_i)\nabla u(x)dx\Big)^2 + \frac{1}{\beta^2}\mathbb{E}\Big(\int_\Omega \delta_{x_i}\nabla\phi_j(x)^\top W_i\nabla u(x)dx\Big)^2\\
&= \frac{1}{\beta^2}\int_\Omega \Big(\nabla\phi_j(x)^\top a(x)\nabla u(x)\Big)^2 dx + \frac{1}{\beta^2}\int_\Omega \mathbb{E}\Big(\nabla\phi_j(x)^\top W_i\nabla u(x)\Big)^2 dx\\
&\leq \frac{1}{\beta^2}\int_\Omega \|\nabla\phi_j(x)\|_2^2 \|\!|a(x)|\!\|_{\mathrm{op}}^2 \|\nabla u(x)\|_2^2 dx + \frac{2}{\beta^2}\int_\Omega \|\nabla\phi_j(x)\|_2^2 \cdot \|\nabla u(x)\|_2^2 dx\\
&\leq \Big(1 + \frac{2}{\beta^2}\Big)\max_{j\in[d]}\sup_{x\in\Omega}\|\nabla\phi_j\|_2^2 \cdot \int_\Omega \|\nabla u(x)\|_2^2 dx\\
&\leq \sigma_L^2\|u\|_{\dot{\mathbb{H}}^1}^2,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\langle \phi_j,\, b_i - b\rangle_{\mathbb{H}^1}^2 &\leq \frac{1}{\beta^2}\mathbb{E}\Big(\int_\Omega \delta_{y_i}\phi_j(y)f(y_i)dy\Big)^2 + \frac{1}{\beta^2}\mathbb{E}\Big(\int_\Omega \delta_{y_i}\phi_j(y)g_i dy\Big)^2\\
&= \frac{1}{\beta^2}\int_\Omega \phi_j(y)^2 f(y)^2 dy + \frac{1}{\beta^2}\int_\Omega \phi_j(y)^2 \mathbb{E}[g_i^2]dy\\
&\leq \frac{1}{\beta^2}\sup_{x\in\Omega}|\phi_j|^2 \int_\Omega \big(f(y)^2+1\big)dy\\
&= \frac{\|f\|_{\mathbb{L}^2}^2+1}{\beta^2}\max_{j\in[d]}\sup_{x\in\Omega}|\phi_j|^2.
\end{aligned}$$

Therefore, Assumption 2.1(W) is satisfied with constants (σ_L, σ_b). Invoking Theorem 2.1 completes the proof.

A.9 Additional simulation studies

In this appendix, we present additional details of simulation related to the optimal oracle inequalities. We first describe the simulation setup in the simulation studies shown in Figure 2.1, and then present simulation results related to the statistical error term $\mathcal{E}_n(M, \Sigma^*)$.

A.9.1 Models underlying simulations in Figure 2.1

The simulation results shown in Figure 2.1 are generated by constructing random transition matrices based on the following random graph models:

Erdös-Rényi random graph: Given $d, N \in \mathbb{N}_+$ and $a > 1$, we consider the following sampling procedure. Let G be an Erdös-Rényi random graph with N vertices and edge probability $p = \frac{a}{N}$, and take $\widetilde{G}$ to be its largest connected component. (When $c > 1$, the number of vertices in $\widetilde{G}$ is of order $\Theta(N)$. See the monograph [57] for details.) For each vertex $v \in V(\widetilde{G})$, we associate it with an independent standard Gaussian random vector $\phi_v \sim \mathcal{N}(0, I_d)$. Let $V(\widetilde{G})$ be the state space and let the Markov transition kernel P be the simple random walk on $\widetilde{G}$.

In Figure 2.1 (a), we take the number of vertices to be $N = 3000$ and the feature dimension to be $d = 1000$. The edge density parameter is chosen as $a = 3$. The resulting giant connected component contains 2813 vertices.

Random geometric graph: Given a pair of positive integers (d_0, N) and scalar $r > 0$, we consider the following sampling procedure. For each vertex $i \in [N]$, we associate it with an independent standard Gaussian random vector $x_i \sim \mathcal{N}(0, I_d)$. The graph G is then constructed such that $(i, j) \in E(G)$ if and only if $\|x_i - x_j\|_2 \leq r$. (See the monograph [166] for more details of this random graph model.) Take $\widetilde{G}$

to be the largest connected component of G. We take $V(\widetilde{G})$ as the state space, and for each $v \in V(\widetilde{G})$, we let the feature vector be $\phi_v = x_v$. Finally, we let P be the simple random walk on $\widetilde{G}$.

In Figure 2.1 (b), we take the number of vertices to be $N = 3000$ and the feature dimension to be $d_0 = 2$. The distance threshold is chosen as $r = 0.1$. The resulting giant connected component contains 2338 vertices.

Despite their simplicity, the two random graph models capture distinct types of the behavior of the resulting random walk in feature space: in the former model, the transition kernel makes "big jumps" in the feature space, and the correlation between two consecutive states is small; in the latter model, the transition kernel makes "local moves" in the feature space, leading to large correlation. This is reflected by the fact that the parameter $\kappa(M)$ in Lemma 2.1 is smaller in the former case and larger in the latter ease.

A.9.2 Simulation results on the statistical error $\mathcal{E}_n(M, \Sigma^*)$

We also conduct simulation studies on the statistical error $\mathcal{E}_n(M, \Sigma^*)$ in Theorems 2.1 and 2.3. We consider the averaged stochastic approximation procedures applied to policy evaluation with linear function approximation, resulting in the TD(0) algorithm discussed in Section 2.4.3 . As in the approximation factor studies, the simulations for statistical error are based on Markov reward processes defined by random walks over two types of random graphs: the Erdös-Rényi graphs and the geometric random graphs; see Section A.9.1 for details. We make the following additional setup:

- Let d be the dimension of feature vectors ϕ_v. Given a vector $\beta_* \in \mathbb{R}^d$ and a positive scalar h, we sample the reward function r as follows:

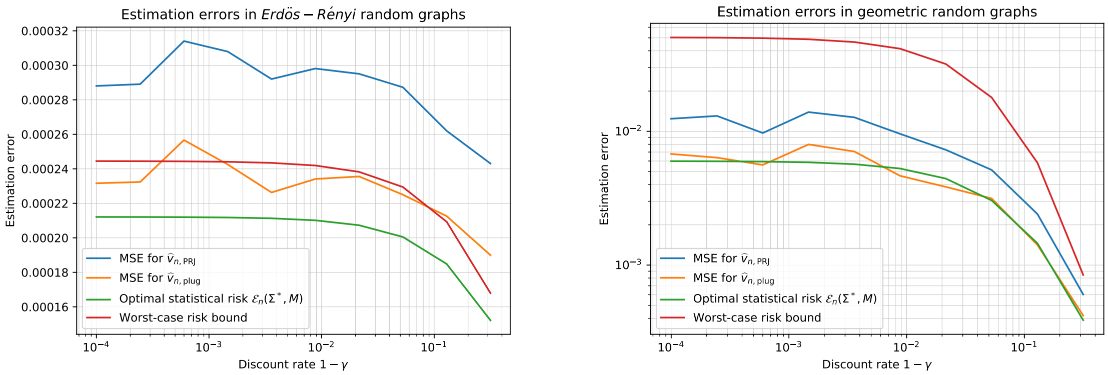
$$r(v) \sim \mathcal{N}(\phi_v^\top \beta_*, h^2), \quad \text{independently for each } v \in V(\widetilde{G}),$$

  and the noisy reward is further given by $R(s) \sim \mathcal{N}(r(s), \sigma^2)$, for some $\sigma > 0$.

- For the geometric random graph, we augment the feature vectors using low-dimensional polynomials. Concretely, instead of taking the feature vector ϕ_v to be the vector x_v itself, for any integer $k > 0$, we let:

$$\phi_v := \Big[\prod_{j=1}^{d_0} x_v^{\alpha_j} \Big]_{\alpha \in \mathbb{N}^d, \ \|\alpha\|_1 \leq k}.$$

  Note that due to the "curse-of-dimensionality", under our setup, the locality behavior of geometric random graph is relevant only if the dimension of the feature vectors is low. Such augmentation in the feature space turns the stochastic approximation problem into a high-dimensional one, thereby rendering it a non-trivial problem.



(a) (b)

Figure A.2: Plots of empirical MSE and theoretical bounds as a function of the discount factor γ in the policy evaluation problem. (See the text for a discussion.) (a) Results for an Erdös-Rényi random graph model with $N = 3000$, projected dimension $d = 100$, and $a = 30$. The resulting number of vertices in the graph $\widetilde{G}$ is 3000. The value of $1-\gamma$ is plotted in log-scale, and the value of approximation factor is plotted on the standard scale. (b) Results for a random geometric graph model with $N = 3000$, projected dimension $d_0 = 3$, $k = 4$, and $r = 1$. The resulting number of vertices in the graph $\widetilde{G}$ is 2991 and the resulting feature dimension is $d = 31$. Both the discount rate $1-\gamma$ and the approximation factor are plotted on the log-scale. The fluctuations in the plot are due to randomness of the estimators.

In our simulation, we generate the Markov reward process as above, on which we run the Polyak–Ruppert averaged stochastic approximation procedure (2.41) and obtain the value function estimator $\widehat{v}_{n,\mathrm{PRJ}}$. In our simulation, we take stepsize $\eta = 1/\sqrt{dn}$ and burn-in period $n_0 = n/5$.

We also consider the plug-in solution $\widehat{v}_{n,\mathrm{plug}}$ defined as:

$$\widehat{\vartheta}_{n,\mathrm{plug}} := \Big(\sum_{t=1}^{n} \big(\psi(s_t)\psi(s_t)^\top - \gamma \psi(s_t)\psi(s_t^+)^\top \big) \Big)^{-1} \cdot \Big(\sum_{t=1}^{n} R_t \psi(s_t) \Big), \quad \text{and}$$

$$\widehat{v}_{n,\mathrm{plug}} := \widehat{\vartheta}_{n,\mathrm{plug}}^\top \psi.$$

To illustrate how well the theoretical predictions match the actual performance, we measure their error of estimating the projected fixed-point $\bar{v}$. In particular, we empirically evaluate the MSE for the estimators $\widehat{v}_{n,\mathrm{PRJ}}$ and $\widehat{v}_{n,\mathrm{plug}}$ by averaging the squared error $\|\widehat{v}_n - \bar{v}\|^2$ over m independent experiments. We also compute the risk functional $\mathcal{E}_n(M, \Sigma^*)$ for each problem instance, as well as its worst-case upper bound $r_{n,\mathrm{worst}} := n^{-1}\mathrm{Tr}(\Sigma^*)/(1-\kappa(M))^2$.

In Figure A.2, we demonstrate the simulation results of empirical MSE and theoretical risk functionals for above problem instances, as a function of the discount rate $1-\gamma$.

The discount rate ranges from 10^{-4} to $10^{-0.5}$, and the reward model described above is generated by setting $\beta_* \sim \mathcal{N}(0, I_d/d)$, $h = 0.5$ and $\sigma = 1$. We take the number of trials m to be 10.

It can be seen from Figure 2.1 that, under both models, the theoretical bound $\mathcal{E}_n(M, \Sigma^*)$ is matched well by both the plug-in estimator $\widehat{v}_{n,\mathrm{plug}}$ and the estimator $\widehat{v}_{n,\mathrm{PRJ}}$ produced by Polyak-Ruppert averaging, while the latter is off by a constant factor. This is because the burn-in period discards a constant fraction of initial iterates. In both cases, the statistical risks increases with discount rate getting close to 0. However, the variation is within a small range (by factor of 2) for Erdös–Rényi graphs, while being significant (by a factor of 20) in geometric random graphs. This phenomenon is similar to that of approximation factor, as both of them are related to the one-step correlation matrix M — with a Markov chain having higher one-step correlation in the feature space, the matrix M becomes closer to the matrix γI_d, making both the approximation factor and the statistical rate large for discount factors close to 1. Finally, we note that the worst-case statistical rate $r_{n,\mathrm{worst}}$ existing in many previous analysis (see e.g. [115]) is an accurate prediction of the actual performance in the Erdös–Rényi, while being off by a factor of 10 in the geometric random graph case. This illustrates the value of establishing instance-dependent bounds for the statistical error in these problems.

Appendix B

Proofs and discussion deferred from Chapter 3

B.1 Additional related work

Chapter 3 analyzes stochastic approximation algorithms based on Markov data, and has consequences for reinforcement learning. So as to put our results into context, we now provide more background on past work in these areas.

B.1.1 Statistical estimation based on Markov data

There is a large body of past work on statistical estimation based on observing a single trajectory of a Markov chain; for example, see [18] for an overview of some classical results. For the problem of functional estimation under the stationary distribution, the asymptotic efficiency of plug-in estimators[1] has been established for discrete-state Markov chains [165, 70] and Itô diffusion processes [113]. In Chapter 3, we provide non-asymptotic bounds, both upper and lower, that depend on a certain instance-dependent functional that also appears in an asymptotic analysis. More recent work has seen non-asymptotic results for statistical estimation with Markovian data, including the estimation of transition kernels [220, 126], mixing times [81], the parameters of Gaussian hidden Markov models [223], as well for certain testing problems [47]. These papers can be roughly divided into two categories. Papers in the first category focus on estimating parameters for each individual state of the Markov chain (e.g., transition kernels), and thus require sample sizes that scale with the complexity of the state space (e.g., its cardinality in the discrete case). By contrast, papers in the second category are concerned with estimating properties of the Markov chain (e.g., the expectation of a functional under the stationary distribution), and the sample complexity of such problems need not depend on the size of the state space. Our results in Chapter 3 falls within the second category.

[1] These papers refer to such methods as “empirical” estimators.

B.1.2 Stochastic approximation methods

The use of recursive stochastic procedures for solving fixed point equations dates back to the seminal work of [174]; see the reference books [20, 11, 111] for more background. By averaging the iterates of the SA procedure, it is known that one can obtain both an improved convergence rate and central limit behavior [169, 186]. A variety of stochastic approximation procedures now serve as the workhorse for modern large-scale machine learning and statistical inference [161, 21], and many algorithmic techniques are known to accelerate their convergence [67, 84, 121]. In particular, non-asymptotic bounds matching the optimal Gaussian limit have been established in a variety of settings [157, 64, 50, 150].

While the instance-dependent nature of this line of investigation aligns with the objective of our work, prior work either assumes an i.i.d. observation model or imposes a martingale difference assumption on the noise.[2] The first study of SA procedures without a martingale difference assumption was initiated by [112], who give a general criteria for convergence, as well as [127, 128], who analyzed linear problems motivated by control and filtering. [142] analyzed general SA problems for controlled Markov processes by applying the Kushner–Clark lemma. In addition to this classical work, stochastic approximation in the Markov setting has attracted much recent attention. [37] provides finite-sample error bounds on the averaged iterate of Markovian linear stochastic approximation, with an optimal leading-order term. Central limit theorems [61] and non-asymptotic convergence rates [93] have been established for controlled Markov processes. In addition to the papers discussed in Section 3.1, several recent works have considered particular aspects of SA with Markov data, including two-time-scale variants [52, 94], observation skipping schemes for bias reduction [108], Lyapunov function-based analysis under general norms [40], and proving guarantees under weaker ergodicity conditions [49].

B.1.3 Application to RL problems

Markovian observations arise naturally in the context of stochastic control and reinforcement learning (RL). See [11] for a historical survey of algorithms for stochastic control and filtering with Markovian stochastic approximation, and the books [14, 200] for more background on the RL setting. In RL problems, SA algorithms are typically used to solve Bellman equations, a class of linear or non-linear fixed-point equations. In policy evaluation problems, temporal difference (TD) methods [198] use linear stochastic approximation to estimate the value function of a given policy, with asymptotic convergence guarantees [48, 204, 24] and non-asymptotic bounds [16, 98]. In the non-linear case, the Q-learning algorithm [216] is a stochastic approximation method that estimates the Q-function of a Markov decision process from data. There is a long line of past work on this algorithm, including convergence guarantees [203, 201, 59], results on linear function approximation for optimal stopping problems [205, 16], and non-asymptotic rates under

[2] In the linear equation setup, the martingale difference noise assumes that $\mathbb{E}[L_{t+1} \mid \mathcal{F}_t] = \bar{L}$ and $\mathbb{E}[b_{t+1} \mid \mathcal{F}_t] = \bar{b}$, which does not cover the Markov case.

general norms in both the i.i.d. setting [211, 19] as well as the Markovian setting [40]. A class of variants of TD and Q-learning are also studied in literature, including actor-critic methods [107], SARSA [185], and methods that employ variance-reduction [190, 98, 212, 99]. A concurrent preprint to this manuscript [125] proves lower bounds on the oracle complexity of policy evaluation with access to temporal difference operators, and develops an acceleration scheme with variance reduction to achieve these lower bounds while retaining the optimal sample complexity.

It should be noted that an important feature of reinforcement learning is function approximation, i.e., using a given function class (e.g. a linear subspace) to approximate the solution to the Bellman equation of interest. This method enables estimation with a sample size depending on the intrinsic complexity of the function class, instead of the cardinality of state-action space. On the other hand, an approximation error is induced by projecting the Bellman equation onto this function class. This trade-off is central to the class of TD algorithms, as studied in a line of past work [204, 226, 15, 158]. Chapter 2 of this dissertation focuses on the i.i.d. setting, and shows that projected linear equations have a non-standard tradeoff between approximation and estimation errors. The results in Chapter 3 is complementary in nature, building on this work by analyzing the more challenging setting of Markov observations. Among the concrete consequences of Chapter 3 are an instance-optimal analysis of TD algorithms in the Markov setting with linear function approximation. This analysis provides the basis for a principled choice of the parameter λ in the broader class of TD(λ) algorithms.

B.2 Auxiliary truncation results

In this section, we present two auxiliary results on the relations between assumptions 3.2, 3.3, and 3.4. These results are based on truncation arguments.

B.2.1 Assumption 3.2 (almost) implies assumption 3.4 under discrete metric

For the discrete metric $\rho(x, y) := \mathbf{1}_{x \neq y}$, the Lipschitz assumption 3.4 is equivalent to the following uniform upper bounds:

$$\|\boldsymbol{L}_{t+1}(s_t) - \bar{L}\|_{\mathrm{op}} \leq \sigma_L d \quad \text{and} \quad \|\boldsymbol{b}_{t+1}(s_t) - \bar{b}\|_2 \leq \sigma_b \sqrt{d}.$$

The following proposition provides uniform high-probability upper bounds on such quantities based on the moment assumption:

Proposition B.1. *Under Assumption 3.2 with $\bar{p} = +\infty$, there exists a universal constant $c > 0$, such that for any $\delta > 0$, the following bounds hold true uniformly over $t = 1, 2, \cdots, n$, with probability $1 - \delta$:*

$$\|\boldsymbol{L}_{t+1}(s_t) - \bar{L}\|_{op} \leq cd \cdot \sigma_L \log \frac{nd}{\delta} \quad \textit{and} \quad \|\boldsymbol{b}_{t+1}(s_t) - \bar{b}\|_2 \leq c\sqrt{d} \cdot \sigma_b \log \frac{nd}{\delta}. \tag{B.1}$$

We prove this proposition at the end of this section.

When the random observations (L_{t+1}, b_{t+1}) are not almost-surely bounded, but satisfies the moment assumption 3.2 with $\bar{p} = +\infty$, we can apply our theorems on the event that Eq (B.1) holds true, and the main theorems hold true conditionally on such an event, with constants (σ_L, σ_b) inflated with a factor $\log(nd/\delta)$.

Proof of Proposition B.1: For a given $t \in [n]$, we note that:

$$\|L_{t+1} - \bar{L}\|_{\mathrm{op}}^2 \leq \|L_{t+1} - \bar{L}\|_F^2 = \sum_{j,\ell=1}^{d} \left[e_j^\top \left(L_{t+1} - \bar{L}\right) e_\ell \right]^2.$$

For each pair $j, \ell \in [d]$, Assumption 3.2 implies that:

$$\mathbb{P}\left(\left| e_j^\top \left(L_{t+1}(s_t) - \bar{L} \right) e_\ell \right| \geq c\sigma_L \log(nd/\delta) \right) \leq \frac{\delta}{2d^2 n}$$

Taking union bound over all the coordinate pairs (j, ℓ) and substituting into above expansion, we have that:

$$\mathbb{P}\left(\|L_{t+1} - \bar{L}\|_{\mathrm{op}} \geq cd \cdot \sigma_L \log(nd/\delta) \right) \leq \delta/(2n).$$

Similarly, for the vector-valued observations b_{t+1}, we have the following bounds with probability $1 - \delta/n$:

$$\|b_{t+1} - \bar{b}\|_2^2 \leq \sum_{j=1}^{d} \left(e_j^\top (b_{t+1} - \bar{b}) \right)^2 \leq c\sigma_b^2 d \cdot \log^2(nd/\delta).$$

Taking union bound over $t = 1, 2, \cdots, n$, we complete the proof of this proposition.

B.2.2 On the stationary tail and boundedness assumption 3.3

Note that in many applications, the Markov chain $(s_t)_{t \geq 0}$ lives in an unbounded state space. However, as long as the stationary distribution ξ of P is sufficiently light-tailed, a simple truncation argument applies, which we illustrate for completeness. Concretely, suppose that there exists a constant $\sigma_\rho > 0$, such that the following bound holds true for any $p \geq 2$:

$$\mathbb{E}_{s \sim \xi}\left[\rho(s, s_0)^p \right] \leq p! \cdot \sigma_\rho^p. \tag{B.2}$$

Given a stationary Markovian trajectory $\{s_t\}_{t=1}^n$, consider the event

$$\mathscr{E}_{n,\delta} = \left\{ \forall t \in [1, n], \rho(s_0, s_t) \leq 2\sigma_\rho \log \tfrac{n}{\delta} \right\}.$$

By the tail assumption (B.2) and a union bound, it directly follows that $\mathbb{P}\big(\mathscr{E}_{n,\delta}\big) \geq 1 - \delta$. Consider a truncated Markov transition kernel P' defined as

$$P'(x, Z) := P\big(x, Z \cap \mathbb{B}\big(0, 2\sigma_\rho \log(n/\delta)\big)\big) + P\big(x, \mathbb{B}\big(0, 2\sigma_\rho \log(n/\delta)\big)^c\big) \mathbf{1}_{s_0 \in Z},$$

for $x \in \mathcal{S}$ and $Z \subseteq \mathcal{S}$.

In words, the Markov chain P' attempts to make the transition from s_t to s_{t+1} according to the transition kernel P'. If the state s_{t+1} lies in the ball $\mathbb{B}\big(0, 2\sigma_\rho \log(n/\delta)\big)^c$, we keep it as is; otherwise, we let the next-step transition be deterministically s_0.

Given a trajectory $\{s'_t\}_{t=1}^n$ of the Markov chain P', there exists a coupling such that

$$\mathbb{P}\big(\{s_t\}_{t=1}^n \neq \{s'_t\}_{t=1}^n\big) \leq \mathbb{P}\big(\mathscr{E}_{n,\delta}^c\big) \leq \delta.$$

One can then proceed by working on the high probability event $\mathscr{E}_{n,\delta}$, where the Markov chain has a effective diameter of $O\big(\sigma_\rho \log \frac{n}{\delta}\big)$.

B.3 Auxiliary results underlying Proposition 3.1

This appendix is devoted to the proofs of auxiliary lemmas that are used in the proof of Proposition 3.1.

B.3.1 Proof of Lemma 3.4

Throughout the proof, we let $x \in \mathcal{S}$ be an arbitrary but fixed state. Note that any positive integer τ can be represented as $\tau = kt_{\text{mix}} + q$ with $k \in \mathbb{N}_+$ and $0 \leq q \leq t_{\text{mix}} - 1$. We show the desired claim by induction over $k \geq 0$.

Base case: When $k = 0$, Assumption 3.3 implies that

$$\mathcal{W}_{1,\rho}(\delta_x P^\tau, \xi) \leq \sup_{s,s'} \rho(s, s') \leq 1 \leq c_0,$$

so that the base case ($k = 0$) holds for our induction proof.

Induction step: At step k of the argument, the induction hypothesis ensures that

$$\mathcal{W}_{1,\rho}\big(\delta_x P^{kt_{\text{mix}}+q}, \xi\big) \leq c_0 \cdot 2^{-k}, \quad \text{for } q = 0, 1, \cdots, t_{\text{mix}} - 1. \tag{B.3}$$

We now need to show that the result holds for any $\tau = (k+1)t_{\text{mix}} + q$, where $q \in \{0, 1, \ldots, t_{\text{mix}} - 1\}$ is arbitrary. We do so via a coupling argument. Take a random initial state $y \sim \xi$, and consider two processes $\{s_t\}_{t\geq 0}$ and $\{s'_t\}_{t\geq 0}$ starting from x and y, respectively. Their joint distribution is defined as follows: choose the coupling between the law of $s_{kt_{\text{mix}}+q}$ and $s'_{kt_{\text{mix}}+q}$ to satisfy the identity $\mathbb{E}\big[\rho(s_{kt_{\text{mix}}+q}, s'_{kt_{\text{mix}}+q})\big] = \mathcal{W}_{1,\rho}\big(\delta_x P^{kt_{\text{mix}}+q}, \xi\big)$. Conditionally on $(s_{kt_{\text{mix}}+q}, s'_{kt_{\text{mix}}+q})$, Assumption 3.1 guarantees the existence of a coupling between $\delta_{s_{kt_{\text{mix}}+q}} P^{t_{\text{mix}}}$ and $s'_{kt_{\text{mix}}+q} P^{t_{\text{mix}}}$ such that

$$\mathbb{E}\big[\rho\big(s_{(k+1)t_{\text{mix}}+q}, s'_{(k+1)t_{\text{mix}}+q}\big) \mid (s_{kt_{\text{mix}}+q}, s'_{kt_{\text{mix}}+q})\big] \leq \tfrac{1}{2}\rho(s_{kt_{\text{mix}}+q}, s'_{kt_{\text{mix}}+q}).$$

Taking expectation on both sides and substituting with equation (B.3), we find that

$$\mathcal{W}_{1,\rho}\big(\delta_x P^{(k+1)t_{\text{mix}}+q}, \xi\big) \leq \mathbb{E}\big[\rho\big(s_{(k+1)t_{\text{mix}}+q}, s'_{(k+1)t_{\text{mix}}+q}\big)\big] \leq c_0 \cdot 2^{-(k+1)},$$

which completes the proof of the induction step.

B.3.2 Proof of Lemma 3.5

Our proof is based on the following intermediate claim

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \leq e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big). \tag{B.4}$$

This bound, which we return to prove at the end of this section, is a weaker form of the claim in the lemma.

We now use the bound (B.4) to prove the lemma. Applying Minkowski's inequality to the recursive relation (3.46), we find that for any $p \geq 2$, the p^{th} moment is upper bounded as

$$\begin{aligned}&\big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p}\\ &\leq \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}.\end{aligned}$$

For the martingale part of the noise, we take the decomposition $L_{t+\ell+1} = \boldsymbol{L}(s_{t+\ell}) + Z_{t+\ell+1}$. By Assumption 3.2 and Hölder's inequality, we have the bounds

$$\mathbb{E}\big[\|Z_{t+\ell+1}\Delta_{t+\ell}\|_2^p \mid \mathcal{F}_t\big] \leq d^{\frac{p}{2}}\sum_{j=1}^d \mathbb{E}\big[\langle e_j,\, Z_{t+\ell+1}\Delta_{t+\ell}\rangle^p \mid \mathcal{F}_t\big] \leq \big(p\sigma_L\sqrt{d}\big)^p\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p \mid \mathcal{F}_t\big],$$

$$\mathbb{E}\big[\|\zeta_{t+\ell+1}\|_2^p \mid \mathcal{F}_t\big] \leq d^{\frac{p}{2}}\sum_{j=1}^d \mathbb{E}\big[\langle e_j,\, \zeta_{t+\ell+1}\rangle^p \mid \mathcal{F}_t\big] \leq (p\sqrt{d})^p \cdot \big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big)^p.$$

Similarly, for the Markov part of the noise, we have:

$$\mathbb{E}\big[\|\nu_{t+\ell+1}\|_2^p\big] \leq (p\sqrt{d})^p \cdot \big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big)^p.$$

On the other hand, the Lipschitz condition (3.4) and the boundedness condition (3.3) of the metric space imply that

$$\|L_{t+\ell+1}(s) - \bar{L}\|_{\mathrm{op}} \leq \sigma_L d, \quad \text{and} \quad \|\boldsymbol{b}(s) - \bar{b}\|_2 \leq \sigma_b\sqrt{d} \quad \text{for all } s \in \mathcal{S}.$$

Substituting into the decomposition above, we arrive at the bounds

$$\begin{aligned}\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} &\leq (\gamma_{\max} + \sigma_L p\sqrt{d} + \sigma_L d)\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p}, \quad \text{and}\\ \big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p} &\leq 2p\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big).\end{aligned}$$

Applying equation (B.4) yields

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} &+ e\eta(\gamma_{\max} + \sigma_L d)\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p}\\ &+ 2(1 + 6\eta\ell)\eta p\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big).\end{aligned}$$

Solving this recursion leads to the bound

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \leq e\eta\ell(\gamma_{\max} + \sigma_L d)\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 3\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big),$$

which establishes the first claim.

Since the stepsize is upper bounded as $\eta \leq \big(2e\eta\ell(\gamma_{\max} + \sigma_L d)\big)^{-1}$, we have the lower bound

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} &\geq \big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} - \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} \\ &\geq \tfrac{1}{2}\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} - 3\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big),\end{aligned}$$

which, in conjunction with the bound (B.4), establishes the second claim.

Proof of equation (B.4): Applying Minkowski's inequality to the recursive relation (3.46) yields (for any $p \geq 2$) a bound on the p^{th} conditional moment:

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell+1}\|_2^p\big]\big)^{1/p} \leq \big(\mathbb{E}\big[\|(I - \eta L_{t+\ell+1})\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}. \tag{B.5}$$

Our next step is to bound the two terms above.

Substituting into the recursive relation (B.5), and applying Minkowski's inequality, we find that the moment $\big(\mathbb{E}\big[\|\Delta_{t+\ell+1}\|_2^p\big]\big)^{1/p}$ is upper bounded by

$$(1 + \eta\gamma_{\max})\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + \eta\sigma_L d\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} + 2\eta p\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big).$$

Solving this recursive inequality leads to

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \leq \exp\big(\eta\ell(\gamma_{\max} + \sigma_L d)\big)\Big(\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 2\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big)\Big).$$

For any stepsize $\eta \in \big(0, \frac{1}{(\gamma_{\max}+\sigma_L d)\ell}\big]$, we have

$$\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \leq e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big),$$

which establishes the claim.

B.3.3 Proof of Lemma 3.6

For notational simplicity, we extend the process $(\Delta_t)_{t\geq 0}$ to the entire set $\mathbb{Z}$ of integers, in particular by defining $\Delta_t := \Delta_0$ for negative integer t. Note that under our assumption, Lemma 3.5 and the assumed bound (3.62) both hold true for the extended process, with index set $t \in \mathbb{Z}$. Moreover, as in the proof of Lemma 3.5, for each $p \geq 2$, we have the moment bound

$$\begin{aligned}\big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p} \leq \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p} + \eta\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \\ + \eta\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p}.\end{aligned}$$

Our next step is to exploit the coarse bound (3.62) so as to obtain upper bounds on the second term $\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p}$. Given the time lag $\tau > 0$, we take the decomposition $\Delta_{t+\ell} = \Delta_{t+\ell-\tau} + (\Delta_{t+\ell} - \Delta_{t+\ell-\tau})$, and by Minkowski's inequality, we have that

$$\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \leq \big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p\big]\big)^{1/p} + \big(\mathbb{E}\big[\|L_{t+\ell+1}(\Delta_{t+\ell} - \Delta_{t+\ell-\tau})\|_2^p\big]\big)^{1/p}. \tag{B.6}$$

The latter term of the bound (B.6) can be controlled through Assumption 3.4:

$$\|L_{t+\ell+1}(s_{t+\ell})(\Delta_{t+\ell} - \Delta_{t+\ell-\tau})\|_2 \leq (\gamma_{\max} + \sigma_L d)\|\Delta_{t+\ell} - \Delta_{t+\ell-\tau}\|_2, \quad \text{a.s.}$$

The distance $\|\Delta_{t+\ell} - \Delta_{t+\ell-\tau}\|_2$ is controlled via the coarse bound (3.62). Putting together the pieces, we find that

$$\big(\mathbb{E}\big[\|L_{t+\ell+1}(\Delta_{t+\ell} - \Delta_{t+\ell-\tau})\|_2^p\big]\big)^{1/p} \leq \eta\big(\gamma_{\max} + \sigma_L d\big) \cdot \big(\omega_p\big(\mathbb{E}\big[\|\Delta_{t+\ell-\tau}\|_2^p\big]\big)^{1/p} + \beta_p\bar{\sigma}\big). \tag{B.7}$$

In order to bound the former term $\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p\big]\big)^{1/p}$ in the bound (B.6), we invoke Lemma 3.4, and obtain a random variable $\widetilde{s}_{t+\ell}$, such that

$$\widetilde{s}_{t+\ell} \mid \mathcal{F}_{t+\ell-\tau} \sim \xi, \quad \text{and} \quad \big(\mathbb{E}\big[\rho(s_{t+\ell}, \widetilde{s}_{t+\ell-\tau})^p \mid \mathcal{F}_{t+\ell-\tau}\big]\big)^{1/p} \leq c_0 \cdot 2^{1-\frac{\tau}{2t_{\mathrm{mix}}p}}. \tag{B.8}$$

By Assumption 3.2, we have the bounds

$$\mathbb{E}\big[\|Z_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] \leq (p\sqrt{d}\sigma_L)^p\|\Delta_{t+\ell-\tau}\|_2^p, \quad \text{and} \tag{B.9a}$$

$$\mathbb{E}\big[\|\big(\boldsymbol{L}(\widetilde{s}_{t+\ell-\tau}) - \bar{L}\big) \cdot \Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] \leq (p\sqrt{d}\sigma_L)^p\|\Delta_{t+\ell-\tau}\|_2^p. \tag{B.9b}$$

Invoking the moment bound (B.8) and using the Lipschitz condition (3.4), we find that

$$\begin{aligned}
&\mathbb{E}\big[\|\big(\boldsymbol{L}(\widetilde{s}_{t+\ell-\tau}) - \boldsymbol{L}(s_{t+\ell-\tau})\big) \cdot \Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big] \\
&\leq \mathbb{E}\big[|\!|\!|\boldsymbol{L}(\widetilde{s}_{t+\ell-\tau}) - \boldsymbol{L}(s_{t+\ell-\tau})|\!|\!|_{\mathrm{op}}^p \mid \mathcal{F}_{t+\ell-\tau}\big] \cdot \|\Delta_{t+\ell-\tau}\|_2^p \\
&\leq \big(\sigma_L c_0 d \cdot 2^{1-\frac{\tau}{2t_{\mathrm{mix}}p}}\|\Delta_{t+\ell-\tau}\|_2\big)^p.
\end{aligned} \tag{B.9c}$$

Finally, we have the operator norm bound

$$\|\bar{L}\Delta_{t+\ell-\tau}\|_2 \leq \gamma_{\max}\|\Delta_{t+\ell-\tau}\|_2. \tag{B.9d}$$

Collecting the results from equations (B.9)(a)—(d), we arrive at the bound

$$\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell-\tau}\|_2^p \mid \mathcal{F}_{t+\ell-\tau}\big]\big)^{1/p} \leq \big(2p\sqrt{d}\sigma_L + \gamma_{\max} + \sigma_L c_0 d \cdot 2^{1-\frac{\tau}{2t_{\mathrm{mix}}p}}\big)\|\Delta_{t+\ell-\tau}\|_2. \tag{B.10}$$

According to Lemma 3.5, given a stepsize bounded as $\eta \le \big(6(\gamma_{\max} + \sigma_L d)\tau\big)^{-1}$, we have

$$\big(\mathbb{E}\|\Delta_{t+\ell-\tau}\|_2^p\big)^{1/p} \le 2\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p} + 12\eta p\tau\sqrt{d}\big(\sigma_L\|\bar{\theta}\|_2 + \sigma_b\big).$$

Collecting the bounds (B.7) and (B.10), and substituting into the decomposition (B.6), for $\tau \ge 2t_{\text{mix}} p \log(c_0 d)$, we arrive at the inequality:

$$\begin{aligned}&\big(\mathbb{E}\big[\|L_{t+\ell+1}\Delta_{t+\ell}\|_2^p\big]\big)^{1/p}\\ &\qquad \le 2\Big(\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big) + \eta\omega_p\big(\gamma_{\max} + \sigma_L d\big)\Big) \cdot \Big(\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p} + \eta p\tau\sqrt{d}\bar{\sigma}\Big)\\ &\qquad\qquad + \eta\big(\gamma_{\max} + \sigma_L d\big)\beta_p\bar{\sigma}.\end{aligned}$$

By following the derivation in the proof of Lemma 3.5, we can show that the third term is upper bounded as

$$\big(\mathbb{E}\big[\|\nu_{t+\ell} + \zeta_{t+\ell+1}\|_2^p\big]\big)^{1/p} \le 2p\sqrt{d}(\sigma_L\|\bar{\theta}\|_2 + \sigma_b).$$

Substituting back into the original decomposition, we find that the difference in moments $D := \big(\mathbb{E}\big[\|\Delta_{t+\ell+1} - \Delta_t\|_2^p\big]\big)^{1/p} - \big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p}$ is bounded as

$$\begin{aligned}D \le 2\eta\Big\{\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big) + \eta\omega_p\big(\gamma_{\max} + \sigma_L d\big)\Big\} \cdot \Big(\big(\mathbb{E}\|\Delta_{t+\ell}\|_2^p\big)^{1/p} + \eta p\tau\sqrt{d}\bar{\sigma}\Big)\\ + \Big(2\eta p\sqrt{d} + \eta^2\big(\gamma_{\max} + \sigma_L d\big)\beta_p\Big)\end{aligned}$$

Lemma 3.5 implies that $\big(\mathbb{E}\big[\|\Delta_{t+\ell}\|_2^p\big]\big)^{1/p} \le e\big(\mathbb{E}\big[\|\Delta_t\|_2^p\big]\big)^{1/p} + 6\eta p\ell\sqrt{d}\bar{\sigma}$ and solving the recursion, we arrive at the bound

$$\begin{aligned}&\big(\mathbb{E}\big[\|\Delta_{t+\ell} - \Delta_t\|_2^p\big]\big)^{1/p}\\ &\le 12\eta\ell\Big(\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big) + \eta\omega_p\big(\gamma_{\max} + \sigma_L d\big)\Big) \cdot \Big(\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + \eta p(\tau + \ell)\sqrt{d}\bar{\sigma}\Big)\\ &\qquad + \Big(2\eta p\sqrt{d} + \eta^2\big(\gamma_{\max} + \sigma_L d\big)\beta_p\Big)\ell\bar{\sigma}\\ &\le \eta\Big(12\big(p\sqrt{d}\sigma_L + \gamma_{\max}\big)\ell + \tfrac{\omega_p}{2}\Big)\Big(\big(\mathbb{E}\|\Delta_t\|_2^p\big)^{1/p} + \eta p(\tau + \ell)\sqrt{d}\bar{\sigma}\Big) + \eta\big(2p\ell\sqrt{d} + \tfrac{1}{2}\beta_p\big)\bar{\sigma},\end{aligned}$$

for any $\tau \ge 2t_{\text{mix}} p \log(c_0 d)$ and stepsize choice $\eta \le \frac{c}{48(\gamma_{\max} + \sigma_L d)}$.

B.4 Auxiliary results underlying Theorem 3.1

In this appendix, we prove two auxiliary lemmas that were used in the proof of Theorem 3.1.

B.4.1 Proof of Lemma 3.9

According to Lemma 3.4, given $\tau > 0$ fixed, for any $t \geq \tau + k_m$, there exists a random variable $\widetilde{s}_{t-k_m}$ such that $\widetilde{s}_{t-k_m} \mid \mathcal{F}_{t-k_m-\tau} \sim \xi$, and $\mathbb{E}\big[\rho(s_{t-k_m}, \widetilde{s}_{t-k_m}) \mid \mathcal{F}_{t-\tau-k_m}\big] \leq c_0 \cdot 2^{1-\frac{\tau}{t_{\mathrm{mix}}}}$. By Assumption 3.1, conditionally on the pair of states $(s_{t-k_m}, \widetilde{s}_{t-k_m})$, we have the following bound for $j \in [m]$:

$$\mathcal{W}_{\rho,1}\big(P^{k_j-k_{j-1}}\delta_{s_{t-k_j}}, P^{k_j-k_{j-1}}\delta_{\widetilde{s}_{t-k_j}}\big) \leq c_0 \cdot \rho\big(s_{t-k_j}, \widetilde{s}_{t-k_j}\big), \quad \text{a.s.}$$

Consequently, there exists a sequence of random variables $(\widetilde{s}_{t-k_j})_{0 \leq j \leq m-1}$, such that the following relations hold true for $j = 1, 2, \cdots, m$:

$$\begin{aligned}&\widetilde{s}_{t-k_{j-1}} \mid \mathcal{F}_{t-k_m} \sim P^{k_j-k_{j-1}}\delta_{\widetilde{s}_{t-k_j}}, \quad \text{and}\\ &\mathbb{E}\big[\rho\big(\widetilde{s}_{t-k_{j-1}}, s_{t-k_{j-1}}\big) \mid \mathcal{F}_{t+k-\ell}\big] \leq c_0^{m+1-j} \cdot \rho\big(s_{t-k_m}, \widetilde{s}_{t-k_m}\big).\end{aligned}$$

Based on above construction, we consider the following decomposition:

$$\begin{aligned}\big(\prod_{j=0}^{m} N_{t-k_j}\big)\Delta_{t-k_m} &= \big(\prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\widetilde{s}_{t-k_j})\big)\Delta_{t-k_m-\tau} + \big(\prod_{j=0}^{m} N(\widetilde{s}_{t-k_j})\big) \cdot \Delta_{t-k_m-\tau}\\ &\quad + \big(\prod_{j=0}^{m} N(s_{t-k_j})\big) \cdot \big(\Delta_{t-k_m} - \Delta_{t-\tau-k_m}\big) := Q_1(t) + Q_2(t) + Q_3(t).\end{aligned} \tag{B.11}$$

In the following, we bound the moments for the summation of the three terms above, respectively. For the first term, we note the telescoping equation:

$$\prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\widetilde{s}_{t-k_j}) = \sum_{q=0}^{m} \big(\prod_{j=0}^{q-1} N(s_{t-k_j})\big)\big(\boldsymbol{L}(s_{t-k_q}) - \boldsymbol{L}(\widetilde{s}_{t-k_q})\big)\big(\prod_{j=q+1}^{m} N(\widetilde{s}_{t-k_j})\big).$$

Note that each matrix in the product has operator norm uniformly bounded by $\sigma_L d$. We can then use the Lipschitz condition 3.4 as well as the bound on the distance $\rho(s_{t-k_q}, \widetilde{s}_{t-k_q})$, and obtain the bound

$$\begin{aligned}&\mathbb{E}\big[\|\!| \prod_{j=0}^{m} N(s_{t-k_j}) - \prod_{j=0}^{m} N(\widetilde{s}_{t-k_j}) \|\!|_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau}\big]\\ &\leq (m+1) \cdot (\sigma_L d)^m \sum_{q=0}^{m} \mathbb{E}\big[\|\!| \boldsymbol{L}(s_{t-k_q}) - \boldsymbol{L}(\widetilde{s}_{t-k_q}) \|\!|_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau}\big]\\ &\leq (m+1)^2 (c_0 \sigma_L d)^{m+1} \cdot 2^{-\frac{\tau}{t_{\mathrm{mix}}}}.\end{aligned}$$

Applying the bound on $\|\Delta_{t-\tau}\|_2$ in Proposition 3.1 and taking $\tau \geq 3mt_{\mathrm{mix}} p \log(c_0 dn)$, we find that

$$\begin{aligned}\mathbb{E}\big[\|Q_1(t)\|_2^2\big] &\leq \mathbb{E}\Big[\mathbb{E}\big[\vert\kern-0.25ex\vert\kern-0.25ex\vert \prod_{j=0}^m N(s_{t-k_j}) - \prod_{j=0}^m N(\widetilde{s}_{t-k_j}) \vert\kern-0.25ex\vert\kern-0.25ex\vert_{\mathrm{op}}^2 \mid \mathcal{F}_{t-k_m-\tau}\big] \cdot \|\Delta_{t-\tau-k_m}\|_2^2\Big] \\ &\leq (m+1)^2 (c_0 \sigma_L d)^{m+1} \cdot 2^{-\frac{\tau}{t_{\mathrm{mix}}}} c\bar{\sigma}^2 \frac{\eta \tau d \log^2 n}{1-\kappa} \;\leq\; \frac{\sigma_L^{m+1}}{n^2} \bar{\sigma}^2. \qquad \text{(B.12)}\end{aligned}$$

Now we turn to bounding the term $Q_2(t)$. First, we note that

$$\begin{aligned}\mathbb{E}\big[\|Q_2(t)\|_2^2\big] &\leq \mathbb{E}\big[\vert\kern-0.25ex\vert\kern-0.25ex\vert \prod_{j=0}^{m-1} N(\widetilde{s}_{t-k_j}) \vert\kern-0.25ex\vert\kern-0.25ex\vert_{\mathrm{op}}^2 \cdot \|N(\widetilde{s}_{t-k_m}) \Delta_{t-k_m-\tau}\|_2^2\big] \\ &\leq (\sigma_L d)^{2m} \mathbb{E}\big[\|N(\widetilde{s}_{t-k_m}) \Delta_{t-k_m-\tau}\|_2^2\big] \leq (\sigma_L d)^{2m} \cdot \sigma_L^2 d \cdot \mathbb{E}\big[\|\Delta_{t-k_m-\tau}\|_2^2\big].\end{aligned}$$

By Proposition 3.1, for $t \geq n_0$ and $n_0 \geq 2(\tau + k_m)$, we have: $\mathbb{E}\big[\|\Delta_{t-k_m-\tau}\|_2^2\big] \leq \frac{c\eta}{1-\kappa} t_{\mathrm{mix}} d\bar{\sigma}^2$. If $m = 0$, we have that $\mathbb{E}\big[N(\widetilde{s}_{t+\tau}) \mid \mathcal{F}_t\big] = 0$ almost surely for each $t \geq n_0$. For $m \geq 1$, the conditional unbiasedness does not hold true, but we still have the following upper bound on the bias

$$\begin{aligned}&\vert\kern-0.25ex\vert\kern-0.25ex\vert \mathbb{E}\big[\prod_{j=0}^m N(\widetilde{s}_{t+k_m+\tau-k_j}) \mid \mathcal{F}_t\big] \vert\kern-0.25ex\vert\kern-0.25ex\vert_{\mathrm{op}} \\ &= \sup_{u,v \in \mathbb{S}^{d-1}} \mathbb{E}\big[\langle u, \prod_{j=0}^m N(\widetilde{s}_{t+k_m+\tau-k_j}) v\rangle\big] \\ &\leq \sup_{u,v \in \mathbb{S}^{d-1}} \mathbb{E}\big[\|N(\widetilde{s}_{t+k_m+\tau})^\top u\|_2 \cdot \vert\kern-0.25ex\vert\kern-0.25ex\vert \prod_{j=1}^{m-1} N(\widetilde{s}_{t+k_m+\tau-k_j}) \vert\kern-0.25ex\vert\kern-0.25ex\vert_{\mathrm{op}} \cdot \|N(\widetilde{s}_{t+\tau}) v\|_2\big] \\ &\leq (\sigma_L d)^{m-1} \sup_{u,v \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}\|N(\widetilde{s}_{t+k_m+\tau})^\top u\|_2^2 \cdot \mathbb{E}\|N(\widetilde{s}_{t+\tau}) v\|_2^2} \\ &\leq (\sigma_L d)^{m-1} \cdot \sigma_L^2 d.\end{aligned}$$

Denote $Y_t := \prod_{j=0}^m N(s_{t-k_j})$ and $\widetilde{Y}_t := \prod_{j=0}^m N(\widetilde{s}_{t-k_j})$ for any $t \geq k_m$. We have the expansion:

$$\begin{aligned}\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} Q_2(t)\|_2^2\big] &\leq 2\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} \mathbb{E}[\widetilde{Y}_t] \cdot \Delta_{t-k_m-\tau}\|_2^2\big] + 2\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} (\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t]) \cdot \Delta_{t-k_m-\tau}\|_2^2\big] \\ &\leq 2n\big(d^m \sigma_L^{m+1}\big)^2 \sum_{t=n_0}^{n} \mathbb{E}\|\Delta_{t-k_m-\tau}\|_2^2 \\ &\qquad + 2 \sum_{n_0 \leq s,t \leq n-1} \mathbb{E}\big[\langle (\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t]) \cdot \Delta_{t-k_m-\tau},\, (\widetilde{Y}_s - \mathbb{E}[\widetilde{Y}_s]) \cdot \Delta_{s-k_m-\tau}\rangle\big].\end{aligned}$$

Note that in the special case of $m = 0$, we have $\mathbb{E}[\widetilde{Y}_t] = 0$ so that the bound holds without the first term on the RHS.

For $t > s + \tau + k_m$, we have the relations

$$\mathbb{E}\big[(\widetilde{Y}_t - \mathbb{E}[\widetilde{Y}_t]) \cdot \Delta_{t-k_m-\tau} \mid \widetilde{\mathcal{F}}_{t-k_m-\tau}\big] = 0, \quad \text{and} \quad (\widetilde{Y}_s - \mathbb{E}[\widetilde{Y}_s]) \cdot \Delta_{s-k_m-\tau} \in \widetilde{\mathcal{F}}_{t-k_m-\tau},$$

meaning that the product term vanishes when $|s - t| > \tau + k_m$. Therefore, we arrive at the bound

$$\begin{aligned}
&\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} Q_2(t) \|_2^2\big] \\
&\quad \leq \begin{cases} \big(2n^2 \big(d^m \sigma_L^{m+1}\big)^2 + 4n(k_m + \tau) \cdot (\sigma_L d)^{2m} \cdot \sigma_L^2 d\big) \cdot \frac{c\eta}{1-\kappa} d t_{\mathrm{mix}} \bar{\sigma}^2 & m \geq 1, \\ 4n\tau \sigma_L^2 d \cdot \frac{c\eta}{1-\kappa} d t_{\mathrm{mix}} \bar{\sigma}^2 & m = 0. \end{cases}
\end{aligned} \tag{B.13}$$

Now we turn to the last term in the decomposition (B.11). We start with the decomposition:

$$\Delta_t - \Delta_{t-\tau} = \eta \sum_{\ell=1}^{\tau} \big(L_{t-\ell+1}(s_{t-\ell}) \Delta_{t-\ell} + \nu_{t-\ell} + \zeta_{t-\ell+1}\big).$$

We therefore have the following decomposition:

$$\begin{aligned}
&\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} Q_3(t) \|_2^2\big] \\
&\leq 4\eta^2 \mathbb{E}\big[\| \sum_{t=n_0}^{n} \big\{ Y_t \cdot \big(\sum_{\ell=1}^{\tau} Z_{t-k_m-\ell+1} \Delta_{t-k_m-\ell}\big) \big\} \|_2^2\big] + 4\eta^2 \mathbb{E}\big[\| \sum_{t=n_0}^{n} \big\{ Y_t \cdot \big(\bar{L} \sum_{\ell=1}^{\tau} \Delta_{t-k_m-\ell}\big) \big\} \|_2^2\big] \\
&\qquad + 4\eta^2 \mathbb{E}\big[\| \sum_{t=n_0}^{n} \big\{ Y_t \cdot \big(\sum_{\ell=1}^{\tau} N_{t-k_m-\ell} \Delta_{t-k_m-\ell}\big) \big\} \|_2^2\big] \\
&\qquad\qquad + 4\eta^2 \mathbb{E}\big[\| \sum_{t=n_0}^{n} \big\{ Y_t \cdot \big(\sum_{\ell=1}^{\tau} (\nu_{t-k_m-\ell} + \zeta_{t-k_m-\ell+1})\big) \big\} \|_2^2\big]
\end{aligned}$$

For the martingale component of the noise, note that each term $\prod_{j=0}^{m} N(s_{t-k_j}) \cdot Z_{t-\ell+1}(s_{t-\ell})$ has zero conditional mean conditioned on $\mathcal{F}_{t-\ell}$. We have that

$$\begin{aligned}
&\mathbb{E}\big[\| \sum_{t=n_0}^{n} Y_t Z_{t-k_m-\ell+1}(s_{t-k_m-\ell}) \Delta_{t-k_m-\ell} \|_2^2\big] = \sum_{t=n_0}^{n-1} \mathbb{E}\big[\| Y_t Z_{t-k_m-\ell+1}(s_{t-k_m-\ell}) \Delta_{t-k_m-\ell} \|_2^2\big] \\
&\quad \leq (\sigma_L d)^{2(m+1)} \sum_{t=n_0}^{n-1} \mathbb{E}\big[\| Z_{t-k_m-\ell+1}(s_{t-k_m-\ell}) \Delta_{t-k_m-\ell} \|_2^2\big] \leq \sigma_L^{2m+4} d^{2m+3} n \cdot \tfrac{c\eta}{1-\kappa} d t_{\mathrm{mix}} \bar{\sigma}^2.
\end{aligned}$$

From the Lipschitz condition (3.4) and the boundedness condition (3.3) on the metric space, it follows that $\|Y_t\|_{\mathrm{op}} \leq (\sigma_L d)^{m+1}$ almost surely. Using this fact, the second term can be bounded as

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\cdot\big(\bar{L}\sum_{\ell=1}^{\tau}\Delta_{t-k_m-\ell}\big)\big\}\|_2^2\big] &\leq n\tau(\sigma_L d)^{2m+2}\gamma_{\max}^2\sum_{t=n_0}^{n-1}\sum_{\ell=1}^{\tau}\mathbb{E}\|\Delta_{t-k_m-\ell}\|_2^2\\ &\leq n^2\tau^2(\sigma_L d)^{2m+2}\gamma_{\max}^2\cdot\tfrac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2.\end{aligned}$$

Collecting equations (B.12) and (B.13) as well as the above bounds for Q_3, we arrive at the upper bound $\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\big(\prod_{j=0}^{m}N_{t-k_j}\big)\Delta_{t-k_m}\|_2^2\big] \leq \sum_{j=1}^{3}T_j$, where

$$\begin{aligned}T_1 &:= n^2 d^{2m}\sigma_L^{2m+2}\big(1+\eta^2\tau^2\gamma_{\max}^2 d^2\sigma_L^2+\eta^2\tau^2 d^3\sigma_L^2/n\big)\cdot\tfrac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2\\ T_2 &:= 4\eta^2\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\big(\sum_{\ell=1}^{\tau}N_{t-k_m-\ell}\Delta_{t-k_m-\ell}\big)\big\}\|_2^2\big], \quad \text{and}\\ T_3 &:= 4\eta^2\mathbb{E}\big[\|\sum_{t=n_0}^{n}\big\{Y_t\big(\sum_{\ell=1}^{\tau}(\nu_{t-k_m-\ell}+\zeta_{t-k_m-\ell+1})\big)\big\}\|_2^2\big].\end{aligned}$$

In the special case of $m=0$, we have:

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}N_t\Delta_t\|_2^2\big] \leq c\sigma_L^2 d\cdot\big(n\tau+n^2\eta^2\sigma_L^2 d\tau^2\big)&\frac{c\eta}{1-\kappa}dt_{\mathrm{mix}}\bar{\sigma}^2\\ &+4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n}N_tN_{t-k_1}\Delta_{t-k_1}\|_2^2\big]\\ &\quad+4\eta^2\tau\sum_{k_1=1}^{\tau}\mathbb{E}\big[\|\sum_{t=n_0}^{n}N_t\big(\nu_{t-k_1}+\zeta_{t-k_1+1}\big)\|_2^2\big].\end{aligned}$$

which completes the proof of this lemma.

B.4.2 Proof of Lemma 3.10

We study the bias and variance of the summation separately. For the bias term, we have:

$$\begin{aligned}&\|\mathbb{E}\big[\big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big)\big]\|_2\\ &=\sup_{z\in\mathbb{S}^{d-1}}\mathbb{E}\big[\langle\big(\prod_{j=0}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k_m}+\zeta_{t-k_m+1}\big),\, z\rangle\big]\\ &\overset{(i)}{\leq}\sup_{z\in\mathbb{S}^{d-1}}\sqrt{\mathbb{E}\|N_t^\top z\|_2^2}\cdot\big[\mathbb{E}\|\big(\prod_{j=1}^{m-1}N_{t-k_j}\big)\big(\nu_{t-k}+\zeta_{t-k+1}\big)\|_2^2\big]^{1/2}\\ &\overset{(ii)}{\leq}\sigma_L\sqrt{d}\cdot(\sigma_L d)^{m-1}\cdot 2\bar{\sigma}\sqrt{d}=2(\sigma_L d)^m\bar{\sigma},\end{aligned}\tag{B.14}$$

where step (i) uses the Cauchy–Schwarz inequality, and step (ii) follows by invoking the moment assumption 3.2 as well as the Lipschitz assumption 3.4.

For $t \in [k_m, n]$, we define

$$\lambda_t := \big(\prod_{j=0}^{m-1} N_{t-k_j}\big)\big(\nu_{t-k_m} + \zeta_{t-k_m+1}\big) - \mathbb{E}\big[\big(\prod_{j=0}^{m-1} N_{t-k_j}\big)\big(\nu_{t-k_m} + \zeta_{t-k_m+1}\big)\big].$$

We have

$$\begin{aligned}\mathbb{E}\big[\|\lambda_t\|_2^2\big] \;\leq\; \mathbb{E}\big[\big(\prod_{j=0}^{m-1} |\!|\!| N_{t-k_j} |\!|\!|_{\mathrm{op}}^2\big) \cdot \|\nu_{t-k_m} + \zeta_{t-k_m+1}\|_2^2\big] &\leq (\sigma_L d)^{2m} \cdot \mathbb{E}\big[\|\nu_{t-k} + \zeta_{t-k+1}\|_2^2\big]\\ &\leq d^{2m+1}\sigma_L^{2m}\bar{\sigma}^2.\end{aligned}$$

For integers $t \geq 0$ and $\ell \geq k_m$, by Lemma 3.4, there exists a random variable $\widetilde{s}_{t+\ell-k_m}$, such that $\widetilde{s}_{t+\ell-k_m} \mid \mathcal{F}_t \sim \xi$, and that $\mathbb{E}\big[\rho(s_{t+\ell-k_m}, \widetilde{s}_{t+\ell-k_m}) \mid \mathcal{F}_t\big] \leq c_0 \cdot 2^{1-\frac{\ell-k_m}{t_{\mathrm{mix}}}}$. By Assumption 3.1, conditionally on the pair of states $(s_{t+\ell-k_m}, \widetilde{s}_{t+\ell-k_m})$, we have the following bound for $j \in [m]$:

$$\mathcal{W}_{\rho,1}\big(P^{k_j-k_{j-1}}\delta_{s_{t+\ell-k_j}}, P^{k_j-k_{j-1}}\delta_{\widetilde{s}_{t+\ell-k_j}}\big) \leq c_0 \cdot \rho\big(s_{t+\ell-k_j}, \widetilde{s}_{t+\ell-k_j}\big), \quad \text{a.s.}$$

Consequently, there exists a sequence of random variables $(\widetilde{s}_{t+\ell-k_j})_{0\leq j\leq m-1}$, such that the following relations hold true for $j = 1, 2, \cdots, m$:

$$\begin{aligned}&\widetilde{s}_{t+\ell-k_{j-1}} \mid \mathcal{F}_{t+\ell-k_m} \sim P^{k_j-k_{j-1}}\delta_{\widetilde{s}_{t+\ell-k_j}}, \quad \text{and}\\ &\mathbb{E}\big[\rho\big(\widetilde{s}_{t+\ell-k_{j-1}}, s_{t+\ell-k_{j-1}}\big) \mid \mathcal{F}_{t+\ell-k_m}\big] \leq c_0^{m+1-j} \cdot \rho\big(s_{t+\ell-k_m}, \widetilde{s}_{t+\ell-k_m}\big).\end{aligned}$$

Given the random variables constructed above, we can then construct the proxy random variable for $\lambda_{t+\ell}$:

$$\begin{aligned}\widetilde{\lambda}_{t+\ell} := \big(\prod_{j=0}^{m-1} N(\widetilde{s}_{t+\ell-k_j})\big)\big(\nu(\widetilde{s}_{t+\ell-k_m}) + \zeta_{t+\ell-k_m+1}(\widetilde{s}_{t+\ell-k_m})\big)&\\ - \mathbb{E}\big[\big(\prod_{j=0}^{m-1} N_{t-k_j}\big)\big(\nu_{t-k_m} + \zeta_{t-k_m+1}\big)\big]&.\end{aligned}$$

By stationarity, we have $\mathbb{E}\big[\widetilde{\lambda}_{t+\ell} \mid \mathcal{F}_t\big] = 0$ almost surely. In order to bound the difference, we note the telescope relation: $\widetilde{\lambda}_{t+\ell} - \lambda_{t+\ell} = \sum_{q=0}^{m-1} E_q^{(mix)} + \bar{E}^{(mix)}$, where

$$\begin{aligned}E_q^{(mix)} := \big(\prod_{j=0}^{q-1} N(s_{t+\ell-k_j})\big)\big(\bar{L}(\widetilde{s}_{t+\ell-k_q}) - \boldsymbol{L}(s_{t+\ell-k_q})\big)\big(\prod_{j=q+1}^{m-1} N(\widetilde{s}_{t+\ell-k_j})\big)&\\ \cdot \big(\nu(\widetilde{s}_{t+\ell-k_m}) + \zeta_{t+\ell-k_m+1}(\widetilde{s}_{t+\ell-k_m})\big)&,\end{aligned}$$

and

$$\begin{aligned}\bar{E}^{(mix)} := \prod_{j=0}^{m-1} N(s_{t+\ell-k_j})\\ \cdot \Big(\nu(\widetilde{s}_{t+\ell-k_m}) + \zeta_{t+\ell-k_m+1}(\widetilde{s}_{t+\ell-k_m}) - \nu(s_{t+\ell-k_m}) + \zeta_{t+\ell-k_m+1}(s_{t+\ell-k_m})\Big).\end{aligned}$$

Using the Wasserstein distance bounds and Lipschitz condition 3.4, we find the conditional expectation $A = \mathbb{E}\big[\|E_q^{(mix)}\|_2 \mid \mathcal{F}_t\big]$ is bounded as

$$\begin{aligned}A &\leq (\sigma_L d)^{m-1}\mathbb{E}\big[\|\boldsymbol{L}(s_{t+\ell-k_q}) - \boldsymbol{L}(\widetilde{s}_{t+\ell-k_q})\|_{\mathrm{op}} \cdot \|\nu(\widetilde{s}_{t+\ell-k}) + \zeta_{t+\ell-k+1}(\widetilde{s}_{t+\ell-k})\|_2 \mid \widetilde{\mathcal{F}}_t\big]\\ &\leq (\sigma_L d)^m \sqrt{\mathbb{E}[\rho(s_{t+\ell-k_q}, \widetilde{s}_{t+\ell-k_q})^2 \mid \widetilde{\mathcal{F}}_t]} \cdot \sqrt{\mathbb{E}[\|\nu(\widetilde{s}_{t+\ell-k}) + \zeta_{t+\ell-k+1}(\widetilde{s}_{t+\ell-k})\|_2^2 \mid \widetilde{\mathcal{F}}_t]}\\ &\leq (\sigma_L d)^m c_0 \cdot 2^{1-\frac{\ell-k_q}{2t_{\mathrm{mix}}}} \cdot 2d\bar{\sigma},\end{aligned}$$

and the conditional expectation $B = \mathbb{E}\big[\|\bar{E}^{(mix)}\|_2 \mid \mathcal{F}_t\big]$ is bounded as

$$\begin{aligned}B &\leq (\sigma_L d)^m \sqrt{\mathbb{E}\big[\|\zeta_{t+\ell-k+1}(s_{t+\ell-k}) - \zeta_{t+\ell-k+1}(\widetilde{s}_{t+\ell-k})\|_2^2 \mid \mathcal{F}_t\big]}\\ &\quad + (\sigma_L d)^m \sqrt{\mathbb{E}\big[\|\nu(s_{t+\ell-k}) - \nu(\widetilde{s}_{t+\ell-k})\|_2^2 \mid \mathcal{F}_t\big]}\\ &\leq (\sigma_L d)^m d\bar{\sigma} c_0 \cdot 2^{1-\frac{\ell-k_m}{2t_{\mathrm{mix}}}}.\end{aligned}$$

Consequently, we can bound the cross term as

$$\begin{aligned}\mathbb{E}\big[\langle\lambda_t,\, \lambda_{t+\ell}\rangle\big] &= \mathbb{E}\big[\langle\lambda_t,\, \mathbb{E}\big[\widetilde{\lambda}_{t+\ell} \mid \mathcal{F}_t\big]\rangle\big] + \mathbb{E}\big[\langle\lambda_t,\, \mathbb{E}\big[\lambda_{t+\ell} - \widetilde{\lambda}_{t+\ell} \mid \mathcal{F}_t\big]\rangle\big]\\ &\leq 0 + \mathbb{E}\big[\|\lambda_t\|_2 \cdot \mathbb{E}\big[\|\lambda_{t+\ell} - \widetilde{\lambda}_{t+\ell}\|_2 \mid \mathcal{F}_t\big]\big]\\ &\leq 12c_0 d^{m+1}\sigma_L^m\bar{\sigma} \cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}} \cdot \sqrt{\mathbb{E}\|\lambda_t\|_2^2}\\ &\leq 12c_0 d^{2m+2}\sigma_L^{2m}\bar{\sigma}^2 \cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}}.\end{aligned}$$

Taking $\tau = 16t_{\mathrm{mix}}\log(c_0 d)$, we can control the cross terms in two different ways:

$$\mathbb{E}\big[\langle\lambda_t,\, \lambda_{t+\ell}\rangle\big] \leq \begin{cases}\sqrt{\mathbb{E}\|\lambda_t\|_2^2} \cdot \sqrt{\mathbb{E}\|\lambda_{t+\ell}\|_2^2} \leq d^{2m+1}\sigma_L^{2m}\bar{\sigma}^2, & 0 \leq \ell \leq k_m + \tau,\\ 12c_0 d^{2m+2}\sigma_L^{2m}\bar{\sigma}^2 \cdot 2^{-\frac{\ell-k}{2t_{\mathrm{mix}}}} \leq d^{2m}\sigma_L^{2m}\bar{\sigma}^2 & \ell \geq k_m + \tau.\end{cases}$$

Summing them up these terms yields

$$\begin{aligned}\mathbb{E}\big[\|\sum_{t=n_0}^{n-1}\lambda_t\|_2^2\big] &= \sum_{t=n_0}^{n-1}\mathbb{E}\|\lambda_t\|_2^2 + 2\sum_{n_0 \leq t_1 < t_2 \leq n-1}\mathbb{E}\big[\langle\lambda_{t_1},\, \lambda_{t_2}\rangle\big]\\ &\leq (k+\tau+1)nd^{2m+1}\sigma_L^{2m}\bar{\sigma}^2 + n^2 d^{2m}\sigma_L^{2m}\bar{\sigma}^2.\end{aligned}$$

Combining with the bound (B.14), we find that

$$\begin{aligned}
&\mathbb{E}\big[\| \sum_{t=n_0}^{n-1} \big(\prod_{j=0}^{m-1} N_{t-k_j} \big) \big(\nu_{t-k_m} + \zeta_{t-k_m+1} \big) \|_2^2 \big] \\
&= \| \sum_{t=n_0}^{n-1} \mathbb{E}\big[\big(\prod_{j=0}^{m-1} N_{t-k_j} \big) \big(\nu_{t-k_m} + \zeta_{t-k_m+1} \big) \big] \|_2^2 + \mathbb{E}\big[\| \sum_{t=n_0}^{n-1} \lambda_t \|_2^2 \big] \\
&\leq c \big(n^2 + (k_m + \tau) n d \big) \sigma_L^{2m} d^{2m} \bar{\sigma}^2,
\end{aligned}$$

for a universal constant $c > 0$.

B.5 Proof of Theorem 3.2

Our strategy is to prove a Bayes risk lower bound. We construct a prior distribution over transition kernels by perturbing the base matrix P_0 appropriately. We then apply the Bayesian Cramér–Rao lower bound to obtain our result.

Let us describe the construction in more detail. For each $s \in \mathcal{S}$, suppose we have a perturbation vector $h_s \in \mathbb{R}^{\mathcal{S}}$. Use these to define the perturbed transition kernel

$$P_h(x, y) := \tfrac{P_0(x,y)\, e^{h_x(y)}}{\sum_{z \in \mathcal{S}} P_0(x,z) e^{h_x(z)}} \qquad \text{for each } x, y \in \mathcal{S}.$$

Note that by construction, for any $x \in \mathcal{S}$ and any $h_x \in \mathbb{R}^{\mathcal{S}}$, we have $\mathrm{supp}\big(P_h(x, \cdot)\big) = \mathrm{supp}\big(P_0(x, \cdot)\big)$. Since P_0 is irreducible and aperiodic, so is P_h. Therefore, the stationary distribution ξ_h of P_h exists and is unique. When the perturbation is small enough, a quantitative perturbation principle can be obtained, which we collect in Lemma B.1 below.

It remains to specify how the perturbation vectors are generated. We parameterize h with a linear transformation, writing $h = Qw$ for a linear operator Q to be specified shortly, and a random vector $w \in \mathbb{R}^d$ drawn from a distribution ρ. In particular, given a collection of vectors $\{q_x(y)\}_{x,y \in \mathcal{S}} \subseteq \mathbb{R}^d$, we consider the linear transformation $Q : \mathbb{R}^d \to \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ given by $w \mapsto \big[\langle w,\, q_x(y) \rangle\big]_{x,y \in \mathcal{S}}$.

Next we specify the prior ρ, and along with some associated notation. Define the subspace $\mathbb{H}_h := \big\{ f \in \mathbb{R}^{\mathcal{S}} : \mathbb{E}_{\xi_h}[f(s)] = 0 \big\}$, and note that P_h maps $\mathbb{H}_h$ to itself. Furthermore, since P_h is irreducible and aperiodic, the mapping $(I - P_h)$ is invertible on $\mathbb{H}_h$. Consequently, for any function $f : \mathcal{S} \to \mathbb{R}$, the following Green function operator is well-defined:

$$\mathcal{A}_h f := (I - P_h)^{-1}\big|_{\mathbb{H}_h} \cdot \big(f - \mathbb{E}_{\xi_h}[f] \big) \in \mathbb{R}^{\mathcal{S}}.$$

We also define an operator $\mathcal{P}_h$ on the space of real-valued functions on $\mathcal{S}$ as follows:

$$\mathcal{P}_h f(x) := \mathbb{E}_{Y \sim P_h(x, \cdot)}[f(Y)].$$

Importantly, $\mathcal{P}_h$ is an operator mapping functions to functions, and distinct from the matrix P_h. It is straightforward to see that the operator $\mathcal{P}_h$ commutes with the operator $\mathcal{A}_h$, for any perturbation matrix h. Finally, for any $h \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ and for all $x \in \mathcal{S}$, we define

$$\boldsymbol{g}_h(x) = \big(I_d - \mathbb{E}_{\xi_h}[\boldsymbol{L}(s)]\big)^{-1} \big(\mathcal{A}_h \boldsymbol{L}(x) \cdot \bar{\theta}(P_h) + \mathcal{A}_h \boldsymbol{b}(x)\big). \tag{B.15}$$

Since the proof works under the perturbed probability transition kernel P_h, it is useful to study the effect of small perturbation on its stationary distribution. The following lemma provides non-asymptotic bounds on the mixing time of perturbed Markov chain and its stationary distribution ξ_h, which will be useful throughout the proof.

Lemma B.1. *Under the setup above, suppose that* $h_{\max} := \max_{x \in \mathcal{S}} \|h_x\|_\infty < \frac{1}{128 t_{\text{mix}}}$. *Then the perturbed transition kernel satisfies the following.*

- *The Markov transition kernel* P_h *satisfies the mixing condition (Assumption 3.1) with the discrete metric and mixing time* $4t_{\text{mix}}$.
- *The stationary distribution* ξ_h *satisfies the bound*

$$\max_{s \in \mathcal{S}} \big\{ \log \tfrac{\xi_0(s)}{\xi_h(s)}, \ \log \tfrac{\xi_h(s)}{\xi_0(s)} \big\} \leq t_{\text{mix}} \big(2 + \log h_{\max}^{-1} + \log \tfrac{1}{\min_x \xi_0(x)}\big) h_{\max}.$$

See Section B.5.1 for the proof of this lemma.

With this notation in hand, we are ready to construct the prior distribution on w. We begin with the following one-dimensional density function, taken from [206]:

$$\mu(t) := \cos^2\left(\tfrac{\pi t}{2}\right) \cdot \mathbf{1}_{t \in [-1,1]}. \tag{B.16a}$$

Also, define the positive-definite matrix $\Lambda := \mathbb{E}_{X \sim \xi_0}\big[\operatorname{cov}_{Y \sim P_0(X,\cdot)}\big(\boldsymbol{g}_0(Y) \mid X\big)\big]$, and let $\Lambda = UDU^\top$ denote its eigen-decomposition. For a random variable $\psi \sim \mu^{\otimes d}$, define the perturbation parameter

$$w = \tfrac{1}{\sqrt{n}} U D^{-1/2} \psi, \tag{B.16b}$$

and let its density denote the prior distribution ρ. Note that for any $w \in \operatorname{supp}(\rho)$, we have

$$\|\Lambda w\|_2 = \|UD^{1/2}\psi\|_2 = \|D^{1/2}\psi\|_2 \leq \sqrt{\operatorname{trace}(D)/n} = \sqrt{\operatorname{trace}(\Lambda)/n}. \tag{B.16c}$$

The final ingredient in our construction is to specify the linear transformation Q. For each $x, y \in \mathcal{S}$, we set

$$q_x(y) := \boldsymbol{g}_0(y) - \mathbb{E}_{s' \sim P_0(x,\cdot)}\big[\boldsymbol{g}_0(s')\big], \tag{B.16d}$$

where the Green function $\boldsymbol{g}$ is defined in equation (B.15). Recall that $h = Qw$ for $w \sim \rho$. This specifies our prior over transition kernels, and concludes the construction.

Next, we state a Bayesian Cramér-Rao lower bound used in the proof. Consider a class $\mathcal{P}_\Theta = \big(\mathbb{P}_\eta : \eta \in \Theta\big)$ of probability distributions of sample $X \in \mathbb{X}$, parameterized by $\eta \in \Theta$, where Θ is an open subset of $\mathbb{R}^d$. Recalling Proposition A.1 from Appendix A implies that

$$\mathop{\mathbb{E}}_{\eta\sim\rho} \mathop{\mathbb{E}}_{X\sim p_\eta} \|\widehat{T}(X) - T(\eta)\|_2^2 \geq \frac{\Big(\int \mathrm{trace}\left(\frac{\partial T}{\partial \eta}(\eta)\right)\rho(\eta)d\eta\Big)^2}{\int \mathrm{trace}\Big(I(\eta)\Big)\rho(\eta)d\eta + \int \|\nabla \log \rho(\eta)\|_2^2 \rho(\eta)d\eta}. \tag{B.17}$$

In order to complete the proof, we provide non-asymptotic estimates on the three quantities involved in the right-hand-side of Eq (B.17). These require a few technical lemmas, whose proofs can be found at the end of the section.

Bounds on the term $\mathrm{trace}\left(\nabla_w \bar{\theta}\right)$: We state two technical lemmas that are helpful in bounding this quantity. The first computes the Jacobian matrix of the desired functional $\bar{\theta}(h)$ with respect to the parameter w.

Lemma B.2. *Under the given set-up, for any $w \in \mathbb{R}^d$, we have*

$$\nabla_w \bar{\theta}(P_h) = \mathbb{E}_{X\sim\xi_h}\Big[\mathrm{cov}_{Y\sim P_h(X,\cdot)}\left\{\boldsymbol{g}_h(Y) - \mathcal{P}_h\boldsymbol{g}_h(X), \boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\right\}\Big]. \tag{B.18}$$

See Section B.5.2 for the proof of this lemma. Next, we control the RHS of equation (B.18) by replacing $\boldsymbol{g}_h$ with $\boldsymbol{g}_0$.

Lemma B.3. *Under the given set-up and for a sample size lower bounded as $n \geq \frac{ct_{\mathrm{mix}}^2\sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2}$ and $\max_{x\in\mathcal{S}} \|h_x\|_\infty \leq \frac{1}{128t_{\mathrm{mix}}}$, we have*

$$\mathbb{E}_{Z\sim\xi_h}\left[\|\boldsymbol{g}_h(Z) - \boldsymbol{g}_0(Z)\|_2^2\right] \leq \tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\mathrm{mix}}^4 d^2}{(1-\kappa)^4 n} \log^6 \tfrac{d}{\min_x \xi_0(x)}.$$

Furthermore, for any w in the support of ρ, we have

$$\|\bar{\theta}(P_h) - \bar{\theta}(P_0)\|_2 \leq \tfrac{3}{2}\sqrt{\mathrm{trace}(\Lambda)/n} + \sqrt{\tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\mathrm{mix}}^4 d^3}{(1-\kappa)^4 n^2} \log^6 \tfrac{d}{\min_x \xi_0(x)}}.$$

See Section B.5.3 for the proof of this lemma.

Combining these two lemmas yields

$$\begin{aligned}
&\mathrm{trace}\left(\nabla_w \bar{\theta}\right)\\
&\geq \mathbb{E}_{X\sim\xi_h}\Big[\mathrm{var}_{Y\sim P_h(X,\cdot)}\left(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\right)\Big]\\
&\qquad - \mathbb{E}_{X\sim\xi_h}\Big[\sqrt{\mathrm{var}_{Y\sim P_h(X,\cdot)}\left(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\right)}\Big] \cdot \sqrt{\mathbb{E}_{Z\sim\xi_h}\left[\|\boldsymbol{g}_h(Z) - \boldsymbol{g}_0(Z)\|_2^2\right]}\\
&\geq \mathrm{trace}\left(\Lambda\right) - \sqrt{\mathrm{trace}\left(\Lambda\right)} \cdot \tfrac{c(1+\sigma_L)\bar{\sigma} t_{\mathrm{mix}}^2 d}{(1-\kappa)^2\sqrt{n}} \log^3 \tfrac{d}{\min_x \xi_0(x)}.
\end{aligned}$$

Now given a sample size lower bounded as $n \geq \frac{ct_{\mathrm{mix}}^2\sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2} + \frac{2c(1+\sigma_L^2)\bar{\sigma}^2 t_{\mathrm{mix}}^4 d^2}{(1-\kappa)^4\,\mathrm{trace}(\Lambda)} \log^6 \frac{d}{\min_x \xi_0(x)}$, we can conclude that

$$\mathrm{trace}\left(\nabla_w \bar{\theta}\right) \geq \tfrac{1}{2}\,\mathrm{trace}(\Lambda) \qquad \text{for any } w \text{ in the support of } \rho. \tag{B.19}$$

Bounds on the Fisher information $I^{(n)}(w)$: We now state an upper bound on the Fisher information of the observed trajectory:

Lemma B.4. *Under the given set-up, for any $w \in \mathbb{R}^d$, if $h_{\max} := \max_x \|h\|_\infty$ satisfies the inequality $h_{\max}^{-1} \geq ct_{\mathrm{mix}}\big(\log h_{\max}^{-1} + \log(\min \xi_0)^{-1}\big)$, we have*

$$I^{(n)}(w) := \mathbb{E}_h\big[\nabla_w \log \mathbb{P}_h\big(s_0^n\big)\nabla_w \log \mathbb{P}_h\big(s_0^n\big)^\top\big] \preceq \tfrac{3n}{2}\mathbb{E}_{X\sim\xi_h}\big[\operatorname{cov}_{Y\sim P_h(X,\cdot)}\big(q_X(Y) \mid X\big)\big].$$

See Section B.5.4 for the proof of this lemma.

In order to apply the preceding lemma, we must verify the condition on $h_{\max}$ for our setting. Under our construction, we have $\max_{x\in\mathcal{S}} \|h_x\|_\infty = \max_{x,y\in\mathcal{S}}\langle \boldsymbol{g}_0(y) - \mathcal{P}_0\boldsymbol{g}_0(x),\, w\rangle$. Note that Assumption 3.2 and Lemma B.7 in Section B.5.7 together imply the following bound for any $\delta > 0$:

$$\xi_0\big(s : |\langle \boldsymbol{g}_0(s),\, w\rangle| \leq \tfrac{c\bar{\sigma}t_{\mathrm{mix}}\|w\|_2}{1-\kappa} \cdot \log^3 \tfrac{d}{\delta}\big) > 1 - \delta.$$

Taking $\delta := \frac{1}{2}\min_{s\in\mathcal{S}} \xi_0(s) > 0$, we have the uniform bound

$$\max_{s\in\mathcal{S}} |\langle \boldsymbol{g}_0(s),\, w\rangle| \leq \tfrac{c\bar{\sigma}t_{\mathrm{mix}}\|w\|_2}{1-\kappa} \log^3\big(d/\min_s \xi_0(s)\big).$$

Note that $\mathcal{P}_0$ is a probability transition kernel, for any $s \in \mathcal{S}$, the vector $\mathcal{P}_0\boldsymbol{g}_0(s)$ lies in the convex hull of $\big(\boldsymbol{g}_0(s')\big)_{s'\in\mathcal{S}}$. So we have the bound $\max_{s\in\mathcal{S}} |\langle \mathcal{P}_0\boldsymbol{g}_0(s),\, w\rangle| \leq \max_{s\in\mathcal{S}} |\langle \boldsymbol{g}_0(s),\, w\rangle| \leq \frac{c\bar{\sigma}t_{\mathrm{mix}}\|w\|_2}{1-\kappa} \log^3\big(d/\min_s \xi_0(s)\big)$. Putting them together leads to the bound

$$\max_{x\in\mathcal{S}} \|h_x\|_\infty \leq 2c\bar{\sigma}t_{\mathrm{mix}}\|w\|_2 \log^3\big(d/\min_s \xi_0(s)\big).$$

Now given a sample size

$$n \geq ct_{\mathrm{mix}}^3\bar{\sigma}^2 \cdot \operatorname{trace}(\Lambda) \cdot \log^3 \tfrac{d}{\min_s \xi_0(s)}, \tag{B.20}$$

we have that $\max_x \|h_x\|_\infty < \frac{1}{128t_{\mathrm{mix}}}$. This satisfies the condition in Lemma B.1 in the appendix. Applying this lemma, we see that the condition

$$h_{\max}^{-1} \geq ct_{\mathrm{mix}}\big(\log h_{\max}^{-1} + \log(\min \xi_0)^{-1}$$

is satisfied, so that Lemma B.4 guarantees that

$$\begin{aligned}
\operatorname{trace}\big(I^{(n)}(w)\big) &\preceq \tfrac{3n}{2}\mathbb{E}_{X\sim\xi_h}\big[\operatorname{var}_{Y\sim P_h(X,\cdot)}\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\big)\big] \\
&\preceq \big(\tfrac{3}{2}\big)^3 n \cdot \mathbb{E}_{X\sim\xi_0}\big[\operatorname{var}_{Y\sim P_0(X,\cdot)}\big(\boldsymbol{g}_0(Y) \mid X\big)\big] \\
&= \tfrac{27n}{8}\operatorname{trace}\big(\Lambda\big).
\end{aligned} \tag{B.21}$$

The last inequality follows because $\xi_h \preceq \frac{3}{2}\xi_0$ $P_h(x,\cdot) \preceq \frac{3}{2}P_0(x,\cdot)$ for all $x \in \mathcal{S}$.

Bounds on the prior Fisher information: From Lemma A.2 in Appendix A, the density ρ of w has Fisher information

$$I(\rho) = U D^{1/2} I\big(\mu^{\otimes d}\big) D^{1/2} U^\top = n\pi\Lambda. \tag{B.22}$$

Consequently, we have $\int \|\nabla \log \rho(w)\|_2^2 \rho(w) \; dw \, \mathrm{trace}\,\big(I(\rho)\big) = n\pi \cdot \mathrm{trace}(\Lambda)$.

Putting together the pieces: Combining the bounds (B.19), (B.21), and (B.22) and applying Proposition A.1, we obtain the lower bound

$$\inf_{\widehat{\theta}_n} \int_{\mathbb{R}^d} \mathbb{E}_{X_1^n \sim \mathbb{P}_{Qw}} \big[\|\widehat{\theta}_n - \bar{\theta}(P_{Qw})\|_2^2\big] \rho(dw) \geq \tfrac{1}{4(5+\pi)n} \,\mathrm{trace}(\Lambda). \tag{B.23}$$

It remains to relate the matrix Λ to the local complexity ε_n in the theorem. In order to do so, we require the following lemma.

Lemma B.5. *Under the setup above, for any function $f : \mathcal{S} \rightarrow \mathbb{R}$ such that $\mathbb{E}_{\xi_0}[f(s)] = 0$, we have $\mathbb{E}_{X \sim \xi_0, Y \sim P_0(X,\cdot)}\big[\big(\mathcal{A}_0 f(Y) - \mathcal{P}_0 \mathcal{A}_0 f(X)\big)^2\big] = \sum_{k=-\infty}^{\infty} \mathbb{E}\big[f(s_0) f(s_k)\big]$, where $(s_k)_{k \in \mathbb{Z}}$ is a stationary Markov chain following P_0.*

See Section B.5.5 for the proof of this lemma.

Applying Lemma B.5 with $f_j(s) = \langle (I_d - \bar{L}^{(0)})^{-1}\big(\boldsymbol{L}(s)\bar{\theta}(P_0) + \boldsymbol{b}(s)\big), \, e_j \rangle$ for $j = 1, 2, \cdots, d$ respectively, we arrive at the chain of equalities

$$\begin{aligned}
\mathrm{trace}(\Lambda) &= \sum_{j=1}^{d} \mathbb{E}_{X \sim \xi_0, Y \sim P_0(X,\cdot)}\big[\big(\mathcal{A}_0 f_j(Y) - \mathcal{P}_0 \mathcal{A}_0 f_j(X)\big)^2\big] \\
&= \sum_{j=1}^{d} \sum_{k=-\infty}^{\infty} \mathbb{E}\big[f_j(s_0) f_j(s_k)\big] = \mathrm{trace}\big((I - \bar{L}^{(0)})^{-1} {\Sigma^*}_{\mathrm{Mkv}} (I - \bar{L}^{(0)})^{-\top}\big) = n\varepsilon_n^2.
\end{aligned}$$

Thus, the right-hand-side of equation (B.23) is exactly $\frac{\varepsilon_n^2}{4(5+\pi)}$.

It remains to bound the size of the neighborhood. Given a sample size n satisfying the bound (B.20), Lemma B.3 implies that $\|\bar{\theta}(P_h) - \bar{\theta}(P_0)\|_2 \leq \sqrt{\frac{\mathrm{trace}(\Lambda)}{n}}$. Consequently, for any w on the support of ρ, we have $P_{Qw} \in \mathfrak{N}_{\mathrm{Est}}(P_0, 2\varepsilon_n)$.

On the other hand, for any $w \in \mathrm{supp}(\rho)$ and any $x \in \mathcal{S}$ and perturbation $h = Qw$, we have

$$\begin{aligned}
\chi^2\left(P_h(x,\cdot) \;||\; P_0(x,\cdot)\right) &= \mathbb{E}_{Y \sim P_0(x,\cdot)}\big[\big(\tfrac{P_h(x,Y)}{P_0(x,Y)} - 1\big)^2\big] \\
&= \mathrm{var}_{Y \sim P_0(x,\cdot)}\left(\tfrac{e^{h_x(Y)}}{\sum_{z \in \mathcal{S}} P_0(x,z) e^{h_x(z)}}\right) \\
&\overset{(i)}{\leq} \mathrm{var}_{Y \sim P_0(x,\cdot)}\left(e^{h_x(Y)}\right) \\
&\leq \mathbb{E}_{Y \sim P_0(x,\cdot)}\big[\big(e^{h_x(Y)} - 1\big)^2\big] \\
&\overset{(ii)}{\leq} e \cdot \mathbb{E}_{Y \sim P_0(x,\cdot)}\big[h_x(Y)^2\big],
\end{aligned}$$

where step (i) follows by using Jensen's inequality to assert that

$$\sum_{z \in \mathcal{S}} P_0(x,z) e^{h_x(z)} \geq e^{\sum_{z \in \mathcal{S}} P_0(x,z) h_x(z)} = 1,$$

and step (ii) follows from the inequality $|e^x - 1| \leq e \cdot |x|$, valid for $x \in [-1, 1]$.

Accordingly, the average χ^2-divergence admits the bound

$$\sum_{x \in \mathcal{S}} \xi_0(x) \chi^2 \left(P_h(x, \cdot) \;||\; P_0(x, \cdot)\right) \leq e \cdot \mathbb{E}_{X \sim \xi_0, Y \sim P_0(X, \cdot)} \left[\langle w, \, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X) \rangle^2\right]$$
$$\leq e \cdot w^\top \Lambda w \leq \tfrac{ed}{n}.$$

For any w on the support of ρ, we thus have $P_{Qw} \in \mathfrak{N}_{\mathrm{Prob}}(P_0, e\sqrt{\frac{d}{n}})$, as claimed. The Bayes risk lower bound (B.23) then implies the desired minimax lower bound.

B.5.1 Proof of Lemma B.1

The proof relies on a total variation distance bound on the transition kernel. In particular, for each $s \in \mathcal{S}$, we have

$$d_{\mathrm{TV}}\big(P_0(x, \cdot), P_h(x, \cdot)\big) \leq \sqrt{\tfrac{1}{2} \chi^2 \left(P_0(x, \cdot) \;||\; P_h(x, \cdot)\right)} = \sqrt{\tfrac{1}{2} \sum_{y \in \mathcal{S}} P_0(x,y) \cdot \big(\tfrac{P_h(x,y)}{P_0(x,y)} - 1\big)^2}$$
$$\overset{(i)}{\leq} \sqrt{\tfrac{1}{2} \big(e^{\|h_x\|_\infty} - 1\big)^2} \overset{(ii)}{\leq} e \cdot \max_{x \in \mathcal{S}} \|h_x\|_\infty. \tag{B.24}$$

In step (i), we use the fact

$$\frac{P_h(x,y)}{P_0(x,y)} = \frac{e^{h_x(y)}}{\sum_{z \in \mathcal{S}} P_0(x,y) e^{h_x(z)}} \in [e^{-\|h_x\|_\infty}, e^{\|h_x\|_\infty}],$$

and in step (i), we use the fact $\|h_x\|_\infty < 1$.

Next, we turn to proofs of the two claims. We first prove the mixing time bound. Note that the non-expansive condition (3.5)(b) is automatically satisfied with $c_0 = 1$ for total variation distance (by a naïve coupling). Given a fixed pair $x, y \in \mathcal{S}$, invoking Lemma 3.4 with $\tau = 4 t_{\mathrm{mix}}$ yields the existence of a joint distribution over the random sequence $\{x_k\}_{0 \leq k \leq \tau}$ and $\{y_k\}_{0 \leq k \leq \tau}$, such that $\{x_k\}$ and $\{y_k\}$ follows the Markov chain P_0, starting from $x_0 = x$ and $y_0 = y$, respectively. Furthermore, we have the bound $\mathbb{P}\big(x_\tau \neq y_\tau\big) \leq \frac{1}{4}$.

Now we construct a coupling between the original chain and perturbed chain. Taking the initial point $\widetilde{x}_0 = x$, we iteratively construct the sequence $\{\widetilde{x}_k\}_{0 \leq k \leq \tau}$ as follows: given $\widetilde{x}_k$ and x_k, we construct the conditional distribution of $\widetilde{x}_{k+1}$ as follows:

- If $x_k = \widetilde{x}_k$, we let $\mathbb{P}\big(\widetilde{x}_{k+1} \neq x_{k+1} \mid x_k, \widetilde{x}_k\big) = d_{\mathrm{TV}}\big(P_0(x_k, \cdot), P_h(x_k, \cdot)\big)$.

- If $x_k \neq \widetilde{x}_k$, we simply take $\widetilde{x}_{k+1}$ and x_{k+1} to be conditionally independent, following their respective transition kernels.

We construct the sequence $\{\widetilde{y}_k\}_{0 \leq k \leq \tau}$ in a similar fashion.

By the union bound, it follows that

$$\mathbb{P}\big(x_\tau \neq \widetilde{x}_\tau\big) \leq \sum_{k=0}^{\tau-1} \mathbb{E}\big[\mathbb{P}\big(x_{k+1} \neq \widetilde{x}_{k+1} \mid x_k = \widetilde{x}_k\big)\big] = \sum_{k=0}^{\tau-1} \mathbb{E}\big[d_{\mathrm{TV}}\big(P_0(x_k, \cdot), P_h(x_k, \cdot)\big)\big]$$
$$\leq 4et_{\mathrm{mix}} \cdot \max_{x \in \mathcal{S}} \|h_x\|_\infty < \tfrac{1}{8}.$$

In the last step, we have used the total variation distance bound (B.24).

Similarly, the process $\{\widetilde{y}_k\}$ satisfies the bound $\mathbb{P}\big(y_\tau \neq \widetilde{y}_\tau\big) < \frac{1}{8}$. Putting together the pieces, we conclude that

$$d_{\mathrm{TV}}\big(\delta_x P_h^\tau, \delta_y P_h^\tau\big) \leq \mathbb{P}\big(\widetilde{x}_\tau \neq \widetilde{y}_\tau\big) \leq \mathbb{P}\big(\widetilde{x}_\tau \neq x_\tau\big) + \mathbb{P}\big(x_\tau \neq y_\tau\big) + \mathbb{P}\big(y_\tau \neq \widetilde{y}_\tau\big)$$
$$< \tfrac{1}{8} + \tfrac{1}{4} + \tfrac{1}{8} = \tfrac{1}{2},$$

which shows that the perturbed chain P_h satisfies the condition (3.5)(a) with mixing time $\tau = 4t_{\mathrm{mix}}$.

Next, we prove the perturbation result for the stationary distribution. Given any fixed initial distribution π_0, note that for any deterministic sequence $(x_0, x_2, \cdots, x_n)$, we have the following expression for the Radon-Nikodym derivative:

$$\frac{d\mathbb{P}_h\big(x_0, x_1, \cdots, x_n\big)}{d\mathbb{P}_0\big(x_0, x_1, \cdots, x_n\big)} = \prod_{k=0}^{n-1} \frac{P_h(x_k, x_{k+1})}{P_0(x_k, x_{k+1})} = \prod_{k=0}^{n-1} \frac{e^{h_{x_k}(x_{k+1})}}{\sum_{y \in \mathcal{S}} e^{h_{x_k}(y)} P(x_k, y)}.$$

We then have the max-divergence bound

$$D_\infty\left(\mathbb{P}_h\big(x_0^n\big) \mid\mid \mathbb{P}_0\big(x_0^n\big)\right) := \sup_{x_0^n \in \mathcal{S}^n} \left| \log \frac{d\mathbb{P}_h\big(x_0, x_1, \cdots, x_n\big)}{d\mathbb{P}_0\big(x_0, x_1, \cdots, x_n\big)} \right| \leq n \cdot \max_x \|h_x\|_\infty.$$

Taking the marginal distribution, we see that the bound $D_\infty\left(\pi_0 P_h^n \mid\mid \pi_0 P_0^n\right) \leq n \cdot h_{\max}$ holds for any initial distribution π_0 and any $n > 0$.

To obtain the desired claim, we take the initial distribution to be the stationary distribution ξ_h of the chain P_h, and let $n = t_{\mathrm{mix}} \log\big(\frac{2}{h_{\max} \cdot \min_x \xi_0(x)}\big)$. Note that $\xi_h P_h^n = \xi_h$ in such case. On the other hand, by Lemma 3.4, the total variation distance can be upper bounded as $d_{\mathrm{TV}}\big(\xi_h P_0^n, \xi_0\big) \leq 2^{1 - \frac{n}{t_{\mathrm{mix}}}} \leq h_{\max} \cdot \min_{x \in \mathcal{S}} \xi_0(x)$. Therefore, for any $x \in \mathcal{S}$, we have

$$\left| \frac{\xi_h P_0^n(x)}{\xi_0(x)} - 1 \right| \leq \frac{d_{\mathrm{TV}}\big(\xi_h P_0^n, \xi_0\big)}{\min_{x \in \mathcal{S}} \xi_0(x)} \leq h_{\max} < \frac{1}{2}.$$

Invoking the inequality $|\log z| \leq 2|z - 1|$ for $|z| \leq 1/2$, we can translate the bound into a max-divergence bound

$$D_\infty\left(\xi_h P_0^n \mid\mid \xi_0\right) = \max_{x \in \mathcal{S}} \left| \log \frac{\xi_h P_0^n(x)}{\xi_0(x)} \right| \leq 2h_{\max}.$$

Finally, applying the triangle inequality yields

$$\begin{aligned} D_\infty\left(\xi_h \mid\mid \xi_0\right) &\leq D_\infty\left(\xi_h P_h^n \mid\mid \xi_h P_0^n\right) + D_\infty\left(\xi_h P_0^n \mid\mid \xi_0\right) \\ &\leq (n+2) h_{\max} \ \leq \ t_{\mathrm{mix}} \big(2 + \log h_{\max}^{-1} + \log \tfrac{1}{\min_x \xi_0(x)}\big) h_{\max}, \end{aligned}$$

which proves the second claim.

B.5.2 Proof of Lemma B.2

We first consider the functional $h \mapsto \bar{\theta}(P_h) := \big(I - \mathbb{E}_{\xi_h}[\boldsymbol{L}(s)]\big)^{-1} \mathbb{E}_{\xi_h}[\boldsymbol{b}(s)]$. Note that the stationary distribution ξ_h satisfies the identity $\xi_h P_h = \xi_h$. Taking derivatives, we obtain the following equality for all $x, y \in \mathcal{S}$:

$$\frac{\partial \xi_h}{\partial h_x(y)} \cdot (I - P_h) = \xi_h \cdot \frac{\partial P_h}{\partial h_x(y)} = \xi_h(x) P_h(x, y) \cdot \big[\mathbf{1}_{z=y} - P_h(x, z)\big]_{z \in \mathcal{S}}.$$

Note that the linear operator $(I - P_h)$ is invertible on the subspace $\mathbb{H}_h$. For any $f \in \mathbb{H}_h$, we have

$$\begin{aligned} \frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[f(s)\big] &= \sum_{z \in \mathcal{S}} \frac{\partial \xi_h(z)}{\partial h_x(y)} \cdot f(s) \\ &= \xi_h(x) P_h(x, y) \cdot \big[\mathbf{1}_{z=y} - P_h(x, z)\big]_{z \in \mathcal{S}} \cdot \big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h} \cdot f. \end{aligned}$$

In the above expression, the notation $\big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h}$ denotes the inverse of the operator $I - P_h$ within the subspace $\mathbb{H}_h$, a bounded linear operator on this space. Note that the derivative is invariant under translation. For any $f \in \mathbb{R}^{\mathcal{S}}$, define the auxiliary function $\widetilde{f} := f - \mathbb{E}_{\xi_h}[f]$, and write

$$\begin{aligned} &\frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[f(s)\big] \\ &= \frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[\widetilde{f}(s)\big] = \xi_h(x) P_h(x, y) \cdot \big[\mathbf{1}_{z=y} - P_h(x, z)\big]_{z \in \mathcal{S}} \cdot \big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h} \cdot \widetilde{f} \\ &= \xi_h(x) P_h(x, y) \cdot \big[\mathbf{1}_{z=y} - P_h(x, z)\big]_{z \in \mathcal{S}} \cdot \big(I - P_h\big)^{-1}\big|_{\mathbb{H}_h} \cdot \big(f - \mathbb{E}_{\xi_h}[f]\big) \\ &= \xi_h(x) P_h(x, y) \cdot \big(\mathcal{A}_h f(y) - \sum_{z \in \mathcal{S}} P_h(x, z) \mathcal{A}_h f(z)\big). \end{aligned} \tag{B.25}$$

On the other hand, we can express the desired functional $\bar{\theta}(P_h)$ in the form above. In particular, setting $\bar{L}^{(h)} := \mathbb{E}_{\xi_h}\big[\boldsymbol{L}(s)\big]$ and $\bar{b}^{(h)} := \mathbb{E}_{\xi_h}\big[\boldsymbol{b}(s)\big]$, we see that for any $x, y \in \mathcal{S}$, we have

$$\begin{aligned} \frac{\partial \bar{\theta}(P_h)}{\partial h_x(y)} &= \big(I - \bar{L}^{(h)}\big)^{-1} \frac{\partial \bar{L}^{(h)}}{\partial h_x(y)} \big(I - \bar{L}^{(h)}\big)^{-1} \bar{b}^{(h)} + \big(I - \bar{L}^{(h)}\big)^{-1} \frac{\partial \bar{b}^{(h)}}{\partial h_x(y)} \\ &= \big(I - \bar{L}^{(h)}\big)^{-1} \big(\big(\frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[\boldsymbol{L}(s)\big]\big) \cdot \bar{\theta}(P_h) + \frac{\partial}{\partial h_x(y)} \mathbb{E}_{\xi_h}\big[\boldsymbol{b}(s)\big]\big). \end{aligned}$$

Following the formula (B.25), we conclude that

$$\begin{aligned}\frac{\partial\bar{\theta}(P_h)}{\partial h_x(y)} &= \xi_h(x)P_h(x,y)\big(I-\bar{L}^{(h)}\big)^{-1}\big[\mathcal{A}_h\big(\boldsymbol{L}(y)\bar{\theta}(P_h)+\boldsymbol{b}(y)\big)\big]\\ &\qquad - \xi_h(x)P_h(x,y)\sum_{z\in\mathcal{S}}P_h(x,z)\big(I-\bar{L}^{(h)}\big)^{-1}\big[\mathcal{A}_h\big(\boldsymbol{L}(z)\bar{\theta}(P_h)+\boldsymbol{b}(z)\big)\big]. \qquad \text{(B.26)}\end{aligned}$$

Recall the shorthand notation from before, where for each $s\in\mathcal{S}$, we defined

$$\boldsymbol{g}_h(s) = \big(I-\bar{L}^{(h)}\big)^{-1}\big[\mathcal{A}_h\big(\boldsymbol{L}(s)\bar{\theta}(P_h)+\boldsymbol{b}(s)\big)\big].$$

Given $w\in\mathbb{R}^d$, if we parameterize the perturbation as $h=Qw$, the chain rule yields

$$\begin{aligned}&\nabla_w\bar{\theta}(P_h) = Q^\top\cdot\nabla_h\bar{\theta}(P_h)\\ &= \sum_{x\in\mathcal{S}}\xi_h(x)\Big(\sum_{y\in\mathcal{S}}P_h(x,y)\boldsymbol{g}(y)q_x(y)^\top - \Big(\sum_{y\in\mathcal{S}}P_h(x,y)\boldsymbol{g}(y)\Big)\Big(\sum_{y\in\mathcal{S}}P_h(x,y)\boldsymbol{g}_h(y)q_x(y)\Big)^\top\Big)\\ &= \mathbb{E}_{X\sim\xi_h}\big[\operatorname{cov}_{Y\sim P_h(X,\cdot)}\big(\boldsymbol{g}_h(Y)-\mathcal{P}_h\boldsymbol{g}_h(X), q_X(Y)\mid X\big)\big],\end{aligned}$$

as claimed. □

B.5.3 Proof of Lemma B.3

The following technical lemma is used throughout the proof, and proved in Section B.5.6.

Lemma B.6. *Given a perturbation vector w satisfying $\|w\|_2\le\frac{1-\kappa}{2ct_{\mathrm{mix}}\sigma_L\sqrt{d\cdot|||\Lambda|||_{op}}\log d}$, for $h=Qw$, the matrix $I-\bar{L}^{(h)}$ is invertible, with $|||\big(I-\bar{L}^{(h)}\big)^{-1}|||_{op}\le\frac{2}{1-\kappa}$.*

Before proceeding with the proof, we note two direct consequences of Lemma B.7 from Section B.5.7. First, by taking $f(x):=\langle e_j,\,\boldsymbol{L}(x)u\rangle$ and $f(x):=\langle e_j,\,\boldsymbol{b}(x)\rangle$, applying the tail assumption 3.2 and the boundedness assumption 3.4, we have the following second moment estimate for any $u\in\mathbb{S}^{d-1}$ and $j\in[d]$:

$$\mathbb{E}_{X\sim\xi_h}\big[\langle e_j,\,\mathcal{A}_h\boldsymbol{L}(X)u\rangle^2\big]\le ct_{\mathrm{mix}}^2\sigma_L^2\log^2 d,\quad\text{and}\qquad\text{(B.27a)}$$

$$\mathbb{E}_{X\sim\xi_h}\big[\langle e_j,\,\mathcal{A}_h\boldsymbol{b}(X)\rangle^2\big]\le ct_{\mathrm{mix}}^2\sigma_b^2\log^2 d.\qquad\text{(B.27b)}$$

Second, by taking $f_j(x):=\langle e_j,\,\boldsymbol{L}(x)\bar{\theta}(P_h)+\boldsymbol{b}(x)\rangle$, for any integer $p\ge1$ and $K>0$, Markov's inequality yields the bound

$$\mathbb{P}_{X\sim\xi_h}\Big[\mathcal{A}_hf_j(X)\ge K\Big]\le K^{-2p}\mathbb{E}_{X\sim\xi_h}\big[\mathcal{A}_hf_j(X)^{2p}\big]\le\Big(\tfrac{cp^2t_{\mathrm{mix}}(\sigma_L\|\bar{\theta}\|_2+\sigma_b)\log d}{K}\Big)^{2p}.$$

By taking $K=2cp^2t_{\mathrm{mix}}(\sigma_L\|\bar{\theta}\|_2+\sigma_b)\log d$ and $p=-2\log\min_{x\in\mathcal{S}}\xi_0(x)$, we find that

$$\mathbb{P}_{X\sim\xi_h}\Big[\mathcal{A}_hf_j(X)\ge 8ct_{\mathrm{mix}}(\sigma_L\|\bar{\theta}\|_2+\sigma_b)\log^3\big(\tfrac{d}{\min_{x\in\mathcal{S}}\xi_0(x)}\big)\Big]<\frac{1}{2}\min_{x\in\mathcal{S}}\xi_0(x)\le\min_{x\in\mathcal{S}}\xi_h(x),$$

Since ξ_h is a discrete measure, this high-probability bound implies a deterministic bound

$$\mathcal{A}_h f_j(x) \leq 8 c t_{\mathrm{mix}} (\sigma_L \|\bar{\theta}\|_2 + \sigma_b) \log^3 \big(\tfrac{d}{\min_{x' \in \mathcal{S}} \xi_0(x')}\big) \qquad \text{for all } x \in \mathcal{S}.$$

Combining the estimates for all j coordinates yields the bound

$$\max_{x \in \mathcal{S}} \|\boldsymbol{g}_h(x)\|_2 \leq \tfrac{1}{1-\kappa} \max_{x \in \mathcal{S}} \|\mathcal{A}_h \big[f_j(x)\big]_{j \in [d]}\|_2 \leq \tfrac{c t_{\mathrm{mix}} (\sigma_L \|\bar{\theta}\|_2 + \sigma_b) \sqrt{d}}{1-\kappa} \log^3 \big(\tfrac{d}{\min_{x \in \mathcal{S}} \xi_0(x)}\big). \tag{B.28}$$

Given the two lemmas and facts derived above, we now proceed to the proof of Lemma B.3. Taking derivatives on both sides of equation (B.15), we obtain

$$\begin{aligned}
\nabla_w \boldsymbol{g}_h(z) = {} & \big(I_d - \bar{L}^{(h)}\big)^{-1} \cdot \mathcal{A}_h \boldsymbol{L}(z) \cdot \nabla_w \bar{\theta}(P_h) \\
& + \big(I_d - \bar{L}^{(h)}\big)^{-1} \cdot \big(\nabla_w \mathcal{A}_h\big) \big(\boldsymbol{L}(z) \bar{\theta}(P_h) + \boldsymbol{b}(z)\big) \\
& - \big(I_d - \bar{L}^{(h)}\big)^{-1} \nabla_w \big(\bar{L}^{(h)}\big) \big(I_d - \bar{L}^{(h)}\big)^{-1} (\mathcal{A}_h \boldsymbol{L}(z) \cdot \bar{\theta}(P_h) + \mathcal{A}_h \boldsymbol{b}(z)) \\
=: {} & J_1(h, z) + J_2(h, z) + J_3(h, z).
\end{aligned}$$

We then have the integral relation

$$\boldsymbol{g}_h(z) - \boldsymbol{g}_0(z) = \int_0^1 \nabla_w \boldsymbol{g}_{sh}(z) \cdot w \; ds = \int_0^1 \big(J_1(sh, z) + J_2(sh, z) + J_3(sh, z)\big) \cdot w \; ds.$$

It thus suffices to prove individual upper bounds on the terms $J_1(sh, z) \cdot w$, $J_2(sh, z) \cdot w$ and $J_3(sh, z) \cdot w$.

Bounds on the term $J_1(sh, z) \cdot w$: Invoking Lemma B.2, we have

$$\nabla_w \bar{\theta}(P_h) = \mathbb{E}_{X \sim \xi_h, Y \sim P_h(X, \cdot)} \big[\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h \boldsymbol{g}_h(X)\big) \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top\big].$$

Consequently,

$$\begin{aligned}
& \|\nabla_w \bar{\theta}(P_h) w\|_2 \\
& \leq \| \operatorname{cov}_{X \sim \xi_h, Y \sim P_h(X, \cdot)} \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big) \cdot w\|_2 \\
& + \|\mathbb{E}_{X \sim \xi_h} \big[\operatorname{cov}_{Y \sim P_h(X, \cdot)} \big(\boldsymbol{g}_h(Y) - \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X) + \mathcal{P}_0 \boldsymbol{g}_0(X), \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)\big] w\|_2.
\end{aligned}$$

For perturbation matrix h satisfying the condition $\max_{x \in \mathcal{S}} \|h_x\|_\infty \leq \frac{1}{128 t_{\mathrm{mix}}}$, Lemma B.1 implies the sandwich relations

$$\tfrac{1}{2} \xi_0 \preceq \xi_h \preceq \tfrac{3}{2} \xi_0, \quad \text{and} \quad \tfrac{1}{2} P_0(x) \preceq P_h(x, \cdot) \preceq \tfrac{3}{2} P_0(x), \quad \text{for all } x \in \mathcal{S}.$$

For the first term in above decomposition, we have

$$\begin{aligned}
& \| \operatorname{cov}_{X \sim \xi_h, Y \sim P_h(X, \cdot)} \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big) \cdot w\|_2 \\
& \leq \tfrac{3}{2} \| \operatorname{cov}_{X \sim \xi_0, Y \sim P_0(X, \cdot)} \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big) \cdot w\|_2 \\
& = \tfrac{3}{2} \|\Lambda w\|_2 \leq \tfrac{3}{2} \sqrt{\operatorname{trace}(\Lambda) / n},
\end{aligned}$$

where the last inequality is due to the bound (B.16c).

For the second term in the decomposition, we have

$$\begin{aligned}
&\|\mathbb{E}\big[\operatorname{cov}\big(\boldsymbol{g}_h(Y) - \boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) + \mathcal{P}_0\boldsymbol{g}_0(X), \boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) \mid X\big)\big] \cdot w\|_2 \\
&= \sup_{v \in \mathbb{S}^{d-1}} \mathbb{E}\big[\big(\boldsymbol{g}_h(Y) - \boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X) + \mathcal{P}_0\boldsymbol{g}_0(X)\big)^\top v \cdot \big(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X)\big)^\top w\big] \\
&\leq \sup_{v \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}\big[\langle\big(\boldsymbol{g}_h(X) - \boldsymbol{g}_0(X)\big),\, v\rangle^2\big]} \cdot \sqrt{\mathbb{E}\big[\big(\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0\boldsymbol{g}_0(X)\big)^\top w\big)^2\big]} \\
&\leq \tfrac{3}{2}\sqrt{w^\top \Lambda w}\sqrt{\mathbb{E}\|\boldsymbol{g}_h(X) - \boldsymbol{g}_0(X)\|_2^2},
\end{aligned}$$

where $X \sim \xi_h, Y \sim P_h(X, \cdot)$.

By equation (B.16b), on the support of the prior density, we have the bound $w^\top \Lambda w = n^{-1}\psi^\top D^{-1/2} U^\top \Lambda U D^{-1/2}\psi \leq \frac{d}{n}$. Consequently, we have the upper bound

$$\|\nabla_w \bar{\theta}(P_h) w\|_2 \leq \tfrac{3}{2}\sqrt{\tfrac{\mathrm{trace}(\Lambda)}{n}} + \tfrac{3}{2} \cdot \sqrt{\tfrac{d}{n} \cdot \mathbb{E}_{X \sim \xi_h}\|\boldsymbol{g}_h(X) - \boldsymbol{g}_0(X)\|_2^2}. \tag{B.29}$$

Collecting the bounds above and invoking equation (B.27) and Lemma B.6, we obtain the following bound on the desired term:

$$\begin{aligned}
&\mathbb{E}_{Y \sim \xi_h}\big[\|J_1(\ell h, Y) w\|_2^2\big] \\
&\leq \big\|\big(I_d - \bar{L}^{(\ell h)}\big)^{-1}\big\|_{\mathrm{op}}^2 \cdot \mathbb{E}_{Y \sim \xi_h}\big[\|\mathcal{A}_{\ell h}\boldsymbol{L}(Y) \cdot \nabla_w \bar{\theta}(P_{\ell h}) w\|_2^2\big] \\
&\leq \tfrac{4}{(1-\kappa)^2} \cdot \tfrac{3}{2}\mathbb{E}_{Y \sim \xi_{\ell h}}\big[\|\mathcal{A}_{\ell h}\boldsymbol{L}(Y) \cdot \nabla_w \bar{\theta}(P_{\ell h}) w\|_2^2\big] \\
&\leq \tfrac{6}{(1-\kappa)^2} \cdot c t_{\mathrm{mix}}^2 \sigma_L^2 d \log^2 d \cdot \|\nabla_w \bar{\theta}(P_{\ell h}) w\|_2^2 \\
&\leq \tfrac{c t_{\mathrm{mix}}^2 \sigma_L^2 d \log^2 d}{(1-\kappa)^2} \cdot \tfrac{\mathrm{trace}(\Lambda)}{n} + \tfrac{c t_{\mathrm{mix}}^2 \sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2 n} \sup_{0 \leq \ell \leq 1} \mathbb{E}_{X \sim \xi_{\ell h}}\|\boldsymbol{g}_{\ell h}(X) - \boldsymbol{g}_0(X)\|_2^2.
\end{aligned}$$

Bounds on the term $J_2(sh, z) \cdot w$: For any function $\mathcal{S} \to \mathbb{R}^d$ and $x, y \in \mathcal{S}$, we note that

$$\begin{aligned}
\tfrac{\partial}{\partial h_x(y)}\mathcal{A}_h f &= -(I - \mathcal{P}_h)^{-1}|_{\mathbb{H}_h} \cdot \tfrac{\partial \mathcal{P}_h}{\partial h_x(y)} \cdot (I - \mathcal{P}_h)^{-1}|_{\mathbb{H}_h} f \\
&= -\mathcal{A}_h \cdot \big[\mathbf{1}_{s=x} P_h(x, y) \cdot \big(\mathbf{1}_{s'=y} - P_h(x, s')\big)\big]_{s, s' \in \mathcal{S}} \cdot \mathcal{A} f \\
&= -\mathcal{A}_h \cdot \big[\mathbf{1}_{s=x} P_h(x, y) \cdot \big(\mathcal{A}_h f(y) - \sum_{s'} P_h(x, s')\mathcal{A}_h f(s')\big)\big]_{s \in \mathcal{S}}.
\end{aligned}$$

We can then derive the formula for derivative with respect to the parameter w, as

$$\begin{aligned}
&\big(\nabla_w \mathcal{A}_h\big) f(z) = \sum_{x, y \in \mathcal{S}} \big(\frac{\partial}{\partial h_x(y)}\mathcal{A}_h f(z)\big) \cdot q_x(y)^\top \\
&= -\sum_{x, y \in \mathcal{S}} P_h(x, y)\mathcal{A}_h \mathbf{1}_x(z) \cdot \big(\mathcal{A}_h f(y) - \mathcal{P}_h \mathcal{A}_h f(x)\big) \cdot \big(\boldsymbol{g}_0(y) - \mathcal{P}_0\boldsymbol{g}_0(x)\big)^\top \\
&= -\sum_{x, y \in \mathcal{S}} \sum_{t=0}^{\infty} \big(P_h^t(z, x) - \xi_h(x)\big) P_h(x, y) \big(\mathcal{A}_h f(y) - \mathcal{P}_h \mathcal{A}_h f(x)\big) \big(\boldsymbol{g}_0(y) - \mathcal{P}_0\boldsymbol{g}_0(x)\big)^\top.
\end{aligned}$$

Substituting $f(z) = \boldsymbol{L}(z)\bar{\theta}(P_h) + \boldsymbol{b}(z)$, we note that $\mathcal{A}_h f = \boldsymbol{g}_h$, and consequently,

$$\begin{aligned}
&\big(\nabla_w \mathcal{A}_h\big)\big(\boldsymbol{L}(z)\bar{\theta}(P_h) + \boldsymbol{b}(z)\big) \\
&= \sum_{t=0}^{\infty} \Big(\mathbb{E}_{X\sim P_h^t(z,\cdot), Y\sim P_h(X,\cdot)}\big[\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h \boldsymbol{g}_h(X)\big)\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top\big] \\
&\qquad\qquad - \mathbb{E}_{X\sim \xi_h, Y\sim P_h(X,\cdot)}\big[\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h \boldsymbol{g}_h(X)\big)\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top\big]\Big) \\
&=: \sum_{t=0}^{\infty} D_t(z).
\end{aligned}$$

Next, we estimate the difference term above in two different ways, depending on the value of t. On the one hand, note that

$$\begin{aligned}
&\mathbb{E}_{Z\sim\xi_h}\|\mathbb{E}_{X\sim P_h^t(Z,\cdot), Y\sim P_h(X,\cdot)}\big[\big(\boldsymbol{g}_h(Y) - \mathcal{P}_h \boldsymbol{g}_h(X)\big)\big(\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\big)^\top\big]w\|_2^2 \\
&\leq \sup_{x,y\in\mathcal{S}} \|\boldsymbol{g}_h(y) - \mathcal{P}_h \boldsymbol{g}_h(x)\|_2^2 \cdot \mathbb{E}_{X\sim\xi_h, Y\sim P_h(X,\cdot)}\big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\rangle^2\big] \\
&\leq 4 \sup_{x\in\mathcal{S}} \|\boldsymbol{g}_h(x)\|_2^2 \cdot \mathbb{E}_{X\sim\xi_h, Y\sim P_h(X,\cdot)}\big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\rangle^2\big],
\end{aligned}$$

where the bound for the factor $\sup_{x\in\mathcal{S}} \|\boldsymbol{g}_h(x)\|_2^2$ follows from equation (B.28). For the latter term in the display above, we note that

$$\begin{aligned}
\mathbb{E}_{X\sim\xi_h, Y\sim P_h(X,\cdot)}\big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\rangle^2\big] &\leq 2\mathbb{E}_{X\sim\xi_0, Y\sim P_0(X,\cdot)}\big[\langle w,\, \boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X)\rangle^2\big] \\
&\leq 2 w^\top \Lambda w = \tfrac{2d}{n}.
\end{aligned}$$

Putting together the pieces yields the first estimate

$$\mathbb{E}_{Z\sim\xi_h}\big[\|D_t(Z)w\|_2^2\big] \leq \tfrac{c t_{\mathrm{mix}}^2 \bar{\sigma}^2 d^2}{(1-\kappa)^2 n} \log^6\big(\tfrac{d}{\min_{x\in\mathcal{S}} \xi_0(x)}\big).$$

On the other hand, given $z \in \mathcal{S}$ and the Markov chain $(s_t)_{t\geq 0}$ starting from $s_0 = z$, for any $t > 0$, there exists a random state $\widetilde{s}_t$ such that $\widetilde{s}_t \sim \xi_h$, and we have $\mathbb{P}\big(\widetilde{s}_t \neq s_t\big) \leq 2^{\lfloor \frac{t}{t_{\mathrm{mix}}} \rfloor}$. Define a random variable $\widetilde{s}_{t+1}$ by setting $\widetilde{s}_{t+1} = s_{t+1}$ whenever $s_t = \widetilde{s}_t$, and drawing $\widetilde{s}_{t+1} \sim P(\widetilde{s}_t, \cdot)$ otherwise. From this construction, we have

$$\begin{aligned}
\|D_t(z)w\|_2 &\leq \sup_{u\in\mathbb{S}^{d-1}} \Big\{\mathbb{E}\big[u^\top\big(\boldsymbol{g}_h(s_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(s_t)\big) \cdot w^\top\big(\boldsymbol{g}_0(s_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(s_t)\big) \mid z\big] \\
&\qquad\qquad - \mathbb{E}\big[u^\top\big(\boldsymbol{g}_h(\widetilde{s}_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(\widetilde{s}_t)\big) \cdot w^\top\big(\boldsymbol{g}_0(\widetilde{s}_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(\widetilde{s}_t)\big) \mid z\big]\Big\} \\
&\leq \sup_{u\in\mathbb{S}^{d-1}} \mathbb{E}\big[u^\top\big(\boldsymbol{g}_h(s_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(s_t)\big) \cdot w^\top\big(\boldsymbol{g}_0(s_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(s_t)\big)\mathbf{1}_{s_t\neq\widetilde{s}_t} \mid z\big] \\
&\qquad + \sup_{u\in\mathbb{S}^{d-1}} \mathbb{E}\big[u^\top\big(\boldsymbol{g}_h(\widetilde{s}_{t+1}) - \mathcal{P}_h \boldsymbol{g}_h(\widetilde{s}_t)\big) \cdot w^\top\big(\boldsymbol{g}_0(\widetilde{s}_{t+1}) - \mathcal{P}_0 \boldsymbol{g}_0(\widetilde{s}_t)\big)\mathbf{1}_{s_t\neq\widetilde{s}_t} \mid z\big].
\end{aligned}$$

Applying the Cauchy–Schwarz inequality twice yields

$$
\begin{aligned}
&\mathbb{E}_{Z\sim\xi_h}\big[\|D_t(Z)w\|_2^2\big]\\
&\leq \mathbb{E}\big[\|\boldsymbol{g}_h(s_{t+1})-\mathcal{P}_h\boldsymbol{g}_h(s_t)\|_2^4\big]^{1/2}\cdot\mathbb{E}\big[w^\top\big(\boldsymbol{g}_0(s_{t+1})-\mathcal{P}_0\boldsymbol{g}_0(s_t)\big)^8\big]^{1/4}\cdot\mathbb{E}\big[\mathbf{1}_{s_t\neq\widetilde{s}_t}\big]^{1/4}\\
&\qquad+\mathbb{E}\big[\|\boldsymbol{g}_h(\widetilde{s}_{t+1})-\mathcal{P}_h\boldsymbol{g}_h(s_t)\|_2^4\big]^{1/2}\cdot\mathbb{E}\big[w^\top\big(\boldsymbol{g}_0(\widetilde{s}_{t+1})-\mathcal{P}_0\boldsymbol{g}_0(\widetilde{s}_t)\big)^8\big]^{1/4}\cdot\mathbb{E}\big[\mathbf{1}_{s_t\neq\widetilde{s}_t}\big]^{1/4}\\
&\leq\frac{ct_{\mathrm{mix}}^4}{(1-\kappa)^4}\bar{\sigma}^4d\|w\|_2^2\cdot\log^6 d\cdot 2^{1-\frac{t}{4t_{\mathrm{mix}}}},
\end{aligned}
$$

corresponding to the second estimate.

Finally, setting $\tau=ct_{\mathrm{mix}}\log\frac{t_{\mathrm{mix}}d}{1-\kappa}$ yields

$$
\begin{aligned}
\mathbb{E}_{Z\sim\xi_h}\big[\|\sum_{t=0}^{\infty}D_t(Z)w\|_2^2\big]&\leq\big(\sum_{t=0}^{\infty}e^{-\frac{t}{\tau}}\big)\cdot\big(\sum_{t=0}^{\infty}e^{\frac{t}{\tau}}\mathbb{E}_{Z\sim\xi_h}\big[\|D_t(Z)w\|_2^2\big]\big)\\
&\leq\tfrac{ct_{\mathrm{mix}}^4\bar{\sigma}^2d^2}{(1-\kappa)^2n}\log^6\big(\tfrac{d}{\min_{x\in\mathcal{S}}\xi_0(x)}\big),
\end{aligned}
$$

so that

$$
\mathbb{E}_{Z\sim\xi_h}\big[\|J_2(\ell h,Z)w\|_2^2\big]\leq\tfrac{ct_{\mathrm{mix}}^4\bar{\sigma}^2d^2}{(1-\kappa)^4n}\log^6\big(\tfrac{d}{\min_{x\in\mathcal{S}}\xi_0(x)}\big).
$$

Bounds on the term $J_3(sh,z)\cdot w$: By equation (B.25), for any vector $u\in\mathbb{S}^{d-1}$, we have

$$
\nabla_w\big(\bar{L}^{(h)}u\big)=\sum_{x,y\in\mathcal{S}}\xi_h(x)P_h(x,y)\big(\mathcal{A}_h\bar{L}^{(h)}(y)-\sum_{z\in\mathcal{S}}P_h(x,z)\mathcal{A}_h\bar{L}^{(h)}(z)\big)u\cdot q_x(y)^\top.
$$

For any $z\in\mathcal{S}$, we obtain

$$
\begin{aligned}
&\|\nabla_w\big(\bar{L}^{(h)}\big)\boldsymbol{g}_h(z)w\|_2\\
&=\sup_{u\in\mathbb{S}^{d-1}}\mathbb{E}_{X\sim\xi_h,Y\sim P_h(X,\cdot)}\big[u^\top\big(\mathcal{A}_h\bar{L}^{(h)}(Y)-\mathcal{P}_h\mathcal{A}_h\bar{L}^{(h)}(X)\big)\boldsymbol{g}_h(z)q_X(Y)^\top w\big]\\
&\leq\sup_{u\in\mathbb{S}^{d-1}}\sqrt{\mathbb{E}\big(u^\top\big(\mathcal{A}_h\bar{L}^{(h)}(Y)-\mathcal{P}_h\mathcal{A}_h\bar{L}^{(h)}(X)\big)\boldsymbol{g}_h(z)\big)^2}\cdot\sqrt{\mathbb{E}\big[\big(q_X(Y)^\top w\big)^2\big]}\\
&\leq ct_{\mathrm{mix}}\sigma_L\|\boldsymbol{g}_h(z)\|_2\log d\cdot\sqrt{\tfrac{d}{n}},
\end{aligned}
$$

where the final inequality is due to equation (B.27). Combining with Lemma B.6, we have the bound

$$
\begin{aligned}
\mathbb{E}_{Z\sim\xi_h}\big[\|J_3(\ell h,Z)w\|_2^2\big]&\leq\tfrac{cd^2}{(1-\kappa)^2n}\cdot t_{\mathrm{mix}}^2\sigma_L^2\log^2 d\cdot\mathbb{E}_{Z\sim\xi_h}\big[\|\boldsymbol{g}_h(Z)\|_2^2\big]\\
&\leq\tfrac{c\sigma_L^2\bar{\sigma}^2t_{\mathrm{mix}}^4d^2}{(1-\kappa)^4n}\log^2 d.
\end{aligned}
$$

Finishing the proof. Collecting the bounds for J_1, J_2 and J_3, for $n \geq \frac{ct_{\text{mix}}^2 \sigma_L^2 d^2 \log^2 d}{(1-\kappa)^2}$, we have

$$\sup_{0 \leq \ell \leq 1} \mathbb{E}_{Z \sim \xi_h} \big[\| \boldsymbol{g}_{\ell h}(Z) - \boldsymbol{g}_0(Z) \|_2^2 \big] \leq \tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\text{mix}}^4 d^2}{(1-\kappa)^4 n} \log^6 \big(\tfrac{d}{\min_x \xi_0(x)} \big) \\ + \tfrac{1}{2} \sup_{0 \leq \ell \leq 1} \mathbb{E}_{Z \sim \xi_h} \big[\| \boldsymbol{g}_{\ell h}(Z) - \boldsymbol{g}_0(Z) \|_2^2 \big],$$

which completes the proof of the first claim of the lemma.

For the second claim, we combine the first claim with equation (B.29) and obtain

$$\| \nabla_w \bar{\theta}(P_h) w \|_2 \leq \tfrac{3}{2} \sqrt{\tfrac{\text{trace}(\Lambda)}{n}} + \sqrt{\tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\text{mix}}^4 d^3}{(1-\kappa)^4 n^2} \log^6 \big(\tfrac{d}{\min_x \xi_0(x)} \big)}.$$

Taking the integral yields

$$\| \bar{\theta}(P_h) - \bar{\theta}(P_0) \|_2 \leq \int_0^1 \| \nabla_w \bar{\theta}(P_{\ell h}) w \|_2 d\ell \\ \leq \tfrac{3}{2} \sqrt{\tfrac{\text{trace}(\Lambda)}{n}} + \sqrt{\tfrac{c(1+\sigma_L^2)\bar{\sigma}^2 t_{\text{mix}}^4 d^3}{(1-\kappa)^4 n^2} \log^6 \big(\tfrac{d}{\min_x \xi_0(x)} \big)},$$

which proves the second claim.

B.5.4 Proof of Lemma B.4

We first compute the Fisher information with respect to the perturbation vector h, and then transform this via chain rule into a formula that holds with respect to the parameter w. We are interested in the matrix $I^{(n)}(h) := \mathbb{E}_h \big[\nabla_h \log \mathbb{P}_h(s_0^n) \nabla_h \log \mathbb{P}_h(s_0^n)^\top \big]$. When the Markov chain P_h is run under the initial distribution ξ_0, the joint distribution of the observed trajectory $(s_t)_{t=0}^n$ can be factorized as $\mathbb{P}_h \big(s_0, s_1, \cdots, s_n \big) = \xi_0(s_0) \cdot \prod_{t=1}^n P_h(s_{t-1}, s_t)$.

Let us now study the Fisher information matrix. For any pair $x, y \in \mathcal{S}$ with $P(x, y) > 0$, performing some algebra yields the expression

$$\tfrac{\partial}{\partial h_x(y)} \log \mathbb{P}_h \big(s_0, s_1, \cdots, s_n \big) = \sum_{t=1}^n \mathbf{1}_{s_{t-1}=x} \big(\mathbf{1}_{s_t=y} - P_h(x, y) \big).$$

Consider the natural filtration $\mathcal{F}_t := \sigma(s_0, s_1, \cdots, s_t)$. Note that under the transition kernel P_h, we have the identity

$$\mathbb{E}_h \big[\mathbf{1}_{s_{t-1}=x} \big(\mathbf{1}_{s_t=y} - P_h(x, y) \big) \mid \mathcal{F}_{t-1} \big] = \mathbf{1}_{s_{t-1}=x} \cdot \big(\mathbb{E}_h \big[\mathbf{1}_{s_t=y} \mid s_{t-1} = x \big] - P_h(x, y) \big) = 0.$$

Therefore, the process $\{ \nabla_h \log \mathbb{P}_h(s_0, s_1, \cdots, s_n) \}_{n \geq 0}$ is a martingale adapted to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$. Its second moment is given by

$$S = \mathbb{E} \big[\nabla_h \log \mathbb{P}_h(s_0^n) \cdot \nabla_h^\top \log \mathbb{P}_h(s_0^n) \big] = \sum_{t=1}^n \mathbb{E} \big[\nabla_h \log P_h(s_{t-1}, s_t) \cdot \nabla_h^\top \log P_h(s_{t-1}, s_t) \big].$$

We find that

$$\begin{aligned}S &= \Big[\mathbf{1}_{x_1=x_2} \cdot \sum_{t=1}^{n} \mathbb{E}\big[\mathbf{1}_{x_1=s_{t-1}} \cdot \big(\mathbf{1}_{s_t=y_1} - P_h(x_1,y_1)\big)\big] \cdot \big(\mathbf{1}_{s_t=y_2} - P_h(x_2,y_2)\big)\Big]_{(x_1,y_1),(x_2,y_2)} \\ &= \sum_{t=1}^{n} \operatorname{diag}\Big(\big\{\mathbb{P}_h\big(s_{t-1}=x\big) \cdot P_h(x,y)\big\}_{(x,y)}\Big) \\ &\qquad - \sum_{t=1}^{n} \big[\mathbb{P}_h(s_{t-1}=x) \cdot P_h(x,y_1) \cdot P_h(x,y_2)\big]_{(x,y_1),(x,y_2)}.\end{aligned}$$

Consequently, the Fisher information matrix is a block diagonal matrix $I^{(n)}(h) = \operatorname{diag}\big(\big\{I_x^{(n)}(h)\big\}_{x\in\mathcal{S}}\big)$, where each block matrix $I_x^{(n)}(h) \in \mathbb{R}^{\mathcal{S}\times\mathcal{S}}$ takes the form

$$I_x^{(n)}(h) = \sum_{t=1}^{n} \mathbb{P}_h\big(s_{t-1}=x\big) \cdot \Big[\operatorname{diag}\big(\big\{P_h(x,y)\big\}_{y\in\mathcal{S}}\big) - \big[P_h(x,y)\big]_{y\in\mathcal{S}}\big[P_h(x,y)\big]_{y\in\mathcal{S}}^{\top}\Big].$$

By Lemma B.1, for $h_{\max}$ satisfying the inequality $h_{\max}^{-1} \geq ct_{\mathrm{mix}}\big(\log h_{\max}^{-1} + \log(\min \xi_0)^{-1}\big)$ for some constant $c > 0$, we have the bound $\frac{1}{2}\xi_h \preceq \xi_0 \preceq \frac{3}{2}\xi_h$, and hence $\frac{1}{2}P_h^k\xi_h \preceq P_h^k\xi_0 \preceq \frac{3}{2}P_h^k\xi_h$ for each $k = 0, 1, 2, \ldots$. From this sandwiching, we find that

$$\begin{aligned}I_x^{(n)}(h) &\preceq \tfrac{3}{2}\sum_{t=1}^{n} P_h^{t-1}\xi_h(x) \cdot \Big[\operatorname{diag}\big(\big\{P_h(x,y)\big\}_{y\in\mathcal{S}}\big) - \big[P_h(x,y)\big]_{y\in\mathcal{S}}\big[P_h(x,y)\big]_{y\in\mathcal{S}}^{\top}\Big] \\ &= \tfrac{3n}{2}\xi_h(x)\Big[\operatorname{diag}\big(\big\{P_h(x,y)\big\}_{y\in\mathcal{S}}\big) - \big[P_h(x,y)\big]_{y\in\mathcal{S}}\big[P_h(x,y)\big]_{y\in\mathcal{S}}^{\top}\Big].\end{aligned}$$

Turning to the Fisher information, we compute

$$\begin{aligned}&I^{(n)}(w) = Q^{\top} I^{(n)}(h) Q \\ &\preceq \tfrac{3n}{2}\sum_{x\in\mathcal{S}} \xi_h(x)\Big(\sum_{y\in\mathcal{S}} P_h(x,y)q_x(y)q_x(y)^{\top} - \Big(\sum_{y\in\mathcal{S}} P_h(x,y)q_x(y)\Big)\Big(\sum_{y\in\mathcal{S}} P_h(x,y)q_x(y)\Big)^{\top}\Big) \\ &= \tfrac{3n}{2}\mathbb{E}_{X\sim\xi_h}\Big[\mathbb{E}_{Y\sim P_h(X,\cdot)}\big[q_X(Y)q_X(Y)^{\top}\big] - \mathbb{E}_{Y\sim P_h(X,\cdot)}\big[q_X(Y)\big] \cdot \mathbb{E}_{Y\sim P_h(X,\cdot)}\big[q_X(Y)\big]^{\top}\Big] \\ &= \tfrac{3n}{2}\mathbb{E}_{X\sim\xi_h}\Big[\operatorname{cov}_{P_h(X,\cdot)}\big(q_X(Y) \mid X\big)\Big].\end{aligned}$$

B.5.5 Proof of Lemma B.5

For each $k \in \mathbb{Z}$, by the definition of the Green function, we note that

$$f(s_k) = \mathcal{A}_0 f(s_k) - \mathbb{E}\big[\mathcal{A}_0 f(s_{k+1}) \mid s_k\big] = \mathcal{A}_0 f(s_k) - \mathcal{P}_0\mathcal{A}_0 f(s_k). \tag{B.30}$$

By stationarity, we have

$$\begin{aligned}\sum_{k=-\infty}^{\infty} \mathbb{E}\big[f(s_k)f(s_0)\big] &= \mathbb{E}[f^2(s_0)] + 2\sum_{k=1}^{\infty}\mathbb{E}\big[f(s_k)f(s_0)\big] \\ &\overset{(i)}{=} -\mathbb{E}[f(s_0)^2] + 2\mathbb{E}\Big[f(s_0) \cdot \sum_{k=0}^{\infty}\mathbb{E}\big[f(s_k) \mid s_0\big]\Big]\end{aligned}$$

where step (i) makes use of the dominated convergence theorem, in particular by noting that $\big|\mathbb{E}\big[f(s_k) \mid s_0\big]\big| \leq \|f\|_\infty \cdot 2^{1-k/t_{\mathrm{mix}}}$ from Lemma 3.4. Consequently, we can write

$$\begin{aligned}
&\sum_{k=-\infty}^{\infty} \mathbb{E}\big[f(s_k)f(s_0)\big] \\
&= -\mathbb{E}[f^2(s_0)] + 2\mathbb{E}\big[f(s_0) \cdot \mathcal{A}_0 f(s_0)\big] \\
&\overset{(ii)}{=} -\mathbb{E}\big[\big(\mathcal{A}_0 f(s_0) - \mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big] + 2\mathbb{E}\big[\big(\mathcal{A}_0 f(s_0) - \mathcal{P}_0\mathcal{A}_0 f(s_0)\big) \cdot \mathcal{A}_0 f(s_0)\big] \\
&= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_0)\big)^2\big] - \mathbb{E}\big[\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big],
\end{aligned}$$

where step (ii) follows from equation (B.30).

With $\mathbb{E}$ denoting expectation over $X \sim \xi_0, Y \sim P_0(X, \cdot)$, we have

$$\begin{aligned}
&\mathbb{E}\big[\big(\mathcal{A}_0 f(Y) - \mathcal{P}_0\mathcal{A}_0 f(X)\big)^2\big] \\
&= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_1) - \mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big] \\
&= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_1)\big)^2\big] + \mathbb{E}\big[\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big] - 2\mathbb{E}\big[\big(\mathcal{A}_0 f(s_1)\big) \cdot \big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)\big] \\
&= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_0)\big)^2\big] + \mathbb{E}\big[\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big] - 2\mathbb{E}\big[\mathbb{E}\big[\mathcal{A}_0 f(s_1) \mid s_0\big] \cdot \big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)\big] \\
&= \mathbb{E}\big[\big(\mathcal{A}_0 f(s_0)\big)^2\big] - \mathbb{E}\big[\big(\mathcal{P}_0\mathcal{A}_0 f(s_0)\big)^2\big],
\end{aligned}$$

and combining the pieces completes the proof of this lemma. □

B.5.6 Proof of Lemma B.6

By following the derivation of equation (B.25), we find that

$$\tfrac{\partial}{\partial h_x(y)} \bar{L}^{(h)} = \xi_h(x) P_h(x, y) \Big\{ \mathcal{A}_h \boldsymbol{L}(y) - \sum_{z \in \mathcal{S}} P_h(x, z) \mathcal{A} \boldsymbol{L}(z) \Big\}.$$

Consequently, for any $u \in \mathbb{S}^{d-1}$, we have the bound

$$\begin{aligned}
&\big\|\nabla_w \big(\bar{L}^{(h)} u\big)\big\|_{\mathrm{op}} \\
&\leq \sup_{z, v \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}_{Y \sim \xi_h}\big[\big(z^\top \mathcal{A}_h \boldsymbol{L}(Y) u\big)^2\big]} \cdot \sqrt{\mathbb{E}_{X \sim \xi_h, Y \sim P_h(X, \cdot)}\big[\big((\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X))^\top v\big)^2\big]} \\
&\leq \sup_{v \in \mathbb{S}^{d-1}} \sqrt{\mathbb{E}_{Y \sim \xi_h}\big[\|\mathcal{A}_h \boldsymbol{L}(Y) u\|_2^2\big]} \cdot \frac{3}{2} \sqrt{\mathbb{E}_{X \sim \xi_0, Y \sim P_0(X, \cdot)}\big[\big((\boldsymbol{g}_0(Y) - \mathcal{P}_0 \boldsymbol{g}_0(X))^\top v\big)^2\big]} \\
&\leq c t_{\mathrm{mix}} \sigma_L \sqrt{d \cdot \|\Lambda\|_{\mathrm{op}}} \log d.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
\|\bar{L}^{(h)} - \bar{L}^{(0)}\|_{\mathrm{op}} \leq \sup_{u \in \mathbb{S}^{d-1}} \big\|\big(\bar{L}^{(h)} - \bar{L}^{(0)}\big) u\big\|_2 &\leq \int_0^1 \sup_{u \in \mathbb{S}^{d-1}} \big\|\nabla_w \big(\bar{L}^{(sQw)} u\big) \cdot w\big\|_2 ds \\
&\leq c t_{\mathrm{mix}} \sigma_L \sqrt{d \cdot \mathrm{trace}(\Lambda)} \log d \cdot \|w\|_2.
\end{aligned}$$

Now given a perturbation vector satisfying the bound $\|w\|_2 \leq \frac{1-\kappa}{2ct_{\mathrm{mix}}\sigma_L\sqrt{d\cdot\|\Lambda\|_{\mathrm{op}}}\log d}$, we have the following bound for any $u \in \mathbb{S}^{d-1}$:

$$\|(I-\bar{L}^{(h)})u\|_2 \geq \|(I-\bar{L}^{(0)})u\|_2 - \|(\bar{L}^{(h)}-\bar{L}^{(0)})u\|_2 \geq (1-\kappa) - \|\bar{L}^{(h)}-\bar{L}^{(0)}\|_{\mathrm{op}} \geq \tfrac{1-\kappa}{2},$$

which implies that $\|I-\bar{L}^{(h)})^{-1}\|_{\mathrm{op}} \leq \frac{2}{1-\kappa}$, as claimed.

B.5.7 A useful moment bound

Finally, we state and prove a moment bound that is useful in multiple proofs. Recall that the operator $\mathcal{P}_h$ is a the perturbed probability transition kernel under perturbation matrix h, and the operator $\mathcal{A}_h$ is the Green function operator associated with this transition kernel.

Lemma B.7. *Consider a bounded function $f : \mathcal{S} \to \mathbb{R}$, and a perturbation vector h satisfying the condition in Lemma B.1. There there exists a universal constant $c > 0$, such that for any integer $p \geq 1$*

$$\Big(\mathbb{E}_{X\sim\xi_h}\big[\big(\mathcal{A}_h f(X)\big)^{2p}\big]\Big)^{\frac{1}{2p}} \leq c\,p\,t_{\mathrm{mix}}\Big[\mathbb{E}_{X\sim\xi_h}\big[f(X)^{2p}\big]\Big]^{\frac{1}{2p}} \log\Big\{\frac{\|f\|_\infty^{2p}}{\mathbb{E}_{X\sim\xi_h}\big[f(X)^{2p}\big]}\Big\}$$

The proof is similar to that of Lemma 3.7. For any function $f : \mathcal{S} \to \mathbb{R}$ such that $\mathbb{E}_{\xi_h}[f(X)] = 0$, we first observe that $\mathcal{A}_h f(s) = \sum_{k=0}^{\infty} \mathcal{P}_h^k f(s)$ for all $s \in \mathcal{S}$. Note that Lemma B.1 guarantees that the perturbed chain satisfies Assumption 3.1 with mixing time $4t_{\mathrm{mix}}$. By Lemma 3.4 and the coupling definition of total variation distance, for each $t \geq 0$, there exists a random variable $\widetilde{s}_t$ such that $\widetilde{s}_t \mid s_0 \sim \xi_h$, and $\mathbb{P}\big(\widetilde{s}_t \neq s_t \mid s\big) \leq 2^{-\lfloor \frac{t}{4t_{\mathrm{mix}}} \rfloor}$.

By construction, the state $\widetilde{s}_t$ is independent of s. Consequently, we have the equivalence $\mathcal{A}_h f(s) = \sum_{k=0}^{\infty} \mathbb{E}\big[f(s_k) - f(\widetilde{s}_k) \mid s\big]$, and for any $\alpha > 0$,

$$\begin{aligned}\mathbb{E}_{s\sim\xi_h}\big[\big(\mathcal{A}_h f(s)\big)^{2p}\big] &\leq \Big(\sum_{k=0}^{\infty} e^{2p\alpha t}\mathbb{E}\big(\mathbb{E}\big[f(s_k) - f(\widetilde{s}_k) \mid s\big]\big)^{2p}\Big) \cdot \Big(\sum_{k=0}^{\infty} e^{-\frac{2p}{2p-1}\alpha k}\Big)^{2p-1} \\ &\leq \alpha^{1-2p} \sum_{k=0}^{\infty} e^{2p\alpha k}\mathbb{E}\big[\,|f(s_k) - f(\widetilde{s}_k)|^{2p}\,\big].\end{aligned}$$

We bound the moment of $f(s_k) - f(\widetilde{s}_k)$ for different values of k in two ways. On the one hand, Young's inequality directly leads to the following naive bound

$$\mathbb{E}\big[\,|f(s_k) - f(\widetilde{s}_k)|^{2p}\,\big] \leq 2^{2p-1}\big(\mathbb{E}\big[f(s_k)^{2p}\big] + \mathbb{E}\big[f(\widetilde{s}_k)^{2p}\big]\big) = 2^{2p}\mathbb{E}_{s\sim\xi_h}\big[f(s)^{2p}\big].$$

On the other hand, for any bounded function f, we have

$$\mathbb{E}\big[\,|f(s_k) - f(\widetilde{s}_k)|^{2p}\,\big] \leq \|f\|_\infty^{2p} \cdot \mathbb{P}\big(s_k \neq \widetilde{s}_k\big) \leq \|f\|_\infty^{2p} \cdot 2^{1-\frac{k}{4t_{\mathrm{mix}}}}.$$

Combining the two estimates yields the bound

$$\mathbb{E}\big[(\mathcal{A}_h f(X))^{2p}\big] \leq \alpha^{1-2p}\Big\{2^{2p} \cdot e^{2p\alpha\tau}\tau\mathbb{E}_{s\sim\xi_h}\big[f(s)^{2p}\big] + \|f\|_\infty^{2p} \sum_{k=\tau+1}^{\infty} e^{2p\alpha k} \cdot 2^{1-\frac{k}{4t_{\mathrm{mix}}}}\Big\},$$

valid for any $\alpha > 0$ and $\tau > 0$. Setting $\tau = c\, t_{\mathrm{mix}} \log \frac{\|f\|_\infty^{2p}}{\mathbb{E}[f(X)^{2p}]}$ and $\alpha = \frac{1}{16\tau p}$ yields the claim.

B.6 Proofs for the examples

We collect the proofs of the consequences to specific examples in this section.

B.6.1 Proofs for TD(0)

We stated three corollaries applicable to this method, and in this section, we prove each of them in turn.

B.6.1.1 Proof of Corollary 3.1

The bulk of the proof involves verifying the conditions needed to apply Proposition 3.1 and Theorem 3.1, but some additional care is needed in order to deal with non-orthonormal basis functions $(\phi_j)_{j\in[d]}$. First, we note that the SA procedure (3.27) can be equivalently written as

$$\theta_{t+1} = (1-\eta\beta)\theta_t + \eta\beta\boldsymbol{L}_{t+1}(\Omega_t)\theta_t - \eta\beta\boldsymbol{b}_{t+1}(\Omega_t), \tag{B.31}$$

where $\boldsymbol{L}_{t+1}(\Omega_t) := \big(I_d - \beta^{-1}\phi(s_t)\phi(s_t)^\top + \gamma\beta^{-1}\phi(s_t)\phi(s_{t+1})^\top\big)$, and $\boldsymbol{b}_{t+1}(\Omega_t) := \beta^{-1}R_t(s_t)\phi(s_t)$. This is an SA scheme with stepsize $\eta\beta$.

For any matrix $A \in \mathbb{R}^{d\times d}$, define $\kappa(A) := \frac{1}{2}\lambda_{\max}\big(A + A^\top\big)$. We verify the eigenvalue condition (3.4) by noting that

$$\begin{aligned}\tfrac{1}{2}\lambda_{\max}\big(\bar{L}+\bar{L}^\top\big) &= 1 - \tfrac{1}{\beta}\,\kappa\big(\gamma\mathbb{E}_{s\sim\xi,s^+\sim P(s,\cdot)}\big[\phi(s)\phi(s^+)^\top\big] - \mathbb{E}_\xi\big[\phi(s)\phi(s)^\top\big]\big)\\ &= 1 - \tfrac{1}{\beta}\lambda_{\max}\big(B^{1/2}\big(I_d - \tfrac{M+M^\top}{2}\big)B^{1/2}\big) \;=\; 1 - \tfrac{\mu}{\beta}(1-\kappa) < 1,\end{aligned}$$

and

$$\|\bar{L}\|_{\mathrm{op}} \leq 1 + \tfrac{1}{\beta}\big(\|\mathbb{E}_{s\sim\xi,s^+\sim P(s,\cdot)}\big[\phi(s)\phi(s^+)^\top\big]\|_{\mathrm{op}} + \|\mathbb{E}_\xi\big[\phi(s)\phi(s)^\top\big]\|_{\mathrm{op}}\big) \leq 3.$$

For the two-step sliding-window Markov chain $\Omega_t = (s_t, s_{t+1})$, Assumption 3.1 holds with mixing time $(t_{\mathrm{mix}}+1)$ in the discrete metric, and the metric space has diameter at most 1. It remains to verify the boundedness and moment assumptions.

In order to verify Assumption 3.4, we note that the bounds (3.28a) imply that

$$
\begin{aligned}
&|\!|\!|\boldsymbol{L}_{t+1}(s_t)|\!|\!|_{\mathrm{op}} \leq 1 + \tfrac{1}{\beta}\big(|\!|\!|\phi(s_t)\phi(s_{t+1})|\!|\!|_{\mathrm{op}} + |\!|\!|\phi(s_t)\phi(s_t)^\top|\!|\!|_{\mathrm{op}}\big) \leq (1+\varsigma^2)d, \quad \text{and}\\
&\quad\|\boldsymbol{b}_{t+1}(s_t)\|_2 \leq \tfrac{1}{\beta}|R_t(s_t)| \cdot \|\phi(s_t)\|_2 \leq \varsigma^2\sqrt{d/\beta}.
\end{aligned}
$$

Turning to the moment assumption, given any vector $u \in \mathbb{S}^{d-1}$ and coordinate vector e_j, we have the bounds

$$
\begin{aligned}
\mathbb{E}_{s\sim\xi, s^+\sim P(s,\cdot)}\big[\big(e_j^\top\phi(s)\phi(s^+)^\top u\big)^2\big] &\leq \sqrt{\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top\phi(s)\big)^4\big]}\cdot\sqrt{\mathbb{E}_{s\sim\xi}\big[\big(u^\top\phi(s)\big)^4\big]} \leq \beta^2\varsigma^4,\\
\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top\phi(s)\phi(s)^\top u\big)^2\big] &\leq \sqrt{\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top\phi(s)\big)^4\big]}\cdot\sqrt{\mathbb{E}_{s\sim\xi}\big[\big(u^\top\phi(s)\big)^4\big]} \leq \beta^2\varsigma^4,\\
\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top R_t(s)\phi(s)\big)^2\big] &\leq \varsigma^2\mathbb{E}_{s\sim\xi}\big[\big(e_j^\top\phi(s)\big)^2\big] \leq \beta\varsigma^4.
\end{aligned}
$$

Finally, the quantity $\bar{\sigma}$ from equation (3.29) is bounded as

$$
\begin{aligned}
&\max_{j\in[d]}\mathbb{E}\big[\langle e_j,\, (\boldsymbol{L}_{t+1}(\Omega_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}_{t+1}(\Omega_t) - b)\rangle^2\big]\\
&\qquad\leq \max_{j\in[d]}\sqrt{\mathbb{E}\big[\langle e_j,\, \phi(s_t)\rangle^4\big]}\cdot\sqrt{\big(\mathbb{E}\big[\phi(s_t)^\top\bar{\theta} - \gamma\phi(s_{t+1})^\top\bar{\theta} - R_t(s_t)\big)^4\big]} \leq \bar{\sigma}^2.
\end{aligned}
$$

Invoking equation (3.73) with the test matrix $Q := B$ and substituting with the representation $v(s) = \langle\theta,\, \phi(s)\rangle$ yields the claim.

B.6.1.2 Proof of Corollary 3.2

We prove this corollary by verifying the assumptions used in our main theorem. Assumption 3.2 directly follows from (3.33c) and the boundedness of reward; Assumption 3.1 is exactly the $\mathcal{W}_1$ mixing time bound imposed on the Markov chain. In order to verify that $\boldsymbol{L}(s, s^+) = I_d - \beta^{-1}\big(\phi(s)\phi(s)^\top - \gamma\phi(s)\phi(s^+)^\top\big)$ satisfies Assumption 3.4, we first note that

$$
\begin{aligned}
&|\!|\!|\boldsymbol{L}(s_1, s_1^+) - \boldsymbol{L}(s_2, s_2^+)|\!|\!|_{\mathrm{op}}\\
&\qquad\leq \tfrac{1}{\beta}|\!|\!|\phi(s_1)\phi(s_1)^\top - \phi(s_2)\phi(s_2)^\top|\!|\!|_{\mathrm{op}} + \tfrac{\gamma}{\beta}|\!|\!|\phi(s_1)\phi(s_1^+)^\top - \phi(s_2)\phi(s_2^+)^\top|\!|\!|_{\mathrm{op}}.
\end{aligned}
$$

By adding and subtracting terms, we have the bound

$$
\begin{aligned}
|\!|\!|\phi(s_1)\phi(s_1)^\top - \phi(s_2)\phi(s_2)^\top|\!|\!|_{\mathrm{op}} &\leq \big\{\|\phi(s_1)\|_2 + \|\phi(s_2)\|_2\big\}\ \|\phi(s_1) - \phi(s_2)\|_2\\
&\overset{(i)}{\leq} 2\varsigma^2\beta d\|s_1 - s_2\|_2,
\end{aligned}
$$

The step (i) follows from the Lipschitz condition (3.33b) and boundedness of the metric space $\mathcal{S}$. More precisely, we have $\|\phi(s_1) - \phi(s_2)\|_2 \leq \varsigma\sqrt{\beta d}\|s_1 - s_2\|_2$ and $\|\phi(s_1)\|_2 = \|\phi(s_1) - \phi(0)\|_2 \leq \varsigma\sqrt{\beta d}$. A similar argument yields that

$$
|\!|\!|\phi(s_1)\phi(s_1^+)^\top - \phi(s_2)\phi(s_2^+)^\top|\!|\!|_{\mathrm{op}} \leq \varsigma^2 d\big(\|s_1^+ - s_2^+\|_2 + \|s_1 - s_2\|_2\big).
$$

Putting together the pieces, we have shown that the mapping $L : \mathcal{S} \to \mathbb{R}^{d\times d}$ is $3\varsigma^2 d$-Lipschitz with respect to the metric $\rho\big((s_1, s_1^+), (s_2, s_2^+)\big) = \|s_1 - s_2\|_2 + \|s_1^+ - s_2^+\|_2$.

Similarly, for the vector observation $\boldsymbol{b}_t(s) = R_t(s)\phi(s)$, we note that for any $s_1, s_2 \in \mathcal{S}$,

$$\begin{aligned}\|\boldsymbol{b}_t(s_1) - \boldsymbol{b}_t(s_2)\|_2 &\le |R_t(s_1) - R_t(s_2)| \cdot \|\phi(s_1)\|_2 + |R_t(s_2)| \cdot \|\phi(s_1) - \phi(s_2)\|_2 \\ &\le 2\varsigma\sqrt{d/\beta}\|\phi(s_1) - \phi(s_2)\|_2,\end{aligned}$$

which shows that $b : \mathcal{S} \to \mathbb{R}^{d/\beta}$ is $2\varsigma^2\sqrt{d}$-Lipschitz. Having verified the assumptions, we complete the proof by following the same steps as in the proof as Corollary 3.1.

B.6.1.3 Proof of Corollary 3.3

In order to verify that Assumption 3.4 holds with respect to the discrete metric, note that for any $d_n \ge 1$, we have $\|\boldsymbol{b}_t(s)\|_2 \le \frac{\varsigma}{\beta}\sqrt{\sum_{j=1}^{d_n} \phi_j^2(s)} \le \frac{\varsigma^2}{\beta}\sqrt{d_n}$, and

$$\|\boldsymbol{L}(s_1, s_2)\|_{\text{op}} \le 1 + \tfrac{1}{\beta}\sum_{j=1}^{d_n} \phi_j^2(s_1) + \tfrac{1}{\beta}\sqrt{\sum_{j=1}^{d_n} \phi_j^2(s_1)} \cdot \sqrt{\sum_{j=1}^{d_n} \phi_j^2(s_2)} \le \tfrac{1+\varsigma^2}{\beta} d_n.$$

Turning to the moment condition, let $\mathbb{E}$ denote expectation over a pair $s \sim \xi$ and $s^+ \sim P(s, \cdot)$. Then for any vector $u \in \mathbb{S}^{d_n - 1}$ and index $j \in [d_n]$, we have

$$\begin{aligned}\mathbb{E}\big[\langle e_j,\, \boldsymbol{L}(s, s^+)u\rangle^2\big] &\le 3 + \tfrac{3}{\beta^2}\mathbb{E}\Big[\big(\langle e_j,\, \phi(s)\rangle\, \langle \phi(s^+),\, u\rangle\big)^2\Big] + \tfrac{3}{\beta^2}\mathbb{E}\Big[\big(\langle e_j,\, \phi(s)\rangle\, \langle \phi(s),\, u\rangle\big)^2\Big] \\ &\le 3 + \tfrac{6}{\beta^2}\|\phi_j\|_\infty^2 \cdot \mathbb{E}\big[\langle \phi(s),\, u\rangle^2\big] \\ &\le 3 + \tfrac{6}{\beta}\varsigma^2.\end{aligned}$$

For each $t = 1, 2, \ldots$, we also have $\mathbb{E}\big[\langle e_j,\, \boldsymbol{b}_{t+1}(s_t)\rangle^2\big] \le \frac{1}{\beta^2}\|R_t\|_\infty^2 \cdot \mathbb{E}_{s\sim\xi}\big[\phi_j(s)^2\big] \le \frac{\varsigma^2}{\beta}$, which is an order-one quantity. Following the same steps as in the proof as Corollary 3.1 then yields the claim.

B.6.2 Proofs for TD(λ)

We first prove Proposition 3.2—the mixing time result—and then use it to establish Corollary 3.4.

B.6.2.1 Proof of Proposition 3.2

We prove the claim via a coupling argument. Consider two initial states $\Omega_0 = (s_0, s_1, h_0)$ and $\Omega_0' = (s_0', s_1', h_1')$. By Assumption 3.1 (mixing time) for the original chain in total variation distance, there exists a coupling between a chains $(s_t)_{t\ge1}$ and $(s_t')_{t\ge1}$ starting from s_1 and s_1' respectively, such that $\mathbb{P}\big(s_{(k+1)t_{\text{mix}}+1} \neq s'_{(k+1)t_{\text{mix}}+1} \mid \{s_t, s_t'\}_{t=1}^{kt_{\text{mix}}+1}\big) \le \frac{1}{2}$. Furthermore, whenever $s_t = s_t'$ for some $t \ge 1$, the two processes are always identical from then on. Let $(g_t)_{t\geq 0}$ and $(g'_t)_{t\geq 0}$ be the eligibility trace process (3.37b) associated to $(s_t)_{t\geq 0}$ and $(s'_t)_{t\geq 0}$, respectively, and let $h_t = \frac{1-\lambda\gamma}{\varsigma\sqrt{\beta d}} g_t$ and $h'_t = \frac{1-\lambda\gamma}{\varsigma\sqrt{\beta d}} g'_t$.

Under this coupling, we note that $\mathbb{P}\big(s_{3t_{\text{mix}}+1} \neq s'_{3t_{\text{mix}}+1}\big) \leq \frac{1}{8}$. Conditioning on the event $\mathscr{E} := \big\{s_{3t_{\text{mix}}+1} = s'_{3t_{\text{mix}}+1}\big\}$, for any $t \geq 3t_{\text{mix}} + 1$, we have

$$\|h_{t+1} - h'_{t+1}\|_2 = \gamma\lambda \|h_t - h'_t\|_2 = \cdots = (\gamma\lambda)^{t - 3t_{\text{mix}} - 1} \|h_{3t_{\text{mix}}+1} - h'_{3t_{\text{mix}}+1}\|_2. \tag{B.32}$$

We split the remainder of the proof into two cases.

Case I: $s_1 \neq s'_1$**:** The coupling bound implies that $\mathbb{P}(\mathscr{E}) \geq \frac{7}{8}$. On the event $\mathscr{E}$, for $\tau \geq 3t_{\text{mix}} + 1 + \frac{4}{1-\gamma\lambda}$, we have the bound $\|h_{t+1} - h'_{t+1}\|_2 \leq \frac{1}{16}\|h_{3t_{\text{mix}}+1} - h'_{3t_{\text{mix}}+1}\|_2 \leq \frac{1}{8}$ almost surely. Under this coupling, we may write

$$\begin{aligned}
\mathbb{E}\big[\rho\big((s_\tau, s_{\tau+1}, h_\tau), (s'_\tau, s'_{\tau+1}, h_\tau)\big)\big] &= \tfrac{1}{4}\big(\mathbb{P}\big(s_\tau \neq s'_\tau\big) + \mathbb{P}\big(s_{\tau+1} \neq s'_{\tau+1}\big) + \mathbb{E}\big[\|h_\tau - h'_\tau\|_2\big]\big)\\
&\leq \tfrac{3}{4}\mathbb{P}(\mathscr{E}^c) + \tfrac{1}{4}\mathbb{E}\big[\|h_\tau - h'_\tau\|_2 \mid \mathscr{E}\big]\\
&\leq \tfrac{1}{8} = \tfrac{1}{2}\cdot\tfrac{1}{4}\mathbf{1}_{s_1 \neq s'_1} \leq \tfrac{1}{2}\rho\big((s_0, s_1, h_0), (s'_0, s'_1, h_0)\big),
\end{aligned}$$

which proves the Wasserstein contraction in this case.

Case II: $s_1 = s'_1$ In this case, the coupling construction ensures that $s_t = s'_t$ for any $t \geq 1$. Invoking the bound (B.32) then yields

$$\mathbb{E}\big[\rho\big((s_\tau, s_{\tau+1}, h_\tau), (s'_\tau, s'_{\tau+1}, h_\tau)\big)\big] = \tfrac{1}{4}\mathbb{E}\big[\|h_\tau - h'_\tau\|_2\big] \leq \tfrac{1}{8}\|h_0 - h'_0\|_2 \leq \tfrac{1}{2}\rho(\Omega_0, \Omega'_0),$$

which establishes contraction in this case. Combining the two cases proves the proposition.

B.6.2.2 Proof of Corollary 3.4

We note that the SA procedure (3.37a) can be written as

$$\theta_{t+1} = (1 - \eta\beta)\theta_t + \eta\beta \boldsymbol{L}_{t+1}(\Omega_t)\theta_t - \eta\beta \boldsymbol{b}_{t+1}(\Omega_t),$$

where $\boldsymbol{L}_{t+1}(\Omega_t) = \big(I_d - \frac{1}{\beta}g_t\phi(s_t)^\top + \gamma\frac{1}{\beta}g_t\phi(s_{t+1})^\top\big)$ and $\boldsymbol{b}_{t+1}(\Omega_t) = \frac{1}{\beta}R_t(s_t)g_t$.

Recalling that $M_\lambda = (1-\lambda)\gamma\sum_{t=0}^{\infty}\lambda^t\gamma^{t+1}B^{-1/2}\mathbb{E}\big[\phi(s_0)\phi(s_{t+1})^\top\big]B^{-1/2}$, we first study the eigenvalues of the symmetrized version of M_λ, and relate these back to those of $\bar{L} = \mathbb{E}_{\tilde{\xi}}\big[\boldsymbol{L}_{t+1}(\Omega_t)\big]$. Note that by the Cauchy–Schwarz inequality, for any vector $u \in \mathbb{S}^{d-1}$, we have

$$u^\top B^{-1/2}\mathbb{E}\big[\phi(s_0)\phi(s_t)^\top\big]B^{-1/2}u \leq \sqrt{\mathbb{E}\big[\big(u^\top B^{-1/2}\phi(s_0)\big)^2\big]\cdot\mathbb{E}\big[\big(u^\top B^{-1/2}\phi(s_t)\big)^2\big]} = 1.$$

We therefore have the bound $\frac{1}{2}\lambda_{\min}(M_\lambda + M_\lambda^\top) \leq (1-\lambda)\gamma\sum_{t=0}^{\infty}(\gamma\lambda)^t = \frac{(1-\lambda)\gamma}{1-\lambda\gamma}$. As in the proof of Corollary 3.1, we can deduce that

$$\tfrac{1}{2}\lambda_{\max}\big(\bar{L} + \bar{L}^\top\big) = \tfrac{1}{\beta}\lambda_{\max}\big(B^{1/2}\big(\tfrac{M_\lambda + M_\lambda^\top}{2}\big)B^{1/2}\big) \geq \tfrac{(1-\lambda)\gamma}{1-\lambda\gamma}.$$

Next, we verify Assumption 3.2 on the noise moments. By the update rule (3.37b), under a stationary trajectory, we have the expression $g_t = \sum_{k=0}^{\infty} (\gamma\lambda)^k \phi(s_{t-k})$. For any $u \in \mathbb{S}^{d-1}$, invoking Hölder's inequality yields

$$\mathbb{E}\big[\langle g_t,\, u\rangle^4\big] \leq \Big(\sum_{k=0}^{\infty} (\gamma\lambda)^k\Big)^3 \cdot \sum_{k=0}^{\infty} (\gamma\lambda)^k \mathbb{E}\big[\langle u,\, \phi(s_{t-k})\rangle^4\big] \leq \beta^2 \big(\tfrac{\varsigma}{1-\gamma\lambda}\big)^4.$$

In other words, for all standard basis vectors e_j, we have

$$\begin{aligned}\mathbb{E}\big[\langle e_j,\, \boldsymbol{L}_{t+1}(\Omega_t) u\rangle^2\big] &\leq 1 + \tfrac{2}{\beta^2} \sqrt{\mathbb{E}\big[\langle e_j,\, \phi(s_t)\rangle^4\big]} \cdot \sqrt{\mathbb{E}\big[\langle g_t,\, u\rangle^4\big]} \leq 1 + 2\tfrac{\varsigma^4}{(1-\gamma\lambda)^2},\\ \mathbb{E}\big[\langle e_j,\, \boldsymbol{b}_{t+1}(\Omega_t) u\rangle^2\big] &\leq \tfrac{\varsigma^2}{\beta^2} \mathbb{E}\big[\langle g_t,\, e_j\rangle^2\big] \leq \tfrac{\varsigma^4}{\beta(1-\gamma\lambda)^2}.\end{aligned}$$

It remains to verify Assumption 3.4. Note that for any pair $\Omega = (s, s_+, h)$ and $\Omega' = (s', s'_+, h')$, the operator norm $T := |\!|\!| \boldsymbol{L}_{t+1}(\Omega) - \boldsymbol{L}_{t+1}(\Omega') |\!|\!|_{\mathrm{op}}$ is almost surely upper bounded as

$$\begin{aligned}T &\leq \tfrac{\varsigma\sqrt{d/\beta}}{1-\lambda\gamma} \cdot \big(|\!|\!| h^\top \phi(s) - (h')^\top \phi(s') |\!|\!|_{\mathrm{op}} + |\!|\!| h^\top \phi(s_+) - (h')^\top \phi(s'_+) |\!|\!|_{\mathrm{op}} \big)\\ &\leq \tfrac{\varsigma\sqrt{d/\beta}}{1-\lambda\gamma} \cdot \Big\{ |\!|\!| (h-h')^\top \phi(s') |\!|\!|_{\mathrm{op}}\\ &\qquad + |\!|\!| h^\top (\phi(s') - \phi(s)) |\!|\!|_{\mathrm{op}} + |\!|\!| (h-h')^\top \phi(s'_+) |\!|\!|_{\mathrm{op}} + |\!|\!| h^\top (\phi(s'_+) - \phi(s_+)) |\!|\!|_{\mathrm{op}} \Big\}\\ &\leq \tfrac{2\varsigma^2 d}{1-\lambda\gamma} \big(\mathbf{1}_{s \neq s'} + \mathbf{1}_{s_+ \neq s'_+} + \|h - h'\|_2 \big) = \tfrac{8\varsigma^2 d}{1-\lambda\gamma} \rho(\Omega, \Omega').\end{aligned}$$

Finally, we note that the quantity $\bar{\sigma}$ defined in equation (3.29) satisfies the bound

$$\begin{aligned}&\sup_{j\in[d]} \mathbb{E}\big[\langle e_j,\, (\boldsymbol{L}_{t+1}(\Omega_t) - \bar{L})\bar{\theta} + (\boldsymbol{b}_{t+1}(\Omega_t) - b)\rangle^2\big]\\ &\leq \sup_{j\in[d]} \sqrt{\mathbb{E}\big[\langle e_j,\, g_t\rangle^4\big]} \cdot \sqrt{\Big(\mathbb{E}\big[\phi(s_t)^\top \bar{\theta} - \gamma\phi(s_{t+1})^\top \bar{\theta} - R_t(s_t)\big)^4\Big]} \leq \tfrac{\bar{\sigma}^2}{(1-\gamma\lambda)^2}.\end{aligned}$$

Invoking equation (3.73), with the test matrix $Q := B$ and substituting the expression $v(s) = \langle \theta,\, \phi(s)\rangle$ yields the claim.

B.6.3 Proofs for vector autoregressive estimation

In this section, we present proofs of results on vector autoregressive models, as introduced in Example 3.3.

B.6.3.1 Proof of Proposition 3.3

We prove the claim by a direct construction of the coupling. Given two initial points $\Omega_0 = \big[X_1^\top, X_0^\top, \cdots, X_{-k+1}^\top\big]^\top$ and $\Omega'_0 = \big[X_1'^\top, X_0'^\top, \cdots, X_{-k+1}'^\top\big]^\top$, we consider a pair of stochastic processes $(X_t)_{t\geq 1}$ and $(X'_t)_{t\geq 1}$ starting from Ω_0 and Ω', respectively, driven by the same noise process $(\varepsilon_t)_{t\geq 0}$. Introduce the shorthand $Y_{t+1} = \begin{bmatrix} X_{t+1} & \dots & X_{t-k+2}\end{bmatrix}^\top$ (note that Y_{t+1} is a sliding window with length one unit shorter than Ω_t). We have:

$$\begin{aligned} \|Y_{t+1} - Y'_{t+1}\|_{P_*}^2 \; &= \; \|R_* \left(Y_t - Y'_t\right)\|_{P_*}^2 = \|Y_t - Y'_t\|_{P_*}^2 - \|Y_t - Y'_t\|_{Q_*}^2 \\ &\leq \big(1 - \tfrac{\mu}{\beta}\big)\|Y_t - Y'_t\|_{P_*}^2 . \end{aligned}$$

Consequently, the augmented processes $\Omega_t = (X_{t+1}, X_t, \cdots, X_{t-k+1})$ and $\Omega'_t = (X'_{t+1}, X'_t, \cdots, X'_{t-k+1})$ satisfy the bound

$$\begin{aligned} \|\Omega_t - \Omega'_t\|_2 \leq \|Y_{t+1} - Y'_{t+1}\|_2 + \|Y_t - Y'_t\|_2 &\leq \tfrac{1}{\sqrt{\lambda_{\min}(P_*)}}\big(\|Y_{t+1} - Y'_{t+1}\|_{P_*} + \|Y_t - Y'_t\|_{P_*}\big) \\ &\leq 2\sqrt{\tfrac{\lambda_{\max}(P_*)}{\lambda_{\min}(P_*)}}\big(1 - \tfrac{\mu}{2\beta}\big)^t \|\Omega_0 - \Omega'_0\|_2 \end{aligned}$$

Note that since $P_* \succeq Q_*$, we have $\lambda_{\min}(P_*) \geq \lambda_{\min}(Q_*) = \mu$. Taking $t_{\mathrm{mix}} = c\frac{\beta}{\mu}\big(1 + \log\frac{\beta}{\mu}\big)$ yields the contraction bound $\|\Omega_{t_{\mathrm{mix}}} - \Omega'_{t_{\mathrm{mix}}}\|_2 \leq \frac{1}{2}\|\Omega_0 - \Omega'_0\|_2$. Taking expectations on both sides completes the proof.

B.6.3.2 Proof of Corollary 3.5

We begin by showing norm bounds and moment bounds on the process $(X_t)_{t\geq 0}$. By definition (3.17) of the process and stability, the block vector $Y_t := \begin{bmatrix} X_t & X_{t-1} & \cdots & X_{t-k+1}\end{bmatrix}^\top$ satisfies the recursion $Y_t = \sum_{i=0}^\infty R_*^i \varepsilon_{t-i} e_1$, where e_1 is the standard block basis vector equal to identify on the first block. We therefore have the bound

$$\|X_t\|_2 \leq \tfrac{1}{\mu}\|Y_t\|_{P_*} \leq \sum_{i=0}^\infty \|R_*^i \varepsilon_{t-i} e_1\|_{P_*} \leq \tfrac{1}{\mu}\sum_{i=0}^\infty \big(1 - \tfrac{\mu}{\beta}\big)^i \|\varepsilon_{t-i} e_1\|_{P_*} \leq \tfrac{\beta^2}{\mu^2}\varsigma\sqrt{m}.$$

Moreover, for each $u \in \mathbb{S}^{m-1}$, we have

$$\begin{aligned} \mathbb{E}\big[\langle X_t,\, u\rangle^4\big] &\leq \big(\sum_{i=0}^\infty e^{-\frac{i\mu}{6\beta}}\big)^3 \cdot \sum_{i=0}^\infty e^{\frac{i\mu}{2\beta}}\, \mathbb{E}\big[\langle R_*^i \varepsilon_{t-i} e_1,\, u e_1\rangle^4\big] \\ &\leq c\big(\beta/\mu\big)^3 \cdot \sum_{i=0}^\infty e^{\frac{i\mu}{2\beta}} \cdot \tfrac{\beta^4}{\mu^4} \cdot e^{-\frac{i\mu}{\beta}}\varsigma^4 \;\leq\; c'\big(\tfrac{\beta^2\varsigma}{\mu^2}\big)^4 . \end{aligned}$$

Next, we proceed with verifying the assumptions used in Theorem 3.1. Letting $\nu := 1/|\!|\!| H^* |\!|\!|_{\mathrm{op}}$, the stochastic approximation procedure can be rewritten as

$$\begin{aligned} \theta_{t+1} &= (1 - \tfrac{\eta}{\nu})\theta_t \\ &\quad + \tfrac{\eta}{\nu}\big(\theta_t - \nu\big(\big[X_{t-j}X_{t+1-i}^\top\big]_{i,j\in[m]} \otimes I_m\big)\theta_t \nu \cdot \mathrm{vec}\big(\begin{bmatrix} X_{t+1}X_t^\top & \cdots & X_{t+1}X_{t-k+1}^\top\end{bmatrix}\big)\big). \end{aligned}$$

Observe that the matrix $\bar{L} := I_{km^2} - \nu H^* \otimes I_m$ satisfies the eigenvalue bound

$$\tfrac{1}{2}\lambda_{\max}(\bar{L} + \bar{L}^\top) \leq 1 - \tfrac{\nu}{2}\lambda_{\min}(H^* + (H^*)^\top) \leq 1 - \nu h^* .$$

On the other hand, the empirical observations satisfy the almost-sure bounds

$$\begin{aligned} \|\!|\boldsymbol{L}_{t+1}(\Omega_t) - \bar{L}|\!\|_{\mathrm{op}} &\leq \nu \cdot \|\!| \big[X_{t-j} X_{t+1-i}^\top\big]_{i,j \in [m]} |\!\|_{\mathrm{op}} \leq \nu \cdot \tfrac{\beta^4}{\mu^4} \varsigma^2 m k \text{ and} \\ \|\!|\boldsymbol{b}_{t+1}(\Omega_t) - \bar{b}|\!\|_{\mathrm{op}} &\leq \nu \cdot \|\!| \begin{bmatrix} X_{t+1} X_t^\top & \cdots & X_{t+1} X_{t-k+1}^\top \end{bmatrix} |\!\|_F \leq \nu \cdot \tfrac{\beta^4}{\mu^4} \varsigma^2 m \sqrt{k}. \end{aligned}$$

For two collections of matrices $\mathcal{U} = \big(U^{(j)}\big)_{j=1}^k$ and $\mathcal{V} = \big(V^{(j)}\big)_{j=1}^k \subseteq \mathbb{R}^{m \times m}$ such that $\sum_{j=1}^k |\!\|U^{(j)}|\!\|_F^2 = \sum_{j=1}^k |\!\|V^{(j)}|\!\|_F = 1$, the corresponding moment can be bounded as

$$\mathbb{E}\big[\langle \mathrm{vec}(\mathcal{U}),\, \big(\boldsymbol{L}_{t+1}(\Omega_t) - \bar{L}\big)\mathrm{vec}(\mathcal{V})\rangle^2\big] \leq \nu^2 \mathbb{E}\Big[\Big(\sum_{\ell=0}^{k-1} \langle U^{(\ell)},\, \sum_{j=0}^{k-1} V^{(j)} X_{t-j} X_{t-\ell}^\top \rangle_F\Big)^2\Big],$$

which is in turn at most

$$\nu^2 k^2 \sum_{\ell=0}^{k-1} \sum_{j=0}^{k-1} \sqrt{\mathbb{E}\big[X_{t-\ell}^{\otimes 4}\big]\big[(U^{(\ell)})^\top U^{(\ell)}, (U^{(\ell)})^\top U^{(\ell)}\big]} \cdot \sqrt{\mathbb{E}\big[X_{t-j}^{\otimes 4}\big]\big[(V^{(j)})^\top V^{(j)}, (V^{(j)})^\top V^{(j)}\big]}.$$

In order to bound this last quantity, we let $(U^{(\ell)})^\top U^{(\ell)} = \sum_{i=1}^m \lambda_i^2 u_i u_i^\top$ be its singular value decomposition, and note that

$$\begin{aligned} \mathbb{E}\big[X_{t-\ell}^{\otimes 4}\big]\big[(U^{(\ell)})^\top U^{(\ell)}, (U^{(\ell)})^\top U^{(\ell)}\big] &= \mathbb{E}\big[X_{t-\ell}^{\otimes 4}\big]\Big[\sum_{i=1}^m \lambda_i^2 u_i u_i^\top, \sum_{i=1}^m \lambda_i^2 u_i u_i^\top\Big] \\ &= \sum_{i,i'} \mathbb{E}\big[X_{t-\ell}^{\otimes 4}\big][u_i, u_i, u_{i'}, u_{i'}] \cdot \lambda_i^2 \lambda_{i'}^2 \leq c' \big(\tfrac{\beta^2 \varsigma}{\mu^2}\big)^4 \Big(\sum_i \lambda_i^2\Big)^2 = c' \big(\tfrac{\beta^2 \varsigma}{\mu^2}\big)^4 |\!\|U^{(\ell)}|\!\|_F^2. \end{aligned}$$

Putting together the pieces, we have

$$\begin{aligned} &\mathbb{E}\big[\langle \mathrm{vec}(\mathcal{U}),\, \big(\boldsymbol{L}_{t+1}(\Omega_t) - \bar{L}\big)\mathrm{vec}(\mathcal{V})\rangle^2\big] \\ &\qquad \leq \nu^2 k^2 c' \big(\tfrac{\beta^2 \varsigma}{\mu^2}\big)^4 \cdot \sum_{\ell=0}^{k-1} \sum_{j=0}^{k-1} |\!\|U^{(\ell)}|\!\|_F^2 |\!\|V^{(j)}|\!\|_F^2 \leq c \big(\nu \cdot \tfrac{\beta^4 k \varsigma^2}{\mu^4}\big)^2. \end{aligned}$$

Similarly, we can prove analogous moment bounds on $\boldsymbol{b}_{t+1}(\Omega_t)$. In particular, for indices $\ell \in [k]$ and $i, j \in [m]$, we consider the coordinate direction of the (i,j) entry in the ℓ-th matrix to deduce that

$$\begin{aligned} \mathbb{E}\big[\langle e_{\ell,i,j},\, (\boldsymbol{b}_{t+1}(\Omega_t) - \bar{b})\rangle^2\big] &\leq \nu^2 \mathbb{E}\big[\langle e_i e_j^\top,\, X_{t+1} X_{t-\ell+1}\rangle^2\big] \\ &\leq \nu^2 \sqrt{\mathbb{E}\big[\langle e_j^\top,\, X_{t+1}\rangle^4\big]} \cdot \sqrt{\mathbb{E}\big[\langle e_i^\top,\, X_{t-\ell+1}\rangle^4\big]} \leq c' \big(\nu \cdot \tfrac{\beta^2 \varsigma}{\mu^2}\big)^4. \end{aligned}$$

Applying Theorem 3.1 completes the proof of this corollary.

Appendix C

Proofs and discussion deferred from Chapter 4

C.1 Proofs of auxiliary results in the upper bounds

In this section, we state and prove some auxiliary results used in the proofs of our non-asymptotic upper bounds.

C.1.1 Some properties of the estimator $\widehat{\tau}_n^f$

In this appendix, we collect some properties of the estimator $\widehat{\tau}_n^f$ defined in equation (4.5).

Proposition C.1. *Given any deterministic function $f \in \mathbb{L}^2(\xi \times \pi)$ for any $a \in \mathbb{A}$, we have $\mathbb{E}[\widehat{\tau}_n^f] = \tau_g(\mathcal{I})$. Furthermore, if $\langle f(x, \cdot),\, \pi(x, \cdot) \rangle_\lambda = 0$ for any $x \in \mathbb{X}$, we have*

$$n \cdot \mathbb{E}\Big[\, |\widehat{\tau}_n^f - \tau_g(\mathcal{I})|^2 \,\Big] = \mathrm{var}_\xi \Big(\langle g(X, \cdot),\, \mu^*(X, \cdot) \rangle_\lambda \Big) + \int_{\mathbb{A}} \mathbb{E}\Big[\frac{\sigma^2(X,a) g^2(X,a)}{\pi(X,a)} \Big] d\lambda(a)$$
$$+ \int_{\mathbb{A}} \mathbb{E}\Big[\pi(X,a) \Big| f(X,a) - \tfrac{g(X,a)\mu^*(X,a)}{\pi(X,a)} + \langle g(X, \cdot),\, \mu^*(X, \cdot) \rangle_\lambda \Big|^2 \Big] d\lambda(a). \tag{C.1}$$

This decomposition immediately implies the claims given in the text. The only portion of the MSE decomposition (C.1) that depends on f is the third term, and by inspection, this third term is equal to zero if and only if

$$f(x,a) = \tfrac{g(x,a)\mu^*(x,a)}{\pi(x,a)} - \langle g(x, \cdot),\, \mu^*(x, \cdot) \rangle_\lambda \qquad \text{for all } (x,a) \in \mathbb{X} \times \mathbb{A}.$$

Proof. Since the action A_i follows the probability distribution $\pi(X_i, \cdot)$ conditionally on X_i, we have $\mathbb{E}\big[f(X_i, A_i) \mid X_i \big] = \langle \pi(X_i, \cdot),\, f(X_i, \cdot) \rangle_\lambda$, and the estimator $\widehat{\tau}_n^f$ is always unbiased. Since the function f is square-integrable with respect to the measure $\xi \times \pi$, the second moment can be decomposed as follows:

$$
\begin{aligned}
&\mathbb{E}\Big[\Big| \frac{g(X_i, A_i)}{\pi(X_i, A_i)} Y_i - f(X_i, A_i) + \int_{\mathbb{A}} \pi(X_i, a) f(X_i, a) d\lambda(a) \Big|^2 \Big] \\
&= \int_{\mathbb{A}} \mathbb{E}\Big[\pi(X_i, a) \cdot \Big| \frac{g(X_i, a)}{\pi(X_i, a)} Y_i - f(X_i, a) \Big|^2 \Big] d\lambda(a) \\
&= \int_{\mathbb{A}} \mathbb{E}\Big[\frac{\sigma^2(X, a) g^2(X, a)}{\pi(X, a)} \Big] d\lambda(a) + \int_{\mathbb{A}} \mathbb{E}\Big[\pi(X, a) \cdot \Big| \frac{g(X, a)\mu^*(X, a)}{\pi(X, a)} - f(X, a) \Big|^2 \Big] d\lambda(a)
\end{aligned}
$$

Conditionally on the value of X, we have the bias-variance decomposition

$$
\begin{aligned}
&\int_{\mathbb{A}} \pi(X, a) \cdot \Big| \frac{g(X, a)\mu^*(X, a)}{\pi(X, a)} - f(X, a) \Big|^2 d\lambda(a) \\
&= \langle g(X, \cdot),\, \mu^*(X, \cdot) \rangle_\lambda^2 \\
&\qquad + \int_{\mathbb{A}} \pi(X, a) \cdot \Big| f(X, a) - \frac{g(X, a)\mu^*(X, a)}{\pi(X, a)} + \langle g(X, \cdot),\, \mu^*(X, \cdot) \rangle_\lambda \Big|^2 d\lambda(a).
\end{aligned}
$$

Finally, we note that

$$
\begin{aligned}
&\mathbb{E}\Big[\langle g(X, \cdot),\, \mu^*(X, \cdot) \rangle_\lambda^2 \Big] - \tau^2(\mathcal{I}) \\
&\qquad\qquad = \Big(\mathbb{E}\Big[\langle g(X, \cdot),\, \mu^*(X, \cdot) \rangle_\lambda \Big] \Big)^2 = \mathrm{var}_\xi \Big(\langle g(X, \cdot),\, \mu^*(X, \cdot) \rangle_\lambda \Big).
\end{aligned}
$$

Putting together the pieces completes the proof. □

C.1.2 Existence of critical radii

In this section, we establish the existence of critical radii $s_m(\mu)$ and $r_m(\mu)$ defined in equations (4.14a) and (4.14b), respectively.

Proposition C.2. *Suppose that the compatibility condition* ***(CC)*** *holds, and that the Rademacher complexities $\mathcal{S}_m\big((\mathcal{F} - \mu) \cap \mathbb{B}_\omega(r_0)\big)$ and $\mathcal{R}_m\big((\mathcal{F} - \mu) \cap \mathbb{B}_\omega(r_0)\big)$ are finite for some $r_0 > 0$. Then:*

(a) There exists a unique scalar $s_m = s_m(\mu) > 0$ such that inequality (4.14a) *holds for any $s \geq s_m$, with equality when $s = s_m$, and is false when $s \in [0, s_m)$.*

(b) There exists a scalar $r_m = r_m(\mu) > 0$ such that inequality (4.14b) *holds for any $r \geq r_m$.*

Proof. Denote the shifted function class $\mathcal{F}^* := \mathcal{F} - \mu$. Since the class $\mathcal{F}$ is convex by assumption, for positive scalars $r_1 < r_2$ and any function $f \in \mathcal{F}^* \cap \mathbb{B}_\omega(r_2)$, we have $\frac{r_1}{r_2} f \in \mathcal{F}^* \cap \mathbb{B}_\omega(r_1)$.

$$
\frac{1}{r_2} \mathcal{R}_m(\mathcal{F}^* \cap \mathbb{B}_\omega(r_2)) \leq \frac{1}{r_2} \mathcal{R}_m\Big(\frac{r_2}{r_1} \cdot \big(\mathcal{F}^* \cap \mathbb{B}_\omega(r_1) \big) \Big) = \frac{1}{r_1} \mathcal{R}_m(\mathcal{F}^* \cap \mathbb{B}_\omega(r_1)).
$$

So the function $r \mapsto r^{-1}\mathcal{R}_m(\mathcal{F}^* \cap \mathbb{B}_\omega(r))$ is non-increasing in r. A similar argument ensures that the function $s \mapsto s^{-1}\mathcal{S}(\mathcal{F}^* \cap \mathbb{B}_\omega(s))$ is also non-increasing in s.

Since the function class $\mathcal{F}$ is compact in $\mathbb{L}^2_\omega$, we have $D := \mathrm{diam}_\omega(\mathcal{F} \cup \{\mu^*\}) < +\infty$, and hence

$$\mathcal{R}_m(\mathcal{F}^*) = \mathcal{R}_m\big(\mathcal{F}^* \cap \mathbb{B}_\omega(D)\big) \leq \frac{D}{r_0}\mathcal{R}_m\big(\mathcal{F}^* \cap \mathbb{B}_\omega(r_0)\big) < +\infty,$$

which implies that $\mathcal{R}_m\big(\mathcal{F}^* \cap \mathbb{B}_\omega(r)\big) < +\infty$ for any $r > 0$. Similarly, the Rademacher complexity $\mathcal{S}\big(\mathcal{F}^* \cap \mathbb{B}_\omega(s)\big)$ is also finite.

For the inequality (4.14a), the left hand side is a non-increasing function of s, while the right hand side is strictly increasing and diverging to infinity as $s \to +\infty$. Furthermore, the right-hand-side is equal to zero at $s = 0$, while the left-hand side is always finite and non-negative for $s > 0$. Consequently, a unique fixed point $s_m \geq 0$ exists, and we have

$$s^{-1}\mathcal{S}\big(\mathcal{F}^* \cap \mathbb{B}_\omega(s)\big) \begin{cases} < s, & \text{for } s > s_m, \text{ and} \\ > s, & \text{for } s \in (0, s_m). \end{cases}$$

As for inequality (4.14b), the left-hand-side is non-increasing, and we have

$$\lim_{r \to +\infty} r^{-1}\mathcal{R}(\mathcal{F}^* \cap \mathbb{B}_\omega(r)) \leq \lim_{r \to +\infty} r^{-1}\mathcal{R}(\mathcal{F}^*) = 0.$$

So there exists $r_m \geq 0$ such that inequality (4.14b) holds for any $r \geq r_m$.

□

C.1.3 Proof of Lemma 4.3

We define the auxiliary function

$$\phi(t) := \begin{cases} 0 & t \leq 1, \\ t-1 & 1 \leq t \leq 2, \\ 1 & t > 2. \end{cases}$$

First, observe that for any scalar $u > 0$, we have

$$\begin{aligned} \frac{1}{m}\sum_{i=1}^m \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} h^2(X_i, A_i) &\geq \frac{1}{m}\sum_{i=1}^m u^2 \cdot \mathbb{I}\Big[\frac{|g(X_i, A_i)h(X_i, A_i)|}{\pi(X_i, A_i)} \geq u\Big] \\ &\geq \frac{1}{m}\sum_{i=1}^m u^2 \cdot \phi\Big(\frac{|g(X_i, A_i)h(X_i, A_i)|}{\pi(X_i, A_i)u}\Big) =: Z_m^{(\phi)}(h). \end{aligned}$$

Second, for any function $h \in \mathcal{H}$, we have

$$\begin{aligned}
\mathbb{E}\Big[Z_m^{(\phi)}(h)\Big] &= u^2 \cdot \sum_{a \in \mathbb{A}} \mathbb{E}_\xi \Big[\pi(X,a)\phi\Big(\frac{|g(X,a)h(X,a)|}{\pi(X,a)u}\Big)\Big] \\
&\geq u^2 \sum_{a \in \mathbb{A}} \mathbb{E}_\xi \Big[\pi(X,a) \cdot \mathbb{I}\big[\frac{|g(X,a)h(X,a)|}{\pi(X,a)u} \geq 2\big]\Big] \\
&= u^2 \cdot \mathbb{P}_{X \sim \xi, A \sim \pi(X,\cdot)}\Big(\frac{|g(X,A)h(X,A)|}{\pi(X,A)} \geq 2u\Big).
\end{aligned}$$

Recall that the constant α_1 is the constant factor in the small-ball probability condition **(SB)**. Choosing the threshold $u := \frac{c_1}{2}$ and using the equality $\|h\|_\omega = 1$, we see that the small-ball condition implies that

$$\mathbb{P}_{X \sim \xi, A \sim \pi(X,\cdot)}\Big(\frac{|g(X,A)h(X,A)|}{\pi(X,A)} \geq 2u\Big) \geq \alpha_2.$$

Now we turn to study the deviation bound for $Z_m^{(\phi)}(h)$. Using known concentration inequalities for empirical processes [2]—see Proposition C.4 in Appendix C.4 for more detail—we are guaranteed to have

$$\begin{aligned}
&\sup_{h \in \mathcal{H}} \Big(Z_m^{(\phi)}(h) - \mathbb{E}\big[Z_m^{(\phi)}(h)\big]\Big) \\
&\qquad\qquad \leq 2\mathbb{E} \sup_{h \in \mathcal{H}} \Big(Z_m^{(\phi)}(h) - \mathbb{E}\big[Z_m^{(\phi)}(h)\big]\Big) + c\alpha_1^2 \cdot \Big\{\sqrt{\frac{\log(1/\varepsilon)}{m}} + \frac{\log(1/\varepsilon)}{m}\Big\}.
\end{aligned}$$

with probability at least $1 - \varepsilon$.

For the expected supremum term, standard symmetrization arguments lead to the bound

$$\mathbb{E} \sup_{h \in \mathcal{H}} \Big(Z_m^{(\phi)}(h) - \mathbb{E}\big[Z_m^{(\phi)}(h)\big]\Big) \leq \frac{\alpha_1^2}{m} \cdot \mathbb{E}\Big[\sup_{h \in \mathcal{H}} \sum_{i=1}^m \varepsilon_i \phi\Big(\frac{2|h(X_i,A_i)g(X_i,A_i)|}{\alpha_1 \pi(X_i,A_i)}\Big)\Big].$$

Note that since ϕ is a 1-Lipschitz function, we may apply the Ledoux-Talagrand contraction (e.g., equation (5.6.1) in the book [213]) so as to obtain

$$\mathbb{E}\Big[\sup_{h \in \mathcal{H}} \sum_{i=1}^m \varepsilon_i \phi\Big(\frac{2|h(X_i,A_i)g(X_i,A_i)|}{\alpha_1 \pi(X_i,A_i)}\Big)\Big] \leq \frac{4}{\alpha_1} \mathbb{E}\Big[\sup_{h \in \mathcal{H}} \sum_{i=1}^m \frac{\varepsilon_i g(X_i,A_i)}{\pi(X_i,A_i)} h(X_i,A_i)\Big].$$

Combining the pieces yields the lower bound

$$\begin{aligned}
&\frac{1}{m} \sum_{i=1}^m \tfrac{g^2(X_i,A_i)}{\pi^2(X_i,A_i)} h^2(X_i,A_i) \\
&\quad \geq \tfrac{\alpha_2 \alpha_1^2}{4} - \frac{4\alpha_1}{m} \mathbb{E}\Big[\sup_{h \in \mathcal{H}} \sum_{i=1}^m \frac{\varepsilon_i g(X_i,A_i)}{\pi(X_i,A_i)} h(X_i,A_i)\Big] - c\alpha_1^2 \cdot \Big\{\sqrt{\tfrac{\log(1/\varepsilon)}{m}} + \tfrac{\log(1/\varepsilon)}{m}\Big\}, \quad \text{(C.2)}
\end{aligned}$$

uniformly holding true over $h \in \mathcal{H}$, with probability $1 - \varepsilon$, which completes the proof of the lemma.

C.2 Proofs of the corollaries

This section is devoted to the proofs of Corollaries 4.1—4.4, as stated in Section 4.2.4.

C.2.1 Proof of Corollary 4.1

Let us introduce the shorthand $f_\theta(x, a) := \langle \theta,\, \phi(x, a)\rangle$ for functions that are linear in the feature map. Moreover, for a vector $\bar{\theta} \in \mathbb{R}^d$ and radius $r > 0$, we define the recentering function $\bar{\mu}(x, a) := \langle \bar{\theta},\, \phi(x, a)\rangle$.

Our proof strategy is to bound the pair of critical radii (s_m, r_m), and we do so by controlling the associated Rademacher complexities. By a direct calculation, we find that

$$(\mathcal{F} - \bar{\mu}) \cap \mathbb{B}_\omega(r) \subseteq \Big\{ f_\theta \;\mid\; \theta^\top \Sigma \theta \leq r^2 \Big\}, \quad \text{where } \Sigma := \mathbb{E}\Big[\frac{g^2(X, A)}{\pi^2(X, A)} \phi(X, A)\phi(X, A)^\top\Big].$$

We can therefore bound the Rademacher complexities as

$$\begin{aligned} &\mathcal{S}_m^2\Big((\mathcal{F} - \bar{\mu}) \cap \mathbb{B}_\omega(r)\Big) \\ &\leq \mathbb{E}\Big[\sup_{\|\theta\|_\Sigma \leq r} \Big\{\frac{1}{m}\langle \theta,\, \sum_{i=1}^m \frac{\varepsilon_i g^2(X_i, A_i)}{\pi^2(X_i, A_i)} (Y_i - \mu^*(X_i, A_i))\phi(X_i, A_i)\rangle\Big\}^2\Big] \\ &= \frac{r^2}{m} \operatorname{trace}\Big(\Sigma^{-1}\Gamma_\sigma\Big), \end{aligned}$$

and

$$\mathcal{R}_m\Big((\mathcal{F} - \bar{\mu}) \cap \mathbb{B}_\omega(r)\Big) \leq \mathbb{E}\Big[\sup_{\|\theta\|_\Sigma \leq r} \frac{1}{m}\langle \theta,\, \sum_{i=1}^m \frac{\varepsilon_i g(X_i, A_i)}{\pi(X_i, A_i)} \phi(X_i, A_i)\rangle\Big] \leq r\sqrt{\frac{d}{m}}.$$

By definition of the fixed point equations, the critical radii can be upper bounded as

$$s_m \leq \sqrt{m^{-1} \operatorname{trace}\Big(\Sigma^{-1}\Gamma_\sigma\Big)}, \quad \text{and} \quad r_m \leq \begin{cases} +\infty, & m \leq \frac{1024}{\alpha_1^2 \alpha_2^2} d \\ 0, & m > \frac{1024}{\alpha_1^2 \alpha_2^2} d \end{cases}.$$

Combining with Theorem 4.2 completes the proof of this corollary.

C.2.2 Proof of Corollary 4.2

We introduce the shorthand $f_\theta(x, a) = \langle \theta,\, \phi(x, a)\rangle$ for functions that are linear in the feature map. Given any vector $\bar{\theta} \in \mathbb{R}^d$ such that $\|\bar{\theta}\|_1 = R_1$, define the set $S = \operatorname{supp}(\bar{\theta}) \subseteq [d]$ along with the function $\bar{\mu}(x, a) = \langle \bar{\theta},\, \phi(x, a)\rangle$. For any radius $r > 0$ and vector $\theta \in (\mathcal{F} - \bar{\mu}) \cap \mathbb{B}_\omega(r)$, we note that

$$\|\theta_{S^c}\|_1 = \|\theta_{S^c} + \bar{\theta}_{S^c}\|_1 = \|\theta + \bar{\theta}\|_1 - \|\theta_S + \bar{\theta}_S\|_1 \leq R_1 - \|\bar{\theta}_S\|_1 + \|\theta_S\|_1 \leq \|\theta_S\|_1.$$

Recalling that $\Sigma = \mathbb{E}\big[\frac{g^2(X,A)}{\pi^2(X,A)}\phi(X,A)\phi(X,A)^\top\big]$, we have the inclusions

$$\begin{aligned}(\mathcal{F} - \bar{\mu}) \cap \mathbb{B}_\omega(r) &\subseteq r \cdot \Big\{ f_\theta \mid \|\theta_{S^c}\|_1 \leq \|\theta_S\|_1, \ \|\theta\|_\Sigma \leq 1 \Big\} \\ &\subseteq r \cdot \Big\{ f_\theta \ \mid \ \|\theta\|_1 \leq 2\sqrt{|S|/\lambda_{\min}(\Sigma)} \Big\} \\ &\subseteq 2r\sqrt{|S|/\lambda_{\min}(\Sigma)} \cdot \mathrm{conv}\Big(\big\{ \pm \phi_j \big\}_{j=1}^d \Big). \qquad \text{(C.3)}\end{aligned}$$

where the second step follows from the bound $\|\theta_S\|_1 \leq \|\theta_S\|_2\sqrt{|S|} \leq \|\theta\|_\Sigma\sqrt{|S|/\lambda_{\min}(\Sigma)}$, valid for any $\theta \in \mathbb{R}^d$.

For each coordinate $j = 1, \ldots, d$, we can apply the Hoeffding inequality along with the sub-Gaussian tail assumption (4.22) so as to obtain

$$\mathbb{P}\Big[\Big|\frac{1}{m}\sum_{i=1}^m \frac{\varepsilon_i g(X_i, A_i)}{\pi(X_i, A_i)}\phi(X_i, A_i)^\top e_j\Big| \geq t\Big] \leq 2e^{-\frac{2mt^2}{\sigma^2}} \quad \text{for any } t > 0/$$

Taking the union bound over $j = 1, 2, \ldots, d$ and then integrating the resulting tail bound, we find that

$$\mathbb{E}\Big[\max_{j=1,\ldots,d}\Big|\frac{1}{m}\sum_{i=1}^m \frac{g(X_i, A_i)}{\pi(X_i, A_i)}\varepsilon_i\phi_j(X_i, A_i)\Big|\Big] \leq \sigma\sqrt{\frac{\log d}{m}}.$$

Combining with equation (C.3), we conclude that

$$\mathcal{R}_m\big((\mathcal{F} - \bar{\mu}) \cap \mathbb{B}_\omega(r)\big) \leq 2r\sigma\sqrt{\frac{|S| \cdot \log(d)}{m\lambda_{\min}(\Sigma)}},$$

for any $\bar{\mu}(x, a) = \langle\bar{\theta}, \ \phi(x, a)\rangle$ with $\bar{\theta}$ supported on S.

Consequently, defining the constant $c_0 = \frac{4096}{\alpha_1^2\alpha_2^2}$, when the sample size satisfies $m \geq c_0|S|\frac{\sigma^2\log(d)}{\lambda_{\min}(\Sigma)}$, the critical radius r_m is 0.

Now we turn to bound the critical radius s_m. By the sub-Gaussian condition (4.22), we have the Orlicz norm bound

$$\begin{aligned}&\Big\|\frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)}\varepsilon_i\phi_j(X_i, A_i)(Y_i - \mu^*(X_i, A_i))\Big\|_{\psi_1} \\ &\qquad\qquad \leq \Big\|\frac{g(X_i, A_i)}{\pi(X_i, A_i)}\phi_j(X_i, A_i)\Big\|_{\psi_1} \cdot \|\frac{g(X_i, A_i)}{\pi(X_i, A_i)}(Y_i - \mu^*(X_i, A_i))\|_{\psi_1} \leq \sigma\bar{\sigma}.\end{aligned}$$

Invoking a known concentration inequality (see Proposition C.4 in Appendix C.4), we conclude that there exists a universal constant $c_1 > 0$ such that

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^m \frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)}\varepsilon_i\phi_j(X_i, A_i)(Y_i - \mu^*(X_i, A_i))\right| \geq t\right) \\ &\qquad\qquad\qquad\qquad \leq 2\exp\left(\frac{-c_1mt^2}{\sigma^2\bar{\sigma}^2 + t\sigma\bar{\sigma}\log(m)}\right),\end{aligned}$$

for any scalar $t > 0$.

Taking the union bound over $j = 1, 2, \ldots, d$ and integrating out the tail yields

$$\mathbb{E}\left[\max_{j\in[d]}\left|\frac{1}{m}\sum_{i=1}^{m}\frac{g(X_i,A_i)}{\pi(X_i,A_i)}\varepsilon_i\phi_j(X_i,A_i)\right|^2\right] \leq c_2\sigma^2\bar{\sigma}^2\Big\{\sqrt{\frac{\log d}{m}}+\frac{\log d\cdot\log m}{m}\Big\}^2,$$

Given a sample size lower bounded as $m \geq \log^2 d$, the derivation above guarantees that the Rademacher complexity is upper bounded as

$$\mathcal{S}\big((\mathcal{F}-\bar{\mu})\cap\mathbb{B}_\omega(r)\big) \leq cr\sigma\bar{\sigma}\sqrt{\frac{|S|\cdot\log(d)}{m\lambda_{\text{min}}(\Sigma)}},$$

and consequently, the associated critical radius satisfies an upper bound of the form $s_m \leq c\sigma\bar{\sigma}\sqrt{\frac{|S|\log(d)}{m\lambda_{\text{min}}(\Sigma)}}$. Combining with Theorem 4.2 completes the proof of Corollary 4.2.

C.2.3 Proof of Corollary 4.3

Clearly, the function class $\mathcal{F}_k$ is symmetric and convex. Consequently, for any $\bar{\mu} \in \mathcal{F}_k$, we have

$$(\mathcal{F}_k - \bar{\mu}) \cap \mathbb{B}_\omega(r) \subseteq (2\mathcal{F}_k) \cap \mathbb{B}_\omega(r).$$

For any pair $\mu_1, \mu_2 \in (2\mathcal{F}) \cap \mathbb{B}_\omega(r)$, by the sub-Gaussian assumption in equations (4.23), we have that

$$\mathbb{E}\Big[\Big(\frac{g(X_i,A_i)}{\pi(X_i,A_i)}\varepsilon_i(\mu_1-\mu_2)(X_i,A_i)\Big)^2\Big] = \|\mu_1-\mu_2\|_\omega^2, \quad \text{and}$$
$$\|\frac{g(X_i,A_i)}{\pi(X_i,A_i)}\varepsilon_i(\mu_1-\mu_2)(X_i,A_i)\|_{\psi_1} \leq \sigma\|\mu_1-\mu_2\|_\infty.$$

By a known concentration inequality (see Proposition C.4 in Appendix C.4), for any $t > 0$, we have

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m}\frac{g(X_i,A_i)}{\pi(X_i,A_i)}\varepsilon_i(\mu_1-\mu_2)(X_i,A_i)\right|\geq t\right)$$
$$\leq 2\exp\left(\frac{-c_1mt^2}{\|\mu_1-\mu_2\|_\omega^2+t\sigma\|\mu_1-\mu_2\|_\infty\log(m)}\right),$$

We also note that the Cauchy–Schwarz inequality implies that

$$\mathbb{E}\Big[\sup_{\|\mu_1-\mu_2\|_\omega\leq\delta}\frac{1}{m}\sum_{i=1}^{m}\frac{g(X_i,A_i)}{\pi(X_i,A_i)}\varepsilon_i(\mu_1-\mu_2)(X_i,A_i)\Big] \leq \delta.$$

By a known mixed-tail chaining bound (see Proposition C.5 and equation (C.9) in Appendix C.4), we find that

$$\mathcal{R}_m\big((\mathcal{F}_k - \bar{\mu}) \cap \mathbb{B}_\omega(r)\big) \leq \frac{c}{\sqrt{m}} \mathcal{J}_2\big((2\mathcal{F}_k) \cap \mathbb{B}_\omega(r), \|\cdot\|_\omega; [\delta, r]\big) \\ + \frac{c\sigma \log m}{m} \mathcal{J}_1\big((2\mathcal{F}_k) \cap \mathbb{B}_\omega(r), \|\cdot\|_\infty; [\delta, 2]\big) + 2\delta, \quad \text{(C.4)}$$

for any scalar $\delta \in [0, 2]$. Observing the norm domination relation $\|f\|_\omega \leq \sigma \|f\|_\infty$ for any function f, we have $\mathcal{J}_2\big((2\mathcal{F}_k) \cap \mathbb{B}_\omega(r), \|\cdot\|_\omega; [\delta, r]\big) \leq \mathcal{J}_2\big(2\sigma \mathcal{F}_k, \|\cdot\|_\infty; [\delta, r]\big)$. As a result, in order to control the right-hand-side of equation (C.4), it suffices to bound the covering number of the class $\mathcal{F}_k$ under the $\|\cdot\|_\infty$-norm.

In order to estimate the Dudley chaining integral for the localized class, we begin with the classical bound [103]

$$\log N\big(\mathcal{F}_k, \|\cdot\|_\infty; \varepsilon\big) \leq \Big(\frac{c}{\varepsilon}\Big)^{p/k},$$

where $c > 0$ is a universal constant. Using this bound, we can control the Dudley entropy integrals for any $\alpha \in \{1, 2\}$, $q > 0$, and interval $[\delta, u]$ with $u \in \{r, 2\}$. In particular, for any interval $[\delta, u]$ of the non-negative real line, we have

$$\mathcal{J}_\alpha\Big(q\mathcal{F}_k, \|\cdot\|_\infty; [\delta, u]\Big) \leq \int_\delta^u \Big(\frac{cq}{\varepsilon}\Big)^{\frac{p}{\alpha k}} d\varepsilon \leq cq^{\frac{p}{\alpha k}} \cdot \begin{cases} \frac{\alpha k}{\alpha k - p} u^{1 - \frac{p}{\alpha k}} & \text{if } p < \alpha k, \\ \log\big(u/\delta\big) & \text{if } p = \alpha k, \\ \frac{\alpha k}{p - \alpha k} \big(\frac{c}{\delta}\big)^{\frac{p}{\alpha k} - 1} & \text{if } p > \alpha k. \end{cases} \quad \text{(C.5)}$$

We set $\delta = \big(\frac{\sigma}{m}\big)^{k/p}$, and use the resulting upper bound on the Dudley integral to control the Rademacher complexity; doing so yields

$$\mathcal{R}_m\big((\mathcal{F}_k - \bar{\mu}) \cap \mathbb{B}_\omega(r)\big) \leq c_{\sigma, p/k} \cdot \begin{cases} r^{1 - \frac{p}{2k}} / \sqrt{m} + \log m \cdot m^{-k/p} & \text{if } p < 2k, \\ \log(m)/\sqrt{m} & \text{if } p = 2k, \\ m^{-k/p} & \text{if } p > 2k. \end{cases}$$

Solving the fixed point equation (4.14b) yields

$$r_m \leq c'_{\sigma, p/k} m^{-k/p} \cdot \log m,$$

where the constant $c_{\sigma, p/k}$ and $c'_{\sigma, p/k}$ depend on the parameters $(\sigma, p/k)$, along with the small ball constants (α_1, α_2).

Turning to the critical radius s_m, we note that each term in the empirical process associated with the observation noise satisfies

$$\mathbb{E}\Big[\Big\{\frac{g^2(X_i, A_i)}{\pi^2(X_i, A_i)} (Y_i - \mu^*(X_i, A_i)) \varepsilon_i (\mu_1 - \mu_2)(X_i, A_i)\Big\}^2\Big] \\ \leq \sqrt{\mathbb{E}\Big[\Big\{\frac{g(X_i, A_i)}{\pi(X_i, A_i)} (Y_i - \mu^*(X_i, A_i))\Big\}^4\Big]} \cdot \sqrt{\mathbb{E}\Big[\Big\{\frac{g(X_i, A_i)}{\pi(X_i, A_i)} (\mu_1 - \mu_2)(X_i, A_i)\Big\}^4\Big]} \\ \leq \bar{\sigma}^2 M_{2 \to 4} \|\mu_1 - \mu_2\|_\omega^2,$$

and

$$\begin{aligned}
&\|\frac{g^2(X_i,A_i)}{\pi^2(X_i,A_i)}(Y_i-\mu^*(X_i,A_i))\varepsilon_i(\mu_1-\mu_2)(X_i,A_i)\|_{\psi_1}\\
&\qquad\leq \|\frac{g(X_i,A_i)}{\pi(X_i,A_i)}(Y_i-\mu^*(X_i,A_i))\|_{\psi_2}\cdot\|\frac{g(X_i,A_i)}{\pi(X_i,A_i)}(\mu_1-\mu_2)(X_i,A_i)\|_{\psi_2}\\
&\qquad\qquad\leq \bar{\sigma}\sigma\|\mu_1-\mu_2\|_\infty.
\end{aligned}$$

Following the same line of derivation in the bound for the Rademacher complexity $\mathcal{R}_m$, we use the mixed-tail chaining bound to find that

$$\begin{aligned}
\mathcal{S}_m\big((\mathcal{F}_k-\bar{\mu})\cap\mathbb{B}_\omega(r)\big)\leq{}& \frac{c\bar{\sigma}\sqrt{M_{2\to4}}}{\sqrt{m}}\mathcal{J}_2\big((2\mathcal{F}_k)\cap\mathbb{B}_\omega(r),\|\cdot\|_\omega;[\delta,r]\big)\\
&+\frac{c\bar{\sigma}\sigma\log m}{m}\mathcal{J}_1\big((2\mathcal{F}_k)\cap\mathbb{B}_\omega(r),\|\cdot\|_\infty;[\delta,2]\big)+2\delta,
\end{aligned}$$

valid for all $\delta\in[0,2]$. The Dudley integral bound (C.5) then implies

$$\mathcal{S}_m\big((\mathcal{F}_k-\bar{\mu})\cap\mathbb{B}_\omega(r)\big)\leq c_{\sigma,p/k}\bar{\sigma}\cdot\begin{cases}r^{1-\frac{p}{2k}}/\sqrt{m}+\log m\cdot m^{-k/p} & \text{if } p<2k,\\ \log(m)/\sqrt{m} & \text{if } p=2k,\\ m^{-k/p} & \text{if } p>2k,\end{cases}$$

where the constant $c_{\sigma,p/k}$ depends on the parameters $(\sigma,p/k)$ and the constant $M_{2\to4}$. Solving the fixed point equation yields

$$s_m\leq c_{\sigma,p/k}\bar{\sigma}\cdot\begin{cases}m^{-\frac{k}{2k+p}} & \text{if } p<2k,\\ m^{-1/4}\sqrt{\log m} & \text{if } p=2k,\\ m^{-\frac{k}{2p}} & \text{if } p>2k.\end{cases}$$

Combining with Theorem 4.2 completes the proof of Corollary 4.3.

C.2.4 Proof of Corollary 4.4

For any $\bar{\mu}\in\mathcal{F}$, define the function class

$$\mathcal{H}:=\Big\{(x,a)\to\frac{g(x,a)}{\pi(x,a)}f(x,a)\;\mid\;f\in\mathbb{B}_\omega(r)\cap(\mathcal{F}-\bar{\mu})\Big\}.$$

Clearly, the class $\mathcal{H}$ is uniformly bounded by b, and for any $f\in\mathcal{F}$, we have the upper bound $\mathbb{E}\Big[\Big|\frac{g(X,A)}{\pi(X,A)}f(X,A)\Big|^2\Big]=\|f\|_\omega^2\leq r^2$.

Invoking a known bracketing bound on empirical processes (cf. Prop. C.6 in Appendix C.4), we have

$$\mathbb{E}\Big[\sup_{h\in\mathcal{H}}\frac{1}{m}\sum_{i=1}^m\varepsilon_i h(X_i,A_i)\Big]\leq\frac{c}{\sqrt{m}}\mathcal{J}_{\text{bra}}\big(\mathcal{H},\|\cdot\|_{\mathbb{L}^2};[0,r]\big)\Big\{1+\frac{b\mathcal{J}_{\text{bra}}\big(\mathcal{H},\|\cdot\|_{\mathbb{L}^2};[0,r]\big)}{r^2\sqrt{m}}\Big\}\tag{C.6}$$

For functions $\ell, f, u : [0,1] \to \mathbb{R}$, such that f is contained in the bracket $[\ell, u]$, we let:

$$\widetilde{\ell}(x,a) := \frac{g(x,a)}{\pi(x,a)} \Big\{ \ell(\phi(x,a)) \mathbf{1}_{g(x,a)>0} + u(\phi(x,a)) \mathbf{1}_{g(x,a)<0} - \bar{\mu}(x,a) \Big\},$$
$$\widetilde{u}(x,a) := \frac{g(x,a)}{\pi(x,a)} \Big\{ u(\phi(x,a)) \mathbf{1}_{g(x,a)>0} + \ell(\phi(x,a)) \mathbf{1}_{g(x,a)<0} - \bar{\mu}(x,a) \Big\}.$$

It is easily observed that the function $(x,a) \mapsto \frac{g(x,a)}{\pi(x,a)} (f - \bar{\mu})(x,a)$ lies in the bracket $[\widetilde{\ell}, \widetilde{u}]$, and for any probability law $\mathbb{Q}$ on $\mathbb{X} \times \mathbb{A}$, we have $\|\widetilde{u} - \widetilde{\ell}\|_{\mathbb{L}^2(\mathbb{Q})} \leq b \cdot \|u - \ell\|_{\mathbb{L}^2(\mathbb{Q}_\phi)}$, where $\mathbb{Q}_\phi$ is the probability law of $\phi(X, A)$ for $(X, A) \sim \mathbb{Q}$.

It is known (cf. Thm 2.7.5 in the book [208]) that the space of monotonic functions from $[0,1]$ to $[0,1]$ has ε-bracketing number under any $\mathbb{L}^2$-norm bounded by $\exp\big(c/\varepsilon\big)$ for any $\varepsilon > 0$. Substituting back into the bracketing entropy bound (C.6) yields

$$\mathcal{R}_m\big(\mathbb{B}_\omega(r) \cap (\mathcal{F} - \bar{\mu})\big) \leq c \Big\{ \sqrt{\frac{br}{m}} + \frac{b^2}{rm} \Big\}.$$

From the definition of the fixed point equation, we can bound the critical radius r as

$$r_m \leq \frac{cb}{m} + \frac{cb}{\sqrt{m}},$$

where $c > 0$ is a universal constant.

Turning to the squared Rademacher process associated with the outcome noise, we construct the function class

$$\mathcal{H}' := \Big\{ (x,a,y) \to y \cdot \frac{g^2(x,a)}{\pi^2(x,a)} f(x,a) \ \mid \ f \in \mathbb{B}_\omega(r) \cap (\mathcal{F} - \bar{\mu}) \Big\}.$$

For functions $\ell, f, u : [0,1] \to \mathbb{R}$, such that f is contained in the bracket $[\ell, u]$, we can similarly construct

$$\widetilde{\ell}(x,a,y) := y \cdot \frac{g^2(x,a)}{\pi^2(x,a)} \Big\{ \ell(\phi(x,a)) \mathbf{1}_{y>0} + u(\phi(x,a)) \mathbf{1}_{y<0} - \bar{\mu}(x,a) \Big\},$$
$$\widetilde{u}(x,a,y) := y \cdot \frac{g^2(x,a)}{\pi^2(x,a)} \Big\{ u(\phi(x,a)) \mathbf{1}_{y>0} + \ell(\phi(x,a)) \mathbf{1}_{y<0} - \bar{\mu}(x,a) \Big\}.$$

It is easily observed that the function $(x, ay) \mapsto y \cdot \frac{g(x,a)}{\pi(x,a)} (f - \bar{\mu})(x,a)$ lies in the bracket $[\widetilde{\ell}, \widetilde{u}]$, and for any probability law $\mathbb{Q}$ on $\mathbb{X} \times \mathbb{A} \times \mathbb{R}$, we have $\|\widetilde{u} - \widetilde{\ell}\|_{\mathbb{L}^2(\mathbb{Q})} \leq b^2 \cdot \|u - \ell\|_{\mathbb{L}^2(\mathbb{Q}_\phi)}$, where $\mathbb{Q}_\phi$ is the probability law of $\phi(X, A)$ for $(X, A, Y) \sim \mathbb{Q}$. Applying the bracketing bound yields

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{h \in \mathcal{H}'} \frac{1}{m} \sum_{i=1}^m \varepsilon_i h(X_i, A_i, Y_i - \mu^*(X_i, A_i)) \Big] \\
&\leq \frac{c}{\sqrt{m}} \mathcal{J}_{\text{bra}}\big(\mathcal{H}', \|\cdot\|_{\mathbb{L}^2}; [0, br]\big) \Big\{ 1 + \frac{b \mathcal{J}_{\text{bra}}\big(\mathcal{H}', \|\cdot\|_{\mathbb{L}^2}; [0, br]\big)}{(br)^2 \sqrt{m}} \Big\} \\
&\leq cb \Big(\sqrt{\frac{r}{m}} + \frac{1}{rm} \Big).
\end{aligned}$$

Denote $Z_m := \sup_{h \in \mathcal{H}'} \frac{1}{m} \sum_{i=1}^m \varepsilon_i h(X_i, A_i, Y_i - \mu^*(X_i, A_i))$. By a standard functional Bernstein bound (e.g., Thm. 3.8 in the book [213]), we have the tail bound

$$\mathbb{P}\Big[Z_m \geq 2\mathbb{E}[Z_m] + t\Big] \leq 2 \exp\left(\frac{-mt^2}{56(br)^2 + 4b^2 t}\right) \quad \text{for any } t > 0.$$

Combining with the expectation bound, we conclude that $\mathcal{S}_m = \sqrt{\mathbb{E}[Z_m^2]} \leq 2cb\big(\sqrt{\frac{r}{m}} + \frac{1}{rm}\big)$. By definition of fixed point equation, the critical radius can be upper bounded $s_m \leq c\big(\frac{b^2}{m}\big)^{1/3}$, and substituting this bound into Theorem 4.2 completes the proof of this corollary.

C.2.5 Strong shattering for sparse linear models

In this section, we state and prove the claim from Example 4.3 about the size of the fat shattering dimension for the class of sparse linear models.

Proposition C.3. *There is a universal constant $c > 0$ such that the function class $\mathcal{F}_s$ of s-sparse linear models over $\mathbb{R}^p$ satisfies the strong shattering condition (4.42) with fat shattering dimension $D = cs \log(e\, p/s)$ at scale $\delta = 1$.*

Proof. We assume without loss of generality (adjusting constants as needed) that $p/s = 2^k$ is an integer power of two. Our argument involves constructing a set of vectors by dividing the p coordinates into s blocks. Let the matrix $A \in \{0, 1\}^{k \times 2^k}$ be such that by sequentially writing down the elements in j-th column, we get the binary representation of the integer $(j-1)$, for $j = 1, 2, \ldots, 2^k$. Let $(a_i^\top)_{1 \leq i \leq k}$ be the row vectors of the matrix A. For $i \in [k]$ and $j \in s$, we construct the p-dimensional data vector as $x_{i,j} = a_i \otimes e_j$, where the $e_j \in \mathbb{R}^s$ is the indicator vector of j-th coordinate. The cardinality of this set is given by

$$\Big|\Big\{x_{i,j} : \; i \in [k], j \in s\Big\}\Big| = ks = \frac{1}{\log 2} \cdot s \log(p/s).$$

It suffices to construct a hypercube packing for this set. Given a binary vector $v \in \{0, 1\}^k$, we let $J(v) \in \{1, 2, \ldots, 2^k\}$ such that the $J(v)$-th column of the matrix A is equal to v. (Note that our construction ensures that such a column always exists and is unique.)

Given any binary vector $\zeta \in \{0, 1\}^{k \times s}$, we construct the following vector:

$$\beta_\zeta := \sum_{i=1}^s e\big(J(\zeta_{i,1}, \zeta_{i,2}, \ldots, \zeta_{i,k})\big) \otimes e_i$$

where the function $e : [2^k] \to \mathbb{R}^{2^k}$ maps the integer j to the indicator vector of j-th coordinate.

We note that the vector β is supported on s-coordinates, with absolute value of each coordinate bounded by 1. Moreover, our construction ensures that $i \in [k]$ and $j \in [s]$,

$$\beta_\zeta^\top x_{i,j} = a_i^\top e\big(J(\zeta_{i,1}, \zeta_{i,2}, \ldots, \zeta_{i,k})\big) = \zeta_{i,j}.$$

Therefore, we have planted a hypercube $\prod_{i \in [k], j \in s} \big(x_{i,j}, \{0,1\}\big)$ in the graph of the function class $\mathcal{F}_s^{sparse}$, which completes the proof of the claim. □

C.3 Some elementary inequalities and their proofs

In this section, we collect some elementary results used throughout Chapter 4 and this appendix, as well as their proofs.

C.3.1 Bounds on conditional total variation distance

The following lemma is required for the truncation arguments used in the proofs of our minimax lower bounds. In particular, it allows us to make small modifications on a pair of probability laws by conditioning on good events, without inducing an overly large change in the total variation distance.

Lemma C.1. *Let (μ, ν) be a pair of probability distributions over the same Polish space $\mathcal{S}$, and consider a subset $\mathscr{E} \subseteq \mathcal{S}$ such that $\min\big\{\mu(\mathscr{E}), \nu(\mathscr{E})\big\} \geq 1 - \varepsilon$ for some $\varepsilon \in [0, 1/4]$. Then the conditional distributions $(\mu \mid \mathscr{E})$ and $(\nu \mid \mathscr{E})$ satisfy the bound*

$$d_{\mathrm{TV}}(\mu, \nu) - 4\varepsilon \overset{(i)}{\leq} d_{\mathrm{TV}}\big[(\mu \mid \mathscr{E}), (\nu \mid \mathscr{E})\big] \overset{(ii)}{\leq} \tfrac{1}{1-\varepsilon} d_{\mathrm{TV}}(\mu, \nu) + 2\varepsilon. \tag{C.7}$$

Proof. Recall the variational definition of the TV distance as the supremum over functions $f : \mathcal{S} \to \mathbb{R}$ such that $\|f\|_\infty \leq 1$. For any such function f, we have

$$\begin{aligned}
&|\mathbb{E}_\mu[f(X)] - \mathbb{E}_\nu[f(X)]| \\
&\leq |\mathbb{E}_\mu[f(X)\mathbf{1}_{X \in \mathscr{E}}] - \mathbb{E}_\nu[f(X)\mathbf{1}_{X \in \mathscr{E}}]| + \mathbb{E}_\mu[|f(X)|\ \mathbf{1}_{\mathscr{E}^c}] + \mathbb{E}_\nu[|f(X)|\ \mathbf{1}_{X \in \mathscr{E}^c}] \\
&\leq \left| \frac{\mathbb{E}_\mu[f(X)\mathbf{1}_{X \in \mathscr{E}}]}{\mu(\mathscr{E})} - \frac{\mathbb{E}_\nu[f(X)\mathbf{1}_{X \in \mathscr{E}}]}{\nu(\mathscr{E})} \right| + \left| \frac{1}{\mu(\mathscr{E})} - \frac{1}{\nu(\mathscr{E})} \right| \mathbb{E}_\nu[|f(X)|] + 2\varepsilon \\
&\leq d_{\mathrm{TV}}\big((\mu \mid \mathscr{E}), (\nu \mid \mathscr{E})\big) + 4\varepsilon,
\end{aligned}$$

and re-arranging yields the lower bound (i).

On the other hand, in order to prove the upper bound (ii), we note that

$$\begin{aligned}
&\big|\mathbb{E}_{\mu|\mathscr{E}}[f(X)] - \mathbb{E}_{\nu|\mathscr{E}}[f(X)]\big| \\
&= \frac{1}{\mu(\mathscr{E})} \left| \mathbb{E}_\mu[f(X)\mathbf{1}_{X \in \mathscr{E}}] - \mathbb{E}_\nu[f(X)\mathbf{1}_{X \in \mathscr{E}}] \frac{\mu(\mathscr{E})}{\nu(\mathscr{E})} \right| \\
&\leq \frac{1}{\mu(\mathscr{E})} |\mathbb{E}_\mu[f(X)\mathbf{1}_{X \in \mathscr{E}}] - \mathbb{E}_\nu[f(X)\mathbf{1}_{X \in \mathscr{E}}]| + \mathbb{E}_\nu[|f(X)|] \cdot \left| \frac{\mu(\mathscr{E})}{\nu(\mathscr{E})} - 1 \right| \\
&\leq \frac{1}{1-\varepsilon} d_{\mathrm{TV}}(\mu, \nu) + 2\varepsilon,
\end{aligned}$$

which completes the proof. □

C.3.2 A second moment lower bound for truncated random variable

The following lemma is frequently used in our lower bound constructions.

Lemma C.2. *Let X be a real-valued random variable with finite fourth moment, and define the* (2–4)*-moment constant* $M_{2\to 4} := \sqrt{\mathbb{E}[X]^4}/\mathbb{E}[X^2]$*. Then we have the lower bound*

$$\mathbb{E}\Big[X^2 \cdot \mathbf{1}\big\{|X| \leq 2M_{2\to 4}\sqrt{\mathbb{E}[X^2]}\big\}\Big] \geq \frac{1}{2}\mathbb{E}[X^2].$$

Proof. Without loss of generality, we can assume that $\mathbb{E}[X^2] = 1$. Applying Cauchy–Schwarz inequality implies that

$$\mathbb{E}\Big[X^2\mathbf{1}\big\{|X| \geq 2M_{2\to 4}\big\}\Big] \leq \sqrt{\mathbb{E}\big[X^4\big]} \cdot \sqrt{\mathbb{P}\big(|X| \geq 2M_{2\to 4}\big)} \leq M_{2\to 4} \cdot \sqrt{\mathbb{P}\big(|X| \geq 2M_{2\to 4}\big)}.$$

By Markov's inequality, we have

$$\mathbb{P}\Big(|X| \geq 2M_{2\to 4}\Big) \leq \frac{\mathbb{E}[X^2]}{4M_{2\to 4}^2} = \frac{1}{4M_{2\to 4}^2}.$$

Substituting back to above bounds, we conclude that $\mathbb{E}\big[X^2\mathbf{1}\big\{|X| \geq 2M_{2\to 4}\big\}\big] \leq \frac{1}{2}$, and consequently,

$$\mathbb{E}\Big[X^2\mathbf{1}\big\{|X| \leq 2M_{2\to 4}\big\}\Big] = \mathbb{E}[X^2] - \mathbb{E}\Big[X^2\mathbf{1}\big\{|X| \geq 2M_{2\to 4}\big\}\Big] \geq \frac{1}{2},$$

which completes the proof. □

C.4 Empirical process results from existing literature

In this appendix, we collect some known bounds on the suprema of empirical processes.

C.4.1 Concentration for unbounded empirical processes

We use a concentration inequality for unbounded empirical processes. It applies to a countable class $\mathcal{F}$ of measurable functions, and a supremum of the form

$$Z := \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} f(X_i) \right|$$

where $\{X_i\}_{i=1}^n$ is a sequence of independent random variables such that $\mathbb{E}[f(X_i)] = 0$ for any $f \in \mathcal{F}$.

Proposition C.4 (Theorem 4 of [2], simplified)**.** *There exists a universal constant $c > 0$ such that for any $t > 0$ and $\alpha \geq 1$, we have*

$$\mathbb{P}\left[Z > 2\mathbb{E}(Z) + t\right] \leq \exp\left(\frac{-t^2}{4v^2}\right) + 3\exp\Big(-\Big(\frac{t}{c\| \max\limits_{i=1,\ldots,n} \sup\limits_{f \in \mathcal{F}} |f(X_i)| \|_{\psi_{1/\alpha}}}\Big)^{1/\alpha}\Big),$$

where $v^2 := \sup_{f\in\mathcal{F}} \sum_{i=1}^n \mathbb{E}[f^2(X_i)]$ is the maximal variance.

The countability assumption can be easily relaxed for separable spaces. A useful special case of Proposition C.4 is by taking the class $\mathcal{F}$ to be a singleton and letting $\alpha = 1$, in which case the bound becomes

$$\left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right| \leq c\sqrt{\operatorname{var}\big(f(X)\big)\frac{\log(1/\delta)}{n}} + \frac{c\log n}{n}\|f(X)\|_{\psi_1} \cdot \log(1/\delta),$$

with probability $1 - \delta$.

C.4.2 Some generic chaining bounds

We also use a known generic chaining tail bound. It involves a separable stochastic process $(Y_t)_{t\in T}$ and a pair (d_1, d_2) of metrics over the index set T. We assume that there exists some $t_0 \in T$ such that $Y_{t_0} \equiv 0$.

Proposition C.5 (Theorem 3.5 of Dirksen [51])**.** *Suppose that for any pair $s, t \in T$, the difference $Y_s - Y_t$ satisfies the mixed tail bound*

$$\mathbb{P}\Big(|Y_s - Y_t| \geq \sqrt{u}d_1(s,t) + ud_2(s,t)\Big) \leq 2e^{-u} \quad \textit{for any } u > 0. \tag{C.8a}$$

Then for any $\ell \geq 1$, we have the moment bound

$$\Big\{\mathbb{E}\Big[\sup_{t\in T}|Y_t|^\ell\Big]\Big\}^{1/\ell} \leq c\Big(\gamma_2(T, d_1) + \gamma_1(T, d_2)\Big) + 2\sup_{t\in T}\left(\mathbb{E}\,|Y_t|^\ell\right)^{1/\ell}, \tag{C.8b}$$

where $\gamma_\alpha(T, d)$ is the generic chaining functional of order α for the metric space (T, d).

For a set T with diameter bounded by r under the metric d, the generic chaining functional can be upper bounded in terms of the Dudley entropy integral as

$$\gamma_\alpha(T, d) \leq c\mathcal{J}_\alpha\big(T, d; [0, r]\big) \quad \text{for each } \alpha \in \{1, 2\}$$

(e.g., cf. Talagrand [202]). Furthermore, suppose that the norm domination relation $d_1(s, t) \leq a_0 d_2(s, t)$ holds true for any pair $s, t \in T$. Let r_1, r_2 be the diameter of the set T under the metrics d_1, d_2, respectively. If we apply Proposition C.5 to a maximal δ-packing for the set T under metric d_1, we immediately have

$$\begin{aligned}\Big\{\mathbb{E}\Big[\sup_{t\in T}|Y_t|^p\Big]\Big\}^{1/p} &\leq c\Big\{\mathcal{J}_2\big(T, d_1; [\delta, r_1]\big) + \mathcal{J}_1(T, d_2; [\delta/a_0, r_2])\Big\} \\ &\quad + \Big\{\mathbb{E}\sup_{\substack{s,t\in T\\ d_1(s,t)\leq\delta}}|Y_s - Y_t|^p\Big\}^{1/p} + 2\sup_{t\in T}\big(\mathbb{E}\,|Y_t|^p\big)^{1/p}. \qquad \text{(C.9)}\end{aligned}$$

C.4.3 Bracketing entropy bounds

Finally, we use the following bracketing integral bound for empirical processes:

Proposition C.6 (Lemma 3.4.2 of [208])**.** *Let $\mathcal{F}$ be a class of measurable functions, such that $\mathbb{E}[f^2(X)] \leq r^2$ and $|f(X)| \leq M$ almost surely for any $f \in \mathcal{F}$. Given n* i.i.d. *samples $\{X_i\}_{i=1}^n$, we have*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\Big] \leq \frac{c}{\sqrt{n}}\mathcal{J}_{\text{bra}}\big(\mathcal{F}, \|\cdot\|_{\mathbb{L}^2}; [0,r]\big)\Big\{1 + \frac{M\mathcal{J}_{\text{bra}}\big(\mathcal{F}, \|\cdot\|_{\mathbb{L}^2}; [0,r]\big)}{r^2\sqrt{n}}\Big\}.$$

Appendix D

Proofs and discussion deferred from Chapter 5

D.1 Properties of effective dimension

In this section, we develop various bounds on the effective dimension under decay rates on the eigenvalues, along with some regularity conditions on the eigenfunctions.

D.1.1 Regularity conditions on eigenfuncions

The most straightforward assumption on the eigenfunctions is the *uniform boundedness condition*

$$\|\phi_j\|_\infty = \sup_{(x,a)} |\phi_j(x,a)| < \infty \qquad \text{for all } j = 1, 2, \ldots. \tag{D.1}$$

This condition appears frequently in the literature [219, 139, 54, 163], but as noted, it is not satisfied by all kernels. See paper [231] and Appendix D.5.1 for some natural counterexamples.

In this chapter, we consider the following relaxed growth condition: there exists a scalar $\nu \in [0, 1/2)$, such that the sup-norm of eigenfunctions satisfy the bound

$$\phi_{\max} := \sup_{j\geq 1} \sup_{(x,a)} \lambda_j^\nu |\phi_j(x,a)| < \infty. \tag{Eig(ν)}$$

We note that the requirement $\nu \in [0, 1/2)$ is natural, since the kernel boundedness condition (Kbou(κ)) implies that condition (Eig(ν)) holds with $\nu = 1/2$ and $\phi_{\max} = \kappa$. An exponent ν strictly less than $1/2$ guarantees slightly more regularity. The growth condition (Eig(ν)) with $\nu = 0$ is equivalent to the uniform boundedness conditiion (D.1). However, when $\nu > 0$, the relaxed condition, on the other hand, is much weaker. For example, it is shown by Mendelson and Neeman [139] that the counterexample in the paper [231] satisfies equation (Eig(ν)) for any $\nu > 0$.

Under Assumption (Eig(ν)), we have the upper bound

$$D(\rho) \leq \sum_{j=1}^{\infty} \sup_{(x,a)} \frac{\lambda_j \phi_j^2(x,a)}{\lambda_j + \rho} \leq \phi_{\max}^2 \sum_{j=1}^{\infty} \frac{\lambda_j^{1-2\nu}}{\lambda_j + \rho}. \tag{D.2}$$

This upper bound, when combined with decay conditions on the eigenvalue sequence $\{\lambda_j\}_{j=1}^{\infty}$, allows us to derive explicit bounds on the effective dimension. Two natural classes of eigenvalue decay are the polynomial condition

$$\lambda_j \leq \lambda_0 j^{-\alpha} \qquad \text{for some } \alpha > 1, \tag{D.3a}$$

and the *exponential decay*

$$\lambda_j \leq \lambda_0 \exp(-c_0 j) \qquad \text{for some } c_0 > 0. \tag{D.3b}$$

Proposition D.1. *Under Assumptions* (Kbou(κ)) *and* (Eig(ν)), *we have*

(a) For eigenvalues with α-polynomial-decay (D.3a) *for some $\alpha > \frac{1}{1-2\nu}$, we have*

$$D(\rho) \leq c\rho^{-\frac{1}{\alpha}-2\nu}. \tag{D.4a}$$

(b) For eigenvalues with exponential decay (D.3b), *we have*

$$D(\rho) \leq c \log\left(\tfrac{\lambda_0}{\rho}\right). \tag{D.4b}$$

In these bounds, the constant c can depend on problem parameters $(\lambda_0, \alpha, c_0, \nu)$ but is independent of ρ. See Appendix D.1.2 for the proof.

In our main theorems, the bounds on the effective dimension is used to establish the sample size requirement (5.21a) and (5.28a). In order for them to be true, up to logarithmic factors of $(n, \sigma, \underline{\sigma}^{-1}, R)$, we need sample sizes

$$n \gtrsim \left(\sigma/\underline{\sigma}\right)^{\frac{2\alpha}{\alpha-1-2\nu\alpha}} \cdot \left(R/\bar{\sigma}\right)^{\frac{1+2\alpha\nu}{\alpha-1-2\nu\alpha}}, \quad \text{under Proposition D.1(a)},$$
$$n \gtrsim \left(\sigma/\underline{\sigma}\right)^{2}, \quad \text{under Proposition D.1(b)}.$$

In words, the sample size requirement depends on two important objects: the tail conditions of the observation noise $W = Y - \mu^*(X, A)$, measured by the ratio between its largest Orlicz norm and smallest variance; and the richness of the kernel class, measured by the eigenvalue decay rates and the radius of the RKHS ball.

D.1.2 Proof of Proposition D.1

Since the eigenvalue sequence converges to zero, the cut-off integer $J := \sup\{j \geq 1 \mid \lambda_j > \rho\}$ is guaranteed to be finite. By the definition of the effective dimension, we have

$$D(\rho) = \sum_{j=1}^{\infty} \frac{\lambda_j^{1-2\nu}}{\lambda_j + \rho} \leq \sum_{j \leq J} \lambda^{-2\nu} + \rho^{-1} \sum_{j > J} \lambda_j^{1-2\nu} \leq \rho^{-2\nu} J + \rho^{-1} \sum_{j > J} \lambda_j^{1-2\nu}.$$

We prove the results for two cases separately.

For the polynomially-decaying eigenvalues, we have $J \leq \big(\frac{\lambda_0}{\rho}\big)^{1/\alpha}$, and

$$\sum_{j>J} \lambda_j^{1-2\nu} \leq \lambda_0 \sum_{j>J} j^{-\alpha(1-2\nu)} < \frac{1}{\alpha(1-2\nu)-1} J^{1-\alpha(1-2\nu)}.$$

Combining these bounds yields

$$D(\rho) \leq \lambda_0^{1/\alpha} \rho^{-2\nu-1/\alpha} + \frac{1}{\alpha(1-2\nu)-1} \lambda_0^{1/\alpha-1+2\nu} \rho^{-1/\alpha-2\nu} \leq c(\lambda_0, \nu, \alpha) \rho^{-1/\alpha-2\nu}.$$

For exponentially-decaying eigenvalues, we have $J \leq c_0^{-1} \log\big(\frac{\lambda_0}{\rho}\big)$, and

$$\sum_{j>J} \lambda_j^{1-2\nu} \leq \lambda_0 \sum_{j>J} \exp\big(-c_0(1-2\nu)j\big) \leq \frac{\lambda_0}{(1-2\nu)c_0} \exp(-c_0(1-2\nu)J) \leq \frac{\rho}{c_0},$$

which leads to the effective dimension bound

$$D(\rho) \leq c(c_0, \lambda_0, \nu) \log\big(\frac{\lambda_0}{\rho}\big),$$

completing the proof of Proposition D.1.

D.2 Relaxing the effective dimension condition

Recall that Theorem 5.2 requires certain growth conditions on the effective dimension. In this section, we discuss how these conditions can be relaxed, thereby obtaining a bound that remains instance-optimal up to logarithmic factors.

D.2.1 Near-optimal rates

Our result involves the modified regularization parameter

$$\rho_n = \frac{\bar{\sigma}^2}{R^2 n} \vee \frac{32\kappa^2 \log(n/\delta)}{n}, \tag{D.5}$$

along with the modified higher-order term $\mathcal{H}_n' := \big(\sigma + \kappa R\big) \frac{\log(n/\delta)}{\sqrt{n}}$.

Corollary D.1. *Suppose that Assumptions Kbou(κ) and subG(σ) are in force, and we that implement the method with regularization parameter (D.5). Then for for any sample size $n \geq 2$ and any $\delta \in (0,1)$, we have*

$$|\widehat{\tau}_n - \tau^*| \leq c v_\xi(\mu^*) \sqrt{\frac{\log(1/\delta)}{n}} + \frac{c(\sigma + \kappa R)}{\bar{\sigma}} \bar{v}_\sigma\big(\mathbb{B}_{\mathbb{H}}(R); n\big) \frac{\log(n/\delta)}{\sqrt{n}} + c\mathcal{H}_n' \tag{D.6}$$

with probability at least $1 - \delta$.

See Appendix D.2.2 for the proof.

A few remarks are in order. First, Corollary D.1 holds for *any* sample size, and is completely agnostic to conditions on the effective dimension D. Compared to the optimal instance-dependent bounds in Theorem 5.2, Corollary D.1 exhibits two differences:

- The variance functional $V_{\bar{\sigma},n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R))$ is multiplied with a problem-dependent factor $\frac{\sigma+\kappa R}{\bar{\sigma}}$, as well as logarithmic factors in the ratio n/δ.
- The high-order term $\mathcal{H}'_n$ is of order $\Theta(n^{-1/2})$, with additional logarithmic factors. Such a convergence rate is slower than the high-order term bound $\mathcal{H}_n$ established in Theorem 5.2, which decays at a rate $o(n^{-1/2})$

Due to these two major differences, the bound in Corollary D.1 may not be always instance-optimal. However, we remark that the near-optimal rate of convergence (as a function of sample size n) is still preserved. In the high-noise regime where the quantities $(\sigma, \bar{\sigma}, \kappa R)$ are of the same order, the leading-order terms in Theorem 5.2 and Corollary D.1 differ only by logarithmic factors. The term $\mathcal{H}'$ is dominated by the leading-order one, up to logarithmic factors. In combination, results in Corollary D.1 under the weak assumptions can be worse than Theorem 5.2 only by logarithmic factors and problem dependent constants. Note that a variety of convergence rates can be established beyond the classical $\sqrt{n}$-regime (see Section 5.3.3 for concrete examples). These convergence rates, though depending on the intricate properties of the policy π, are automatically achieved without the effective dimension condition.

D.2.2 Proof of Corollary D.1

We use the same notation $(\widehat{u}_n, \beta_n, \beta_*)$ as in the proof of Theorem 5.2. Recall from the decomposition (5.47) that $\widehat{\tau}_n - \tau^* = \langle \widehat{u}_n - \bar{u},\, \beta_* \rangle + \langle \bar{u},\, \widehat{\beta}_n - \beta_* \rangle + \langle \widehat{u}_n - \bar{u},\, \widehat{\beta}_n - \beta_* \rangle$. Since the bound (5.50) does not rely on the effective dimension, it still holds under our current assumptions—that is, we have

$$|\langle \widehat{u}_n - \bar{u},\, \beta_* \rangle| \leq 2V_{\xi^*}(\mu^*)\sqrt{\tfrac{\log(1/\delta)}{n}} + 6\kappa R \tfrac{\log(1/\delta)}{n}, \tag{D.7}$$

with probability $1-\delta$.

The rest of this section is devoted to the control of the other two terms in the decomposition. We use the following lemma, which is analogous to Lemma 5.2.

Lemma D.1. *Uner the assumptions of Corollary D.1, for any fixed $z \in \ell^2$, we have*

$$\Big|\langle z,\, \widehat{\beta}_n - \beta_* \rangle\Big| \leq c \big\| \big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} z \big\|_{\ell^2} \cdot (\sigma + \kappa R) \sqrt{\frac{\log(n/\delta)\log(1/\delta)}{n}}.$$

with probability at least $1-\delta$.

See Appendix D.3.3 for the proof.

Taking this lemma as given, we proceed with the proof of Corollary D.1. Applying Lemma D.1 with $z = \bar{u}$, we have

$$\begin{aligned}
\Big|\langle \bar{u},\, \widehat{\beta}_n - \beta_* \rangle\Big| &\leq c\big\|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2}\bar{u}\big\|_{\ell^2} \cdot (\sigma + \kappa R)\frac{\log(n/\delta)}{\sqrt{n}} \\
&\leq c\big\|\big(I + \frac{\bar{\sigma}^2}{R^2 n}\mathbf{\Lambda}^{-1}\big)^{-1/2} z\big\|_{\ell^2} \cdot (\sigma + \kappa R)\sqrt{\frac{\log(n/\delta)\log(1/\delta)}{n}} \\
&\leq 4c\bar{v}_\sigma\big(\mathbb{B}_{\mathbb{H}}(R); n\big) \cdot \frac{\sigma + \kappa R}{\bar{\sigma}} \cdot \frac{\log(n/\delta)}{\sqrt{n}}, \qquad \text{(D.8)}
\end{aligned}$$

with probability at least $1 - \delta$.

In our next step, we apply equation (5.48b) from Lemma 5.1, as well as condition (Kbou(κ)). Doing so yields

$$\big\|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2}(\widehat{u}_n - \bar{u})\big\|_{\ell^2} \leq c\sqrt{\frac{D(\rho_n)}{n}\log(1/\delta)} \leq c\kappa\sqrt{\frac{\log(1/\delta)}{\rho_n n}} \leq c,$$

with probability $1 - \delta$.

Combining with Lemma D.1 yields

$$\begin{aligned}
\Big|\langle \widehat{u}_n - \bar{u},\, \widehat{\beta}_n - \beta_* \rangle\Big| &\leq c\big\|\big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2}(\widehat{u}_n - \bar{u})\big\|_{\ell^2} \cdot (\sigma + \kappa R)\frac{\log(n/\delta)}{\sqrt{n}} \\
&\leq c'(\sigma + \kappa R)\frac{\log(n/\delta)}{\sqrt{n}}. \qquad \text{(D.9)}
\end{aligned}$$

Combining equations (D.7), (D.8) and (D.9) completes the proof of Corollary D.1.

D.3 Proof of technical lemmas

We collect the proofs of auxiliary lemmas in the proof of Theorem 5.2 in this section.

D.3.1 Proof of Lemma 5.3

We start with the decomposition

$$z^\top\big(\widehat{\beta}_n - \beta_*\big) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} W_i}_{\text{noise part}} - \underbrace{\rho_n z^\top \mathbf{\Lambda}^{-1}\beta_*}_{\text{bias part}},$$

where $W_i := \varepsilon_i\, z^\top\big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1}\phi(X_i, A_i)$.

Beginning with the bias term, we note that

$$\begin{aligned} \Big| z^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} \cdot \rho_n \mathbf{\Lambda}^{-1} \beta_* \Big| &= \rho_n \cdot \Big| \langle \mathbf{\Lambda}^{-1/2} \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} z,\, \mathbf{\Lambda}^{-1/2} \beta_* \rangle \Big| \\ &\leq \rho_n \|\mu^*\|_{\mathbb{H}} \cdot \|\mathbf{\Lambda}^{-1/2} \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} z\|_{\ell^2} \\ &\leq \sqrt{\rho_n} \|\mu^*\|_{\mathbb{H}} \cdot \|\big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} z\|_{\ell^2}. \qquad \text{(D.10)} \end{aligned}$$

Here the final step is based on the fact that $\|\mathbf{A}x\|_{\ell^2} \leq \|\mathbf{B}x\|_{\ell^2}$ for any pair $(\mathbf{A}, \mathbf{B})$ of operators such that $\mathbf{A} \preceq \mathbf{B}$.

For the stochastic part, we note that the noise variables $\{\varepsilon_i\}_{i=1}^n$ are independent conditioned on $(X_i, A_i)_{i=1}^n$. For each $i \in [n]$, the conditional variance takes the form

$$\mathrm{var}\Big(W_i \mid (X_i, A_i)_{i=1}^n\Big) = \sigma^2(X_j, A_j) \cdot \Big|\phi(X_i, A_i)^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} z\Big|^2.$$

Summing up these relations yields

$$\begin{aligned} &\mathrm{var}\Big(\frac{1}{n}\sum_{i=1}^n W_i \mid (X_i, A_i)_{i=1}^n\Big) \\ &= z^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} \sum_{i=1}^n \sigma^2(X_j, A_j) \phi(X_i, A_i) \phi(X_i, A_i)^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} z \\ &\leq \bar{\sigma}^2 \|\widehat{\mathbf{\Gamma}}_n^{1/2} \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} z\|_2^2. \end{aligned}$$

Introducing the shorthand $\mathbf{M} = \widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}$, by the noise tail assumption (subG(σ)) and Adamczak's concentration inequality [2], conditionally on $(X_i, A_i)_{i=1}^n$, we have

$$\begin{aligned} &\Big| z^\top \mathbf{M}^{-1} \cdot \tfrac{1}{n} \sum_{i=1}^n \big\{\varepsilon_i \phi(X_i, A_i)\big\} \Big| \\ &\qquad \leq c\bar{\sigma} \|\widehat{\mathbf{\Gamma}}_n^{1/2} \mathbf{M}^{-1} z\|_{\ell^2} \sqrt{\tfrac{\log(1/\delta)}{n}} + \max_{i \in [n]} \big| z^\top \mathbf{M}^{-1} \phi(X_i, A_i) \big| \tfrac{\sigma \log n \log(1/\delta)}{n}, \end{aligned}$$

with probability $1 - \delta$.

In order to control the max term on the RHS, we invoke the Cauchy–Schwarz inequality, thereby finding that

$$\big| z^\top \mathbf{M}^{-1} \phi(X_i, A_i) \big| \leq \|\mathbf{M}^{-1/2} z\|_{\ell^2} \cdot \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \|\mathbf{M}^{-1/2} \phi(x, a)\|_{\ell^2}.$$

Combining above two bounds yields

$$\begin{aligned} &\Big| z^\top \mathbf{M}^{-1} \cdot \tfrac{1}{n} \sum_{i=1}^n \Big\{\varepsilon_i \phi(X_i, A_i)\Big\} \Big| \\ &\quad \leq c \| \big(\mathbf{M}^{-1/2} z\|_{\ell^2} \Big\{ \bar{\sigma} \sqrt{\tfrac{\log(1/\delta)}{n}} + \sigma \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \|\mathbf{M}^{-1/2} \phi(x, a)\|_{\ell^2} \cdot \tfrac{\log(1/\delta) \log n}{n} \Big\}, \qquad \text{(D.11)} \end{aligned}$$

with probability $1 - \delta$.

Combining equations (D.10) and (D.11), we conclude that

$$\Big|\langle z,\, \widehat{\beta}_n - \beta_* \rangle\Big| \leq c \|\mathbf{M}^{-1/2} z\|_{\ell^2} \cdot \Big\{ \sqrt{\rho_n} \|\mu^*\|_{\mathbb{H}} + \bar{\sigma} \sqrt{\tfrac{\log(1/\delta)}{n}} + \sigma \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \|\mathbf{M}^{-1/2} \phi(x,a)\|_{\ell^2} \cdot \tfrac{\log(1/\delta) \log n}{n} \Big\},$$

which completes the proof of Lemma 5.3.

D.3.2 Proof of Lemma 5.4

For use in this proof, we note that the population-level covariance operator $\mathbf{\Gamma}_{*,q}$ satisfies the sandwich relation

$$\underline{q} I \preceq \mathbf{\Gamma}_{*,q} \preceq \overline{q} I. \tag{D.12}$$

Our argument adopts the approach used in the paper [132], but involves more refined arguments so as to obtain sharper bounds that allow small value of ρ_n. By multiplying with the operator $(\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1})^{-1/2}$ from both the left and the right of equation (5.57), we find that it suffices to bound the operator norm of the following pre-conditioned error operator:

$$\widehat{\Delta}_n := (\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1})^{-1/2} \big(\widehat{\mathbf{\Gamma}}_{n,q} - \mathbf{\Gamma}_{*,q}\big) (\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1})^{-1/2}.$$

Note that $\widehat{\Delta}_n$ is sum of i.i.d. random operators. In order to bound its operator norm, we invoke a known Bernstein inequality in Hilbert spaces. It applies to an i.i.d. sequence $\{X_i\}_{i=1}^n$ of self-adjoint zero-mean operators on a separable Hilbert space $\mathbb{V}$.

Proposition D.2 (Minsker [144]). *Consider a sequence such that*

$$\|\mathbb{E}[X_i^2]\|_{op} \leq \sigma^2, \quad \text{trace}(\mathbb{E}[X_i^2]) \leq V < \infty, \quad \textit{and} \quad \|X_i\|_{op} \leq U, \ \textit{almost surely.}$$

Then we have the concentration inequality

$$\mathbb{P}\Big(\| \sum_{j=1}^n X_i \|_{op} \geq t \Big) \leq \frac{14V}{\sigma^2} \exp\left(-\frac{t^2/2}{n\sigma^2 + tU/3} \right), \quad \textit{for any } t > 0.$$

A form of this result is stated as as Theorem 3.1 in the paper [144]; see also §3.1 of the same paper for the extension to the infinite-dimensional case.

Using this auxiliary result, let us now prove Lemma 5.4. In doing so, we make use the shorthand notation $\phi^i := \phi(X_i, A_i)$ and $q^i := q(X_i, A_i)$, along with the sequence of random linear operators

$$\Delta_i := (\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1})^{-1/2} \big(q^i \phi^i (\phi^i)^\top - \mathbf{\Gamma}_{*,q}\big) (\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1})^{-1/2} \qquad \text{for each } i \in [n].$$

We need to bound the relevant quantities required to apply Proposition D.2. Beginning with the variance, we have

$$\begin{aligned}
&\mathbb{E}\big[\Delta_i^2\big] && \text{(D.13)}\\
&\preceq \mathbb{E}\Big[\Big\{(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}\big(q^i\phi^i(\phi^i)^\top-\mathbf{\Gamma}_{*,q}\big)(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}\Big\}^2\Big]\\
&= \mathbb{E}\Big[\Big\{q^i(\phi^i)^\top(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1}\phi^i\Big\}\cdot\Big\{(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}q^i\phi^i(\phi^i)^\top(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}\Big\}\Big]. && \text{(D.14)}
\end{aligned}$$

Define the quantity $\Phi_{\max} := \sup_{(x,a)}\big|q(x,a)\phi(x,a)^\top(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1}\phi(x,a)\big|$. We can use this uniform bound to control the right-hand-side of the relation (D.14), and obtain

$$\begin{aligned}
\mathbb{E}\big[\Delta_i^2\big] &\preceq \Phi_{\max}\cdot\mathbb{E}\Big[(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}q^i\phi^i(\phi^i)^\top(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}\Big]\\
&= \Phi_{\max}\cdot(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}\mathbf{\Gamma}_{*,q}(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}
\end{aligned}$$

We can then bound the operator norm and trace of $\mathbb{E}[\Delta_i^2]$ as

$$\big\|\mathbb{E}\big[\Delta_i^2\big]\big\|_{\mathrm{op}} \leq \Phi_{\max}\cdot\big\|(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}\mathbf{\Gamma}_{*,q}(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}\big\|_{\mathrm{op}} \leq \Phi_{\max}, \tag{D.15a}$$

and by equation (D.12), we have

$$\mathrm{trace}\Big(\mathbb{E}\big[\Delta_i^2\big]\Big) \leq \Phi_{\max}\cdot\mathrm{trace}\big((\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1}\mathbf{\Gamma}_{*,q}\big) \leq \overline{q}\Phi_{\max}\cdot\mathrm{trace}\big(\underline{q}I+\rho_n\mathbf{\Lambda}^{-1})^{-1}\big). \tag{D.15b}$$

Finally, we note that

$$\begin{aligned}
\|\Delta_i\|_{\mathrm{op}} &\leq \mathrm{trace}\Big((\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}q^i\phi^i(\phi^i)^\top(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1/2}\Big)\\
&= q^i(\phi^i)^\top(\mathbf{\Gamma}_{*,q}+\rho_n\mathbf{\Lambda}^{-1})^{-1}\phi^i \leq \Phi_{\max} && \text{(D.15c)}
\end{aligned}$$

almost surely.

Combining the different parts of equation (D.15) with Proposition D.2 yields the tail bound

$$\mathbb{P}\Big(\|\widehat{\Delta}_n\|_{\mathrm{op}} \leq t\Big) \leq 14\overline{q}\,\mathrm{trace}\big((\underline{q}I+\rho_n\mathbf{\Lambda}^{-1})^{-1}\big)\cdot\exp\left\{\frac{-nt^2/2}{(1+t/3)\Phi_{\max}}\right\},$$

valid for any $t > 0$.

Noting that $\mathrm{trace}\big((\underline{q}I+\rho_n\mathbf{\Lambda}^{-1})^{-1}\big) \leq \rho_n^{-1}\,\mathrm{trace}(\mathbf{\Lambda}) \leq \frac{\kappa^2}{\rho_n}$, for the event $\mathscr{E}_\delta$ defined as

$$\mathscr{E}_\delta := \left\{\|\widehat{\Delta}_n\|_{\mathrm{op}} \leq \sqrt{\frac{2\Phi_{\max}}{n}\log\big(\frac{\kappa^2}{\rho_n\delta}\big)} + \frac{6\Phi_{\max}}{n}\log\big(\frac{\kappa^2}{\rho_n\delta}\big)\right\},$$

we have $\mathbb{P}(\mathscr{E}_\delta) \geq 1-\delta$.

Note that the quantity $\Phi_{\max}$ admits the bound

$$\Phi_{\max} \leq \bar{q} \sup_{x \in \mathbb{X}, a \in \mathbb{A}} \left| \phi(x,a)^\top (\underline{q} I + \rho_n \mathbf{\Lambda}^{-1})^{-1} \phi(x,a) \right| \leq \bar{q}/\underline{q} \cdot D\big(\rho_n / \underline{q}\big).$$

On the event $\mathscr{E}_\delta$, the conditions (5.56) imply that $\lVert\!| \widehat{\Delta}_n \rVert\!|_{\mathrm{op}} \leq \omega$. Therefore, we conclude that the following bound holds true with probability $1 - \delta$:

$$(1-\omega) I \preceq (\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1})^{-1/2} \big(\widehat{\mathbf{\Gamma}}_{n,q} + \rho_n \mathbf{\Lambda}^{-1}\big) (\mathbf{\Gamma}_{*,q} + \rho_n \mathbf{\Lambda}^{-1})^{-1/2} \preceq (1+\omega) I,$$

which completes the proof of Lemma 5.4.

D.3.3 Proof of Lemma D.1

Recall the error decomposition in the proof of Lemma 5.2:

$$\widehat{\beta}_n - \beta_* = \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1} \cdot \frac{1}{n} \sum_{i=1}^n \Big\{ \varepsilon_i \phi(X_i, A_i) - \rho_n \mathbf{\Lambda}^{-1} \beta_* \Big\},$$

where we define the noise function $\varepsilon_i := Y_i - \mu^*(X_i, A_i)$.

Since Lemma 5.3 does not depend on the condition (5.21a) on the effective dimension, conditionally on the state-action pairs $(X_i, A_i)_{i=1}^n$, with probability $1-\delta$, we have

$$\begin{aligned} &\left| z^\top (\widehat{\beta}_n - \beta_*) \right| \leq c \big\| \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} z \big\|_{\ell^2} \\ &\times \Big\{ \sqrt{\rho_n} \|\mu^*\|_{\mathbb{H}} + \bar{\sigma} \sqrt{\tfrac{\log(1/\delta)}{n}} + \sigma \sup_{x \in \mathcal{S}, a \in \mathbb{A}} \big\| \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x,a) \big\|_{\ell^2} \cdot \tfrac{\log(1/\delta) \log n}{n} \Big\}. \end{aligned} \tag{D.16}$$

On the other hand, note that under Assumption (Kbou(κ)), given the regularization parameter choice (D.5), we have

$$\log\Big(\frac{\kappa^2}{\rho_n \delta}\Big) \frac{D(\rho_n)}{n} \leq \log\Big(\frac{\kappa^2}{\rho_n \delta}\Big) \frac{\kappa^2}{n \rho_n} \leq \frac{1}{32},$$

which verifies the condition (5.56) with $\bar{q} = \underline{q} = 1$ and $\omega = 1/2$. Invoking the empirical covariance concentration lemma 5.4 with $q \equiv 1$ yields

$$\frac{1}{2}(I + \rho_n \mathbf{\Lambda}^{-1}) \preceq \widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1} \preceq 2(I + \rho_n \mathbf{\Lambda}^{-1}), \quad \text{with probability } 1 - \delta.$$

We can therefore control the the relevant terms in equation (D.16), leading to the following inequalities with probability $1-\delta$.

$$\begin{gathered} \big\| \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} z \big\|_{\ell^2} \leq 2 \big\| \big(I + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} z \big\|_{\ell^2}, \quad \text{and,} \\ \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \big\| \big(\widehat{\mathbf{\Gamma}}_n + \rho_n \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x,a) \big\|_{\ell^2} \leq \rho_n^{-1/2} \cdot \sup_{(x,a) \in \mathcal{S} \times \mathbb{A}} \| \mathbf{\Lambda}^{1/2} \phi(x,a) \|_{\ell^2} \leq \kappa / \sqrt{\rho_n}, \end{gathered}$$

where the last step follows from the uniform upper bound (Kbou(κ)).

Substituting these results back into equation (D.16), and taking the regularization parameter according to equation (D.5), we conclude that

$$\begin{aligned}\Big|z^\top(\widehat{\beta}_n-\beta_*)\Big| &\leq 4c\big\|\big(I+\rho_n\mathbf{\Lambda}^{-1}\big)^{-1/2}z\big\|_{\ell^2}\Big\{\sqrt{\rho_n}\|\mu^*\|_{\mathbb{H}}+\bar{\sigma}\sqrt{\tfrac{\log(1/\delta)}{n}}+\frac{\sigma\kappa\log(1/\delta)\log n}{n\sqrt{\rho_n}}\Big\}\\ &\leq c'\big\|\big(I+\rho_n\mathbf{\Lambda}^{-1}\big)^{-1/2}z\big\|_{\ell^2}\cdot(\sigma+\kappa R)\sqrt{\frac{\log(n/\delta)\log(1/\delta)}{n}},\end{aligned}$$

with probability $1-\delta$, which proves Lemma D.1.

D.3.4 Proof of Lemma 5.6

As with the proof of Lemma 5.3, we decompose the error into a noise and bias term—namely

$$\langle z,\,\widehat{\beta}_n-\beta_*\rangle=\underbrace{\frac{1}{n}\sum_{i=1}^n W_i}_{\text{noise part}}-\underbrace{\rho_n^{\text{(III)}}z^\top\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}\mathbf{\Lambda}^{-1}\beta_*}_{\text{bias part}},$$

where $W_i:=z^\top\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}\varepsilon_i\widehat{\sigma}_n^{-2}\phi(X_i,A_i)$

For the bias part, applying the Cauchy–Schwarz inequality yields

$$\begin{aligned}\rho_n^{\text{(III)}}\Big|z^\top\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}\mathbf{\Lambda}^{-1}\beta_*\Big| &\leq \rho_n^{\text{(III)}}\|\mathbf{\Lambda}^{-1/2}\beta_*\|_{\ell^2}\cdot\|\mathbf{\Lambda}^{-1/2}\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}z\|_{\ell^2}\\ &\overset{(i)}{\leq}\sqrt{\rho_n^{\text{(III)}}}\|\mu^*\|_{\mathbb{H}}\cdot\|\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1/2}z\|_{\ell^2}\\ &\overset{(ii)}{\leq}\frac{1}{\sqrt{n}}\|\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1/2}z\|_{\ell^2},\end{aligned}\tag{D.17}$$

where in step (i), we use the fact $\|\mathbf{\Lambda}^{-1/2}\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1/2}\|_{\text{op}}\leq(\rho_n^{\text{(III)}})^{-1/2}$, and in step (ii), we substitute with the regularization parameter choice $\rho_n^{\text{(III)}}=\frac{1}{Rn}$.

For the noise part, we use Adamczak's concentration inequality to establish high-probability bounds. We start with the expression for the conditional variance

$$\mathbb{E}\Big[W_i^2\mid X_i,A_i,\widehat{\sigma}_n\Big]=\frac{\sigma^2(X_i,A_i)}{\widehat{\sigma}_n^4(X_i,A_i)}\Big(z^\top\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}\phi(X_i,A_i)\Big)^2,$$

which leads to the bound

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\mathbb{E}\big[W_i^2\mid X_i,A_i,\widehat{\sigma}_n\big] &\leq \max_{i\in[n]}\frac{\sigma^4(X_i,A_i)}{\widehat{\sigma}_n^4(X_i,A_i)}\cdot z^\top\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}\widehat{\mathbf{\Gamma}}_n^\sigma\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}z\\ &\leq 4z^\top\big(\widehat{\mathbf{\Gamma}}_n^\sigma+\rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}z.\end{aligned}\tag{D.18}$$

In the last step, we use the fact $\widehat{\sigma}_n^2(X_i, A_i) \geq \frac{1}{2}\sigma^2(X_i, A_i)$ for any $i \in [n]$.

On the other hand, for any $p > 0$, we have the conditional moment bound

$$\begin{aligned}
&\Big\{\mathbb{E}\Big[|W_i|^p \mid X_i, A_i, \widehat{\sigma}_n\Big]\Big\}^{1/p} \\
&\leq \sqrt{p}\sigma \cdot \Big|z^\top \big(\widehat{\mathbf{\Gamma}}_n^\sigma + \rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1}\widehat{\sigma}_n^{-2}\phi(X_i, A_i)\Big| \\
&\leq \frac{2\sqrt{p}\sigma}{\underline{\sigma}^2}\|\big(\widehat{\mathbf{\Gamma}}_n^\sigma + \rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1/2}z\|_{\ell^2} \cdot \sup_{(x,a)}\|\big(\widehat{\mathbf{\Gamma}}_n^\sigma + \rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(x,a)\|_{\ell^2}. \qquad \text{(D.19)}
\end{aligned}$$

Combining equations (D.18) and (D.19) with Adamczak's inequality, we conclude that

$$\begin{aligned}
|\frac{1}{n}\sum_{i=1}^n W_i| \leq c\|\big(\widehat{\mathbf{\Gamma}}_n^\sigma + \rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1/2}z\|_{\ell^2}\Bigg\{2\sqrt{\frac{\log(1/\delta)}{n}} \\
+ \frac{\log n \log(1/\delta)\sigma}{n\underline{\sigma}^2}\sup_{(x,a)}\|\big(\widehat{\mathbf{\Gamma}}_n^\sigma + \rho_n^{\text{(III)}}\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(x,a)\|_{\ell^2}\Bigg\}. \qquad \text{(D.20)}
\end{aligned}$$

Finally, putting together equations (D.17) and (D.20) completes the proof of this lemma.

D.4 Conditional variance estimation

In this section, we discuss the problem of estimating the conditional variance function $(x, a) \mapsto \sigma^2(x, a)$.

D.4.1 Some conditional variance estimators

In this section, we construct concrete estimators for the conditional variance σ^2 that satisfy the robust pointwise risk property. In combination with the four-stage framework (5.25), these results immediately lead to instance-optimal results in Theorem 5.3.

Kernel ridge regression: Consider a positive semi-definite kernel function $\mathcal{K}_\sigma : (\mathbb{X} \times \mathbb{A}) \times (\mathbb{X} \times \mathbb{A}) \to \mathbb{R}$ that defines an RKHS $\mathbb{H}_\sigma$ with the Mercer decomposition

$$\mathcal{K}_\sigma\big((s_1, a_1), (s_2, a_2)\big) = \sum_{j=1}^\infty \lambda_j \phi_j(s_1, a_1)\phi_j(s_2, a_2). \qquad \text{(D.21)}$$

We assume that the RKHS $\mathbb{H}_\sigma$ satisfies the regularity assumption (Kbou(κ)), and that the true conditional variance function lies in this RKHS, i.e.,

$$\|\sigma^2\|_{\mathbb{H}_\sigma} \leq R^\sigma. \qquad \text{(D.22)}$$

Following the definition (5.19), for any $\rho > 0$ we define $D_\sigma(\rho)$ as the effective dimension associated to the regularization parameter $\rho > 0$ for the RKHS $\mathbb{H}_\sigma$.

We consider the penalized least-square estimator

$$\widehat{\sigma}_n^2 := \arg\min_h \Big\{ \frac{1}{n} \sum_{i=1}^n \big(Z_i - h(X_i, A_i) \big)^2 + \rho \|h\|_{\mathbb{H}_\sigma}^2 \Big\}. \tag{D.23}$$

Proposition D.3. *Let $\rho_0(\varepsilon)$ be the smallest value of ρ such that $\rho D_\sigma(\rho) \leq \big(\varepsilon / R^\sigma\big)^2$, the estimator (D.23) with parameter choice $\rho_n = \rho_0(\varepsilon)$ satisfies the robust pointwise risk property with*

$$m(\varepsilon, \delta) := c \frac{\sigma^4 \log(1/\delta)}{\varepsilon^2} D_\sigma(\rho_0(\varepsilon)) + c \frac{(R^\sigma)^2}{\sigma^4 \rho_0(\varepsilon)}, \quad \textit{and} \quad \bar{b}(\varepsilon, \delta) := \frac{\varepsilon}{c \sqrt{D_\sigma(\rho_0(\varepsilon))}}. \tag{D.24}$$

See Section D.4.2.1 for the proof.

A few remarks are in order. First, Proposition D.3 requires that the effective dimension of the RKHS $\mathbb{H}_\sigma$ to satisfy that $\rho D_\sigma(\rho) \to 0$ for $\rho \to 0^+$. A similar condition is also imposed on the RKHS $\mathbb{H}$ used to estimate the treatment effect function, which can be verified under certain conditions on the eigenfunctions. (see equation (D.2) and Proposition D.1 in the appendix for the statement of such results.) In particular, suppose that the effective dimension satisfies a decay condition $D(\rho) \leq D_0 \rho^{\omega - 1}$ for some scalar $\omega \in (0, 1]$, by seeing the scalars (D_0, R^σ, σ) as constants, we choose $\rho_0(\varepsilon) = \varepsilon^{\frac{2}{\omega}}$ the condition (D.24) becomes

$$m(\varepsilon, \delta) \asymp \varepsilon^{-2/\omega} \log(1/\delta) \quad \text{and} \quad \bar{b}(\varepsilon, \delta) \asymp \varepsilon^{1/\omega},$$

Such a requirement on the sample size m and the bias upper bound $\bar{b}$ may not always achieve the optimal rate for estimating the function σ^2. However, since we only need the estimation error to be smaller than a constant $\underline{\sigma}^2/2$, as required in equation (5.28b), a polynomial dependency on the accuracy level ε and poly-logarithmic dependency on the failure probability δ suffices our purposes.

Local average estimator: Let the statespace $\mathbb{X}$ be a compact subset of $\mathbb{R}^d$ and let the action space $\mathbb{A}$ be discrete. Define the class of L-Lipschitz functions as

$$\mathcal{F}_L = \Big\{ f : \mathbb{X} \to \mathbb{R}, |f(x) - f(y)| \leq L \|x - y\|_2 \text{ for any } x, y \in \mathbb{X} \Big\}$$

We assume that the conditional variances are smooth enough.

$$\sigma^2(\cdot, a) \in \mathcal{F}_L, \quad \text{for each } a \in \mathbb{A}. \tag{D.25}$$

To make estimation possible with random design, we need an additional regularity assumption on the density.

$$\inf_{x \in \mathbb{X}, a \in \mathbb{A}} \big(\xi^* \cdot \pi \big) \Big(\mathbb{B}(x, r) \times \{a\} \Big) \geq p_0 r^{d_0}, \quad \text{for any } r \in (0, r_0). \tag{D.26}$$

Given a tuning parameter $r_n > 0$, we consider the local averaging estimator

$$\widehat{\sigma}_n^2(x_0, a_0) := |\widehat{S}_{x_0,a_0}|^{-1} \sum_{i \in \widehat{S}_{x_0,a_0}} Z_i, \quad \text{where } \widehat{S}_{x_0,a_0} := \Big\{ i \in [n] : X_i \in \mathbb{B}(x_0, r_n), A_i = a_0 \Big\}. \tag{D.27}$$

Proposition D.4. *For any $\varepsilon > 0$ and $\delta \in (0,1)$, there exists a universal constant $c > 0$, such that the estimator (D.27) satisfies the robust pointwise risk property with*

$$m(\varepsilon, \delta) = \frac{L^{2d_0} \log(1/\delta)}{p_0} \Big(\frac{c}{\varepsilon}\Big)^{d_0+2} + \frac{\log(1/\delta)}{\mathbb{L}^2 p_0 r_0^{d_0+1}} L^{d_0} \frac{\log(1/\delta)}{p_0} \log^{d_0+2}(1/\varepsilon), \quad \textit{and}$$
$$\bar{b}(\varepsilon, \delta) = \varepsilon/2.$$

See Section D.4.2.2 for the proof.

A few remarks are in order. Compared to Proposition D.3, the local average estimator only requires the target function $\sigma^2(\cdot, a)$ to be Lipschitz, for any $a \in \mathbb{A}$. In dimension larger than 1, this usually requires less order of smoothness than the RKHS case in Proposition D.3, while being less flexible with the structure of the function class. The regularity condition (D.26) ensures that any small ball in $\mathbb{X}$ and any action a get sufficiently large probability of being sampled. For example, when the probability distribution ξ has a density function uniformly bounded by $\xi_{\min} > 0$, and when the probability of choosing any action a is at least $\pi_{\min}$, the condition (D.26) is satisfied with $p_0 = c_d \xi_{\min} \pi_{\min}$ and $d_0 = d$, for a constant $c_d > 0$ depending only on d. More generally, even if the function $(x, a) \mapsto \xi^*(x)\pi(x, a)$ can attain 0 at some points, as long as appropriate growth conditions are imposed around these points, the condition (D.26) will still be satisfied. Finally, though we only study the Lipschitz case, in literature optimal results for general Hölder classes have been established for the estimation problems of conditional variance [188]. In combination with their results, we can also obtain optimal instance-dependent guarantees in Theorem 5.3.

D.4.2 Proofs of robust pointwise risk properties

In this appendix, we establish the robust pointwise risk properties for various estimators discussed in Appendix D.4.

D.4.2.1 Proof of Proposition D.3

The proof is similar to that of Lemma 5.2, with specific treatment given to the deterministic bias part. Define the infinite-dimensional vectors

$$\beta_* := \Psi(\sigma^2) \quad \text{and} \quad \widehat{\beta}_n := \Psi(\widehat{\sigma}_n^2).$$

We can represent the error using basis functions.

$$\sigma^2(s_0, a_0) - \widehat{\sigma}_n^2(s_0, a_0) = \langle \widehat{\beta}_n - \beta_*, \, \phi(s_0, a_0) \rangle.$$

Defining the noise and bias parts

$$\varepsilon_i := Z_i - \mathbb{E}[Z_i \mid X_i, A_i], \quad \text{and} \quad b(X_i, A_i) := \mathbb{E}[Z_i \mid X_i, A_i] - \sigma^2(X_i, A_i).$$

We also define the empirical covariance operator

$$\widehat{\mathbf{\Gamma}}_n := \frac{1}{n}\sum_{i=1}^n \phi(X_i, A_i)\phi(X_i, A_i)^\top,$$

the error vector $\widehat{\beta}_n - \beta_*$ admits a representation

$$\widehat{\beta}_n - \beta_* = \big(\widehat{\mathbf{\Gamma}}_n + \rho\mathbf{\Lambda}^{-1}\big)^{-1}\frac{1}{n}\sum_{i=1}^n \Big\{\varepsilon_i(X_i, A_i)\phi(X_i, A_i) + b(X_i, A_i)\phi(X_i, A_i) - \rho\mathbf{\Lambda}^{-1}\beta_*\Big\}. \tag{D.28}$$

Define the event

$$\mathscr{E}_{\varepsilon,\delta} := \Big\{\max_{1 \leq i \leq n} |b(X_i, A_i)| \leq \bar{b}(\varepsilon, \delta)\Big\}.$$

Clearly, the error consists of three parts: a part induced by stochastic (unbiased) noise ε_i; a part involving the observation bias $b(X_i, A_i)$; and the bias introduced by the regularization ρ. We claim that the following bounds hold true with probability $1 - \delta$ on the event $\mathscr{E}_{\varepsilon,\delta}$.

$$\left|\phi(s_0, a_0)^\top\big(\widehat{\mathbf{\Gamma}}_n + \rho\mathbf{\Lambda}^{-1}\big)^{-1}\frac{1}{n}\sum_{i=1}^n \varepsilon_i\phi(X_i, A_i)\right| \leq c\big(\sigma^2 + \bar{b}(\varepsilon, \delta)\big)\sqrt{\frac{D_\sigma(\rho)\log(1/\delta)}{n}}. \tag{D.29a}$$

$$\left|\phi(s_0, a_0)^\top\big(\widehat{\mathbf{\Gamma}}_n + \rho\mathbf{\Lambda}^{-1}\big)^{-1}\frac{1}{n}\sum_{i=1}^n b(X_i, A_i)\phi(X_i, A_i)\right| \leq c\bar{b}(\varepsilon, \delta)\sqrt{D_\sigma(\rho)}, \tag{D.29b}$$

$$\left|\phi(s_0, a_0)^\top\big(\widehat{\mathbf{\Gamma}}_n + \rho\mathbf{\Lambda}^{-1}\big)^{-1}\rho\mathbf{\Lambda}^{-1}\beta_*\right| \leq cR^\sigma\sqrt{\rho D_\sigma(\rho)}. \tag{D.29c}$$

Taking these three bounds as given, for any $\varepsilon > 0$ and $\delta \in (0, 1)$, we take $\rho_0(\varepsilon)$ be the smallest value of ρ such that $\rho D_\sigma(\rho) \leq \big(\frac{\varepsilon}{R^\sigma}\big)^2$ (which is guaranteed to exist for if $n \cdot D_\sigma(1/n) \to 0$), the robust pointwise risk condition is satisfied with

$$m(\varepsilon, \delta) = c\frac{\sigma^4\log(1/\delta)}{\varepsilon^2}D_\sigma(\rho_0(\varepsilon)) + c\frac{(R^\sigma)^2}{\sigma^4\rho_0(\varepsilon)}, \quad \text{and} \quad \bar{b} = \frac{\varepsilon}{c\sqrt{D_\sigma(\rho_0(\varepsilon))}},$$

completing the proof of Proposition D.3.

The rest of this section is devoted to the proofs of equations (D.29a)– (D.29c).

Proof of equation (D.29a): By definition, note that the noise satisfies the conditional ψ_1-norm bound

$$\|\varepsilon_i \mid X_i, A_i\|_{\psi_1} \leq 4(\sigma^2 + |b(X_i, A_i)|).$$

Invoking Adamczak's concentration inequality, conditionally on $(X_i, A_i)_{i=1}^n$, with probability $1 - \delta$, we have

$$\begin{aligned}
&\left|\phi(s_0, a_0)^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1} \frac{1}{n} \sum_{i=1}^n \varepsilon_i(X_i, A_i)\phi(X_i, A_i)\right| \\
&\qquad \leq c\big(\sigma^2 + \max_{i \in [n]} |b(X_i, A_i)|\big) \times \Bigg\{ \|\widehat{\mathbf{\Gamma}}_n^{1/2} \big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1} \phi(s_0, a_0)\|_{\ell^2} \sqrt{\frac{\log(1/\delta)}{n}} \\
&\qquad\qquad + \sup_{x', a'} \left|\phi(x', a')^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1} \phi(s_0, a_0)\right| \frac{\log n \log(1/\delta)}{n} \Bigg\}.
\end{aligned}$$

On the event $\mathscr{E}_{\varepsilon,\delta}$, we have $\max_{i \in [n]} |b(X_i, A_i)| \leq \bar{b}(\varepsilon, \delta)$. By Lemma 5.4, with probability $1 - \delta$, we have

$$\sup_{x,a} \|\big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x, a)\|_{\ell^2} \leq 2 \sup_{x,a} \|\big(I + \rho \mathbf{\Lambda}^{-1}\big)^{-1/2} \phi(x, a)\|_{\ell^2} \leq c\sqrt{D_\sigma(\rho)}.$$

Putting together the pieces completes the proof of equation (D.29a).

Proof of equation (D.29b): Applying the Cauchy–Schwarz inequality to the finite summation yields

$$\begin{aligned}
&\left|\phi(s_0, a_0)^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1} \frac{1}{n} \sum_{i=1}^n b(X_i, A_i)\phi(X_i, A_i)\right|^2 \\
&\leq \frac{1}{n} \sum_{i=1}^n \left|\phi(s_0, a_0)^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1} b(X_i, A_i) \phi(X_i, A_i)\right|^2 \\
&= \max_{i \in [n]} b^2(X_i, A_i) \\
&\qquad \times \frac{1}{n} \sum_{i=1}^n \phi(s_0, a_0)^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1} \phi(X_i, A_i) \phi(X_i, A_i)^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1} \phi(s_0, a_0) \\
&\leq \max_{i \in [n]} b^2(X_i, A_i) \cdot \phi(s_0, a_0)^\top \big(\widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1}\big)^{-1} \phi(s_0, a_0)
\end{aligned}$$

On the other hand, we can apply Lemma 5.4 to the empirical covariance operator $\widehat{\mathbf{\Gamma}}_n$ in the RKHS $\mathbb{H}_\sigma$. Given the regularization parameter $\rho \geq c\kappa^2 \frac{\log(n/\delta)}{n}$, with probability $1 - \delta$, we have

$$\frac{1}{2}(I + \rho \mathbf{\Lambda}^{-1}) \preceq \widehat{\mathbf{\Gamma}}_n + \rho \mathbf{\Lambda}^{-1} \preceq 2(I + \rho \mathbf{\Lambda}^{-1}).$$

Consequently, on the event $\mathscr{E}_{\varepsilon,\delta}$, we have the upper bound

$$\begin{aligned}
&\left|\phi(s_0,a_0)^\top\big(\widehat{\mathbf{\Gamma}}_n+\rho\mathbf{\Lambda}^{-1}\big)^{-1}\frac{1}{n}\sum_{i=1}^n b(X_i,A_i)\phi(X_i,A_i)\right|\\
&\leq c\bar{b}(\varepsilon,\delta)\cdot\sup_{x,a}\|\big(\widehat{\mathbf{\Gamma}}_n+\rho\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(x,a)\|_{\ell^2}^2\\
&\leq 2c\bar{b}(\varepsilon,\delta)\cdot\sup_{x,a}\|\big(I+\rho\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(x,a)\|_{\ell^2}^2\\
&=2c\bar{b}(\varepsilon,\delta)D_\sigma(\rho).
\end{aligned}$$

with probability $1-\delta$.

Proof of equation (D.29c): By Cauchy–Schwarz inequality, we note that

$$\begin{aligned}
&\left|\phi(s_0,a_0)^\top\big(\widehat{\mathbf{\Gamma}}_n+\rho\mathbf{\Lambda}^{-1}\big)^{-1}\rho\mathbf{\Lambda}^{-1}\beta_*\right|\\
&\leq\sqrt{\rho}\cdot\|\big(\widehat{\mathbf{\Gamma}}_n+\rho\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(s_0,a_0)\|_{\ell^2}\cdot\vert\!\vert\!\vert\big(\widehat{\mathbf{\Gamma}}_n+\rho\mathbf{\Lambda}^{-1}\big)^{-1/2}(\rho\mathbf{\Lambda}^{-1})^{1/2}\vert\!\vert\!\vert_{\mathrm{op}}\cdot\|\mathbf{\Lambda}^{-1/2}\beta_*\|_{\ell^2}\\
&\leq 2\sqrt{\rho}R^\sigma\|\big(\widehat{\mathbf{\Gamma}}_n+\rho\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(s_0,a_0)\|_{\ell^2}.
\end{aligned}$$

By Lemma 5.4, with probability $1-\delta$, we have

$$\|\big(\widehat{\mathbf{\Gamma}}_n+\rho\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(s_0,a_0)\|_{\ell^2}\leq 2\|\big(I+\rho\mathbf{\Lambda}^{-1}\big)^{-1/2}\phi(s_0,a_0)\|_{\ell^2}\leq 2\sqrt{D_\sigma(\rho)},$$

which proves equation (D.29c).

D.4.2.2 Proof of Proposition D.4

We start with a decomposition of the error

$$\widehat{\sigma}_n^2(x_0,a_0)-\sigma^2(x_0,a_0)=\left|\widehat{S}_{x_0,a_0}\right|^{-1}\sum_{i\in\widehat{S}_{x_0,a_0}}\Big\{\varepsilon_i+b(X_i,A_i)+\big(\sigma^2(X_i,A_i)-\sigma^2(x_0,a_0)\big)\Big\},\tag{D.30}$$

where the noise ε_i is defined as $\varepsilon_i:=Z_i-\mathbb{E}[Z_i|X_i,A_i]$ for each $i\in[n]$.

Recall from the definition of the set $\widehat{S}_{x_0,a_0}$ that for each $i\in\widehat{S}_{x_0,a_0}$, we have $A_i=a_0$ and $X_i\in\mathbb{B}(x_0,r_n)$. Applying the Lipschitz condition D.25 then leads to the bound

$$\left|\widehat{S}_{x_0,a_0}\right|^{-1}\sum_{i\in\widehat{S}_{x_0,a_0}}\left|\sigma^2(X_i,A_i)-\sigma^2(x_0,a_0)\right|\leq Lr_n.\tag{D.31}$$

Defining the event

$$\mathscr{E}_{\varepsilon,\delta}:=\Big\{\max_{i\in[n]}|b(X_i,A_i)|\leq\bar{b}(\varepsilon,\delta)\Big\},$$

on this event, we can control the additional bias in the observations

$$\left|\widehat{S}_{x_0,a_0}\right|^{-1} \sum_{i \in \widehat{S}_{x_0,a_0}} |b(X_i, A_i)| \leq \bar{b}(\varepsilon, \delta). \tag{D.32}$$

For the stochastic noise, we claim the following bound holds true whenever the tuning parameter r_n satisfies $r_n \leq r_0$ and $\frac{p_0 n r_n^{d_0}}{\log^2 n} \geq \log(1/\delta)$.

$$\Big| \sum_{i \in \widehat{S}_{x_0,a_0}} \varepsilon_i \Big| \leq c \sqrt{\frac{\log(1/\delta)}{n p_0 r_n^{d_0}}}, \quad \text{with probability } 1-\delta, \text{ on the event } \mathscr{E}_{\varepsilon,\delta} \tag{D.33}$$

We prove this inequality at the end of this section.

Combining equations (D.31), (D.32), and (D.33), we choose the local radius as

$$r_n := \left\{\frac{\log(1/\delta)}{\mathbb{L}^2 p_0 n}\right\}^{\frac{1}{d_0+2}}.$$

Whenever the sample size n satisfies

$$n \geq \frac{\log(1/\delta)}{\mathbb{L}^2 p_0 r_0^{d_0+1}}, \quad \text{and} \quad \frac{n}{\log^{d_0+2} n} \geq L^{d_0} \frac{\log(1/\delta)}{p_0},$$

on the event $\mathscr{E}_{\varepsilon,\delta}$, we have the upper bound with probability $1-\delta$,

$$\left|\widehat{\sigma}_n^2(x_0, a_0) - \sigma^2(x_0, a_0)\right| \leq \bar{b}(\varepsilon, \delta) + c \cdot L^{\frac{2d_0}{d_0+2}} \left\{\frac{\log(1/\delta)}{p_0 n}\right\}^{\frac{1}{d_0+2}}.$$

Proof of equation (D.33): We start by exhibiting a lower bound on the cardinality of the set $\widehat{S}_{x_0,a_0}$. When the averaging radius r_n satisfies $r_n \leq r_0$, by the density condition (D.26), we have

$$p_* := \mathbb{P}\left(i \in \widehat{S}_{x_0,a_0}\right) \geq p_0 r_n^{d_0}, \quad \text{for each } i \in [n].$$

The indicators $\mathbf{1}_{i \in \widehat{S}_{x_0,a_0}}$ are independent for each $i \in [n]$. By Chernoff bound in the entropy form, we have

$$\mathbb{P}\left(|\widehat{S}_{x_0,a_0}| \leq \frac{n p_*}{2}\right) \leq \exp\Big\{-n D_{\mathrm{KL}}\left(p_*/2 \parallel p_*\right)\Big\} \leq \exp\left(-c p_* n\right),$$

for a universal constant $c > 0$.

Defining the event

$$\mathscr{E}' := \left\{|\widehat{S}_{x_0,a_0}| \geq \frac{n p_*}{2}\right\},$$

the concentration inequality above implies that $\mathbb{P}(\mathscr{E}') \geq 1-\delta$ whenever $np_* \geq c' \log(1/\delta)$.

Let us condition on the state-action pairs $(X_i, A_i)_{i=1}^n$ such that the event $\mathscr{E}_{\varepsilon,\delta} \cap \mathscr{E}'$ holds true, applying Adamczak's concentration inequality to its summation, with probability $1-\delta$ under the conditional law, we have

$$\Big| \sum_{i \in \widehat{S}_{x_0, a_0}} \varepsilon_i \Big| \leq c\big(\sigma^2 + \overline{b}(\varepsilon, \delta)\big) \cdot \Big\{ \sqrt{\frac{\log(1/\delta)}{|\widehat{S}_{x_0,a_0}|}} + \frac{\log(1/\delta) \log n}{|\widehat{S}_{x_0,a_0}|} \Big\},$$

for a universal constant $c > 0$.

Taking into account the random design points $(X_i, A_i)_{i=1}^n$, as long as the sample size and the radius satisfies

$$np_* \geq np_0 r_n^{d_0} \geq \log(1/\delta) \cdot \log^2 n,$$

with probability $1-\delta$, we have

$$\Big| \sum_{i \in \widehat{S}_{x_0, a_0}} \varepsilon_i \Big| \leq c \sqrt{\frac{\log(1/\delta)}{np_0 r_n^{d_0}}},$$

which proves equation (D.33).

D.5 Proofs for the examples

We collect the proofs for the examples in this section.

D.5.1 Proof of Corollary 5.1

We first establish the effective dimension condition (5.23) by verifying the sup-norm growth bound (Eig(ν)). Doing so ensures that the optimal risk is determined (up to universal constant factors) by the risk functionals $V^2_{\xi^*}(\mu^*) + V^2_{\sigma,n}(\xi^*, \pi, g; \mathcal{F})$ and $V^2_{\sigma,n}(\delta_{x_0}, \pi, g; \mathcal{F})$. We then use Theorems 5.1 and 5.3 to prove the bounds (5.34a) and (5.34b), respectively.

D.5.1.1 Establishing the effective dimension condition

We start by establishing tight bounds on the sup-norm growth condition (Eig(ν)), which comes with a non-trivial (yet well-controlled) exponent ν. This result is of independent interest, illustrating the growth of eigenfunctions as a natural phenomenon for RKHS applied to data whose densities have singularities.

Lemma D.2. *Under the set-up above, there exists a pair of positive constants c_1, c_2 that depends only on α, such that for each $j \geq 1$, the eigenfunctions ϕ_j (normalized with $\|\phi_j\|_{\mathbb{L}^2(\xi)} = 1$) associated to eigenvalue λ_j satisfy*

$$c_1 \lambda_j^{-\frac{\alpha}{4(\alpha+2)}} \leq \|\phi_j\|_\infty \leq c_2 \lambda_j^{-\frac{\alpha}{4(\alpha+2)}}.$$

See Section D.5.1.5 for the proof of this lemma.

Consequently, the condition (Eig(ν)) is satisfied with exponent $\nu = \frac{\alpha}{4(\alpha+2)}$ and constant $\phi_{\max}$ depending only on α. With eigenvalue decay $\lambda_j \asymp j^{-2}$ of the first-order Sobolev space (see [218]), Proposition D.1 yields

$$D(\rho) \asymp \rho^{-\frac{\alpha+1}{\alpha+2}}, \quad \text{and} \quad D(\rho_n)/n \asymp n^{-\frac{1}{\alpha+2}}.$$

which ensures the regularity condition (5.21a) for sample size n larger than a threshold depending only on α.

D.5.1.2 Proof of equation (5.34a)

Note that for any function $f \in \mathbb{B}_{\mathbb{H}}(1)$ and $x \in [0,1]$, the Cauchy–Schwarz inequality yields

$$|f(x)| = \left| \int_0^x f'(t)dt \right| \leq \sqrt{x \cdot \int_0^x (f'(t))^2 dt} \leq \sqrt{x}.$$

So we have $\|\mu^*\|_\infty \leq 1$ and consequently $V_{\xi^*}(\mu^*) \leq 1$.

By the definition (5.1), the variance function $V_{\sigma,n}(\xi^*, \pi, g; \mathcal{F})$ is defined (up to universal constant factors) as the optimum value of the following variational problem:

$$\sup_q \left\{ \int_0^1 q(x)dx \right\}, \quad \text{such that} \tag{D.34a}$$

$$q(0) = 0, \quad \int_0^1 (q'(x))^2 dx \leq n, \quad \text{and} \quad \int_0^1 (1-x)^\alpha q^2(x)dx \leq 1. \tag{D.34b}$$

It suffices to establish upper and lower bounds on the variance functional $V_{\sigma,n}(\xi^*, \pi, g; \mathcal{F})$ under different regimes.

Our proof relies on a technical lemma regarding the constraint (D.34b), stated as

Lemma D.3. *Under the constraint* (D.34b)*, we have*

$$\sup_{x \in [0,1]} |q(x)| \leq c_\alpha n^{\frac{1+\alpha}{2(2+\alpha)}}, \quad \textit{and} \quad \sup_{x \in [0,1-\varepsilon]} |q(x)| \leq c_{\alpha,\varepsilon} n^{1/4},$$

for a constant c_α depending on $\alpha > 0$, and a constant $c_{\alpha,\varepsilon}$ depending on $\alpha > 0$ and $\varepsilon \in (0,1)$.

We prove this lemma in Section D.5.1.4.

Taking it as given, we now proceed the proof of equation (5.34a). It suffices to establish upper and lower bounds on the variance functional $V_{\sigma,n}(\xi^*, \pi, g; \mathcal{F})$ under different regimes.

Upper bounds on the variance functional: Given a function q satisfying the constraint (D.34b), for any $\varepsilon \in (0,1)$, we decompose the integral $\int_0^1 q(x)dx$ into parts $\int_0^{1-\varepsilon}$ and $\int_{1-\varepsilon}^1$, and bound them in different ways.

By the Cauchy–Schwarz inequality, we note that

$$\Big(\int_0^{1-\varepsilon} q(x)dx \Big)^2 \leq \int_0^{1-\varepsilon} (1-x)^\alpha q^2(x)dx \cdot \int_0^{1-\varepsilon} \frac{dx}{(1-x)^\alpha} \leq \begin{cases} \frac{1}{1-\alpha} & \alpha < 1, \\ \log(1/\varepsilon) & \alpha = 1, \\ \frac{1}{\alpha-1}\varepsilon^{1-\alpha} & \alpha > 1. \end{cases} \tag{D.35}$$

For the second part, integration-by-parts yields

$$\int_{1-\varepsilon}^1 q(x)dx = q(1) - (1-\varepsilon)q(1-\varepsilon) - \int_{1-\varepsilon}^1 xq'(x)dx = \varepsilon q(1-\varepsilon) + \int_{1-\varepsilon}^1 (1-x)q'(x)dx.$$

For the integral term, applying the Cauchy–Schwarz inequality yields

$$|\int_{1-\varepsilon}^1 (1-x)q'(x)dx| \leq \|q\|_{\mathbb{H}} \cdot \sqrt{\int_{1-\varepsilon}^1 (1-x)^2 dx} \leq \sqrt{\varepsilon^3 n}. \tag{D.36}$$

By Lemma D.3, we have

$$|\varepsilon q(1-\varepsilon)| \leq \varepsilon n^{\frac{1+\alpha}{2(2+\alpha)}}. \tag{D.37}$$

Combining equations (D.35), (D.36), (D.37) yields

$$\left| \int_0^1 q(x)dx \right| \leq \sqrt{\varepsilon^3 n} + c_\alpha \varepsilon n^{\frac{1+\alpha}{2(2+\alpha)}} + c'_\alpha \times \begin{cases} 1 & \alpha < 1, \\ \sqrt{\log(1/\varepsilon)} & \alpha = 1, \\ \varepsilon^{-(\alpha-1)/2} & \alpha > 1. \end{cases}$$

We consider three cases:

- When $\alpha < 1$, we take $\varepsilon = 0$, and obtain that $V_{\sigma,n}(\xi^*, \pi, g; \mathcal{F}) \lesssim_\alpha 1$.
- When $\alpha = 1$, we take $\varepsilon = n^{-1}$, and obtain that $V_{\sigma,n}(\xi^*, \pi, g; \mathcal{F}) \lesssim \sqrt{\log n}$.
- When $\alpha > 1$, we take $\varepsilon = n^{\frac{-1}{\alpha+2}}$, and obtain that $V_{\sigma,n}(\xi^*, \pi, g; \mathcal{F}) \lesssim_\alpha n^{\frac{\alpha-1}{2(\alpha+2)}}$.

Lower bounds on the variance functional: On the lower bound side, for positive scalars $\varepsilon \in (0,1)$ and $h > 0$, we construct the function

$$q_{\varepsilon,h}(x) := \begin{cases} 0 & x \leq 1-\varepsilon, \\ \frac{h}{\varepsilon}(x-1+\varepsilon) & x > 1-\varepsilon. \end{cases}$$

Clearly, we have $q_{\varepsilon,h}(0) = 0$, and straightforward calculation yields

$$\int_0^1 q_{\varepsilon,h}(x)dx = \frac{\varepsilon h}{2}, \quad \int_0^1 (q'_{\varepsilon,h}(x))^2 dx = \frac{h^2}{\varepsilon}, \quad \text{and} \quad \int_0^1 (1-x)^\alpha q^2_{\varepsilon,h}(x)dx \le \varepsilon^{\alpha+1}h^2.$$

For $\alpha > 1$, under the choice of parameters

$$\varepsilon = n^{\frac{-1}{2+\alpha}}, \quad \text{and} \quad h = n^{\frac{1+\alpha}{2(2+\alpha)}},$$

we have that $\int_0^1 q_{\varepsilon,h}(x)dx \asymp n^{\frac{\alpha-1}{2(2+\alpha)}}$.

For $\alpha < 1$, taking $\varepsilon = 1$ and $h = 1$, we have that $\int_0^1 q_{\varepsilon,h}(x)dx \asymp 1$.

Consequently, for $\alpha \neq 1$, we have the lower bounds

$$V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) \gtrsim_\alpha \begin{cases} n^{\frac{\alpha-1}{2(\alpha+2)}}, & \alpha > 1, \\ 1 & \alpha < 1. \end{cases} \tag{D.38}$$

For the case of $\alpha = 1$, we use a different construction. Define the function

$$q_n(x) := \begin{cases} \frac{(\log n)^{-1/2}}{1-x}, & x \in [0, 1-n^{-1/3}], \\ n^{1/3}(\log n)^{-1/2}, & x \in [1-n^{-1/3}, 1]. \end{cases}$$

Straightforward calculation yields

$$\int_0^1 (q'_n(x))^2 dx = \frac{n}{12\log n} < n, \quad \text{and} \quad \int_0^1 q_n^2(x)(1-x)dx = \frac{1}{3} + \frac{1}{2\log n} < 1,$$

which verifies the constraint (D.34b).

Therefore, we can lower bound the quantity $V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R))$ using the value of the variational problem at q_n, leading to the result

$$V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R)) \ge \int_0^1 q_n(x)dx \ge \frac{1}{3}\sqrt{\log n}. \tag{D.39}$$

Combining equations (D.38) and (D.39) completes the proof of the lower bound on $V_{\sigma,n}(\xi^*, \pi, g; \mathbb{B}_{\mathscr{H}}(R))$.

D.5.1.3 Proof of equation (5.34b)

By Theorem 5.1 and the claim (5.32), the minimax risk is determined by the variance functional $V_{\sigma,n}(\delta_{x_0}, \pi, g; \mathcal{F})$, defined as the optimum value of the variational problem

$$\sup_q |q(x_0)|, \quad \text{such that } q(0) = 0, \int_0^1 (q'(x))^2 dx \le n, \quad \text{and} \quad \int_0^1 (1-x)^\alpha q^2(x)dx \le 1. \tag{D.40}$$

If $x_0 = 0$, we have the trivial solution $V_{\sigma,n}(\delta_{x_0}, \pi, g; \mathcal{F}) = 0$. The rest of this section deals with the case of $x_0 = 1$ and $x_0 \in (0, 1)$, respectively.

Case I: $x_0 = 1$**.** By Lemma D.3, we have $|q(1)| \leq c_\alpha n^{\frac{1+\alpha}{2(1+\alpha)}}$ for any function q satisfying the constraints in the variational problem (D.40). On the other hand, consider the function

$$q(x) := \begin{cases} 0 & x \leq 1 - n^{\frac{-1}{2+\alpha}}, \\ n^{\frac{3+\alpha}{2(2+\alpha)}} \big(x - 1 + n^{\frac{-1}{2+\alpha}}\big) & x > 1 - n^{\frac{-1}{2+\alpha}}. \end{cases}$$

Straightforward calculation verifies that the function q satisfies the constraint in the variational problem (D.40), with $q(1) = n^{\frac{1+\alpha}{2(2+\alpha)}}$. Combining with the upper bound establishes that

$$V_{\sigma,n}(\delta_{x_0}, \pi, g; \mathcal{F}) \asymp n^{\frac{1+\alpha}{2(2+\alpha)}}.$$

Case II: $x_0 \in (0,1)$**.** For any function q satisfying the constraint in the variational problem (D.40), Lemma D.3 yields

$$|q(x_0)| \leq c_{\alpha,x_0} n^{-1/4}, \quad \text{for constant } c_{\alpha,x_0} \text{ depending on } \alpha \text{ and } x_0.$$

On the other hand, given $n \geq \max\big({x_0}^{-2}, (1-x_0)^{-1}\big)$, we construct the function

$$q(x) = \max\Big\{\frac{n^{1/4}}{2} - \frac{n^{3/4}}{2}|x - x_0|, 0\Big\}.$$

Straightforward calculation verifies that the function q satisfies the constraint in the variational problem (D.40), with $q(x_0) = n^{1/4}/2$. Combining with the upper bound establishes that

$$V_{\sigma,n}(\delta_{x_0}, \pi, g; \mathcal{F}) \asymp n^{1/4}.$$

Putting together the results under two cases completes the proof of equation (5.34b).

D.5.1.4 Proof of Lemma D.3

For any $x \in [0,1]$ and $y \in [0,x]$, by applying the Cauchy–Schwarz inequality, we find that

$$|q(y)| \geq |q(x)| - |q(y) - q(x)| \geq |q(x)| - \sqrt{(x-y)\int_y^x q'(t)^2 dt} \geq |q(x)| - \sqrt{(x-y)n}.$$

Substituting into the second constraint in equation (D.34b) yields

$$\begin{aligned} 1 &\geq \int_0^1 (1-y)^\alpha q^2(y) dy \geq \int_0^x (1-y)^\alpha \Big[|q(x)| - \sqrt{(x-y)n}\Big]_+^2 dy \\ &\geq \int_0^{\frac{q^2(x)}{n}} (z + 1 - x)^\alpha \big(|q(x)| - \sqrt{nz}\big)^2 dz \geq \int_0^{\frac{q^2(x)}{2n}} z^\alpha \big(|q(x)| - \sqrt{nz}\big)^2 dz \\ &\geq \frac{1}{4(1+\alpha)} \cdot \frac{|q(x)|^{4+2\alpha}}{n^{1+\alpha}}. \end{aligned}$$

Since the choice of $x \in [0,1]$ is arbitrary, it follows that

$$\sup_{x \in [0,1]} |q(x)| \leq c_\alpha n^{\frac{1+\alpha}{2(2+\alpha)}}. \tag{D.41}$$

On the other hand, when x is bounded away from 1, following the same derivation, we have

$$1 \geq \int_0^{\frac{q^2(x)}{n}} (z+1-x)^\alpha \big(|q(x)| - \sqrt{nz}\big)^2 dz$$
$$\geq \int_0^{\frac{q^2(x)}{2n}} (1-x)^\alpha \big(|q(x)| - \sqrt{nz}\big)^2 dz \geq \frac{(1-x)^\alpha}{4n} q^4(x),$$

which implies that

$$\sup_{x \in [0,1-\varepsilon]} |q(x)| \leq c_{\alpha,\varepsilon} n^{1/4} \tag{D.42}$$

Putting together the pieces completes the proof of Lemma D.3.

D.5.1.5 Proof of Lemma D.2

Since we focus on the ratio between $\mathbb{L}^2$-norm and sup-norm of the eigenfunction ϕ_j, we slightly abuse the notation, and use ϕ_j to denote a constant multiple of an eigenfunction of $\mathcal{K}$ under $\mathbb{L}^2$ associated to the eigenvalue λ_j. The orthogonality condition gives

$$\lambda_j \phi_j(x) = \int_0^1 \mathcal{K}(x,y)\phi_j(y)\pi(y)dy. \tag{D.43}$$

Substituting the kernel function $\mathcal{K}$ the integral equation (D.43), we have

$$\lambda_j \phi_j(x) = \int_0^x y\phi_j(y)\pi(y)dy + x \cdot \int_x^1 \phi_j(y)\pi(y)dy.$$

Taking the derivative twice yields the ordinary differential equation

$$\lambda_j \phi_j''(x) + (1-x)^\alpha \phi_j(x) = 0 \quad \text{on } [0,1].$$

Define the auxiliary function $\psi_j(z) := \phi_j\big(\lambda_j^{\frac{1}{2+\alpha}}(1-z)\big)$, the differential equation can be converted into a standard form

$$\psi_j''(z) + z^\alpha \psi_j = 0, \quad \text{for } z \in \big(0, \lambda_j^{\frac{-1}{2+\alpha}}\big)$$

Using $J_\alpha : \mathbb{R}_+ \to \mathbb{R}$ to denote the Bessel function of first kind (see [217]), the ODE above admits the closed-form solution

$$\psi_j(z) = \sqrt{z}\left\{\gamma_1(j) J_{\frac{-1}{\alpha+2}}\Big(\frac{2}{\alpha+2} z^{\alpha/2+1}\Big) + \gamma_2(j) J_{\frac{1}{\alpha+2}}\Big(\frac{2}{\alpha+2} z^{\alpha/2+1}\Big)\right\}, \text{ for } z \in \big(0, \lambda_j^{\frac{-1}{2+\alpha}}\big). \tag{D.44}$$

for a pair of constants $\gamma_1(j)$ and $\gamma_2(j)$ that may depend on j.

Since we focus on the ratio between $\mathbb{L}^2$-norm and sup-norm of ψ_j, we can assume $\gamma_1^2(j) + \gamma_2^2(j) = 1$ without loss of generality. Let ϕ_j be induced by such function ψ_j. Under this setup, we claim the following relations for any $j \geq 1$

$$\underline{c} \leq \|\phi_j\|_\infty \leq \overline{c}, \tag{D.45a}$$

$$\underline{c}\lambda_j^{\frac{\alpha}{4\alpha+8}} \leq \|\phi_j\|_{\mathbb{L}^2(\pi)} \leq \overline{c}\lambda_j^{\frac{\alpha}{4\alpha+8}}, \tag{D.45b}$$

for a pair $(\underline{c}, \overline{c})$ of constants depending only on α.

Renormalizing the eigenfunction ϕ_j to the quantity $\phi_j/\|\phi_j\|_{\mathbb{L}^2(\xi)}$, we conclude that

$$\underline{c}/\overline{c} \cdot \lambda_j^{-\frac{\alpha}{4(\alpha+2)}} \leq \|\frac{\phi_j}{\|\phi_j\|_{\mathbb{L}^2(\xi)}}\|_\infty \leq \overline{c}/\underline{c} \cdot \lambda_j^{-\frac{\alpha}{4(\alpha+2)}},$$

which proves Lemma D.2.

The rest of this section is devoted to the proofs of equations (D.45a) and (D.45b).

Proof of equation (D.45a): For $\alpha > 0$ fixed, by definition, we note have

$$\begin{aligned}\inf_{j\geq 1} \|\phi_j\|_\infty &\geq \inf_{\gamma_1^2+\gamma_2^2=1} \sup_{z\in[1,2]} \left|\gamma_1 J_{\frac{-1}{\alpha+2}}\left(\frac{2}{\alpha+2} z^{\alpha/2+1}\right) + \gamma_2 J_{\frac{1}{\alpha+2}}\left(\frac{2}{\alpha+2} z^{\alpha/2+1}\right)\right| \\ &= \sup_{z\in[1,2]} \left|\gamma_1^* J_{\frac{-1}{\alpha+2}}\left(\frac{2}{\alpha+2} z^{\alpha/2+1}\right) + \gamma_2^* J_{\frac{1}{\alpha+2}}\left(\frac{2}{\alpha+2} z^{\alpha/2+1}\right)\right|,\end{aligned}$$

where the constants (γ_1^*, γ_2^*) minimizes the expression above (the expression is uniformly continuous in (γ_1, γ_2, z), which implies continuity of the supremum in (γ_1, γ_2), and guarantees existence of a minimizer on a compact domain). Since Bessel functions $\mathcal{J}_{\frac{-1}{1+\alpha}}$ and $\mathcal{J}_{\frac{1}{1+\alpha}}$ are linearly independent on any open interval [217], there exists a constant $\underline{c} > 0$ depending only on α, such that

$$\inf_{j\geq 1} \|\phi_j\|_\infty \geq \underline{c}.$$

On the other hand, using the asymptotic formulae for Bessel functions, we note that

$$\left|J_{\frac{\pm 1}{\alpha+2}}\left(\frac{2}{\alpha+2} z^{\alpha/2+1}\right)\right| = \begin{cases} O\big(1/\sqrt{z}\big) & z \to 0, \\ O\big(z^{-\frac{\alpha}{4}-\frac{1}{2}}\big) & |z| \to \infty. \end{cases}$$

Combining with the expression (D.44) implies that the class of functions $\{\phi_j\}_{j\geq 1}$ admits a uniform upper bound $\overline{c}$, which is independent of j.

Proof of equation (D.45b)**:** Define the auxiliary functions

$$\widetilde{\psi}_j(z) = \mathbf{1}_{z>1}\sqrt{\frac{\alpha+2}{\pi}}z^{-\frac{\alpha}{4}}\Big\{\gamma_1(j)\cos\Big(\frac{2}{\alpha+2}z^{\alpha/2+1} - \frac{\alpha\pi}{4(\alpha+2)}\Big) \\ + \gamma_2(j)\cos\Big(\frac{2}{\alpha+2}z^{\alpha/2+1} + \frac{\alpha\pi}{4(\alpha+2)}\Big)\Big\}. \quad \text{(D.46)}$$

By the asymptotic approximation properties for Bessel functions [217], we have

$$\Big|\widetilde{\psi}_j(z) - \psi_j(z)\Big| \leq c'(1+z)^{-1-\frac{3}{4}\alpha}, \quad \text{for any } z \in \mathbb{R}.$$

for a constant $c' > 0$ depending only on α.

Let $\widetilde{\phi}_j := \psi\big(\lambda_j^{\frac{-1}{2+\alpha}}(1-x)\big)$, we have

$$\begin{aligned}\|\phi_j - \widetilde{\phi}_j\|^2_{\mathbb{L}^2(\pi)} &\leq \int_0^1 \Big|\psi_j\big((1/\lambda_j)^{\frac{1}{2+\alpha}}(1-x)\big) - \widetilde{\psi}_j\big((1/\lambda_j)^{\frac{1}{2+\alpha}}(1-x)\big)\Big|^2 \pi(x)dx \\ &\leq c'\int_0^1 \frac{x^\alpha}{1 + \big((1/\lambda_j)^{\frac{1}{2+\alpha}}x\big)^{2+\frac{3}{2}\alpha}}dx \\ &\leq c_2(\alpha)\lambda_j^{\frac{\alpha+1}{\alpha+2}}, \end{aligned} \quad \text{(D.47)}$$

where the constant $c_2(\alpha)$ depends only on α.

For the function $\widetilde{\phi}_j$, we can compute its $\mathbb{L}^2(\pi)$-norm.

$$\begin{aligned}\|\widetilde{\phi}_j\|^2_{\mathbb{L}^2(\pi)} &= \lambda_j^{\frac{1+\alpha}{2+\alpha}} \int_1^{(\frac{1}{\lambda_j})^{\frac{1}{2+\alpha}}} \widetilde{\psi}_j^2(z)z^\alpha dz \\ &= \frac{2}{\alpha+2}\cdot\lambda_j^{\frac{1+\alpha}{2+\alpha}} \int_1^{1/\sqrt{\lambda_j}} \widetilde{\psi}_j^2\big(\theta^{\frac{2}{2+\alpha}}\big)\theta^{\frac{\alpha}{\alpha+2}}d\theta \\ &= \frac{2}{\pi}\cdot\lambda_j^{\frac{1+\alpha}{2+\alpha}} \int_1^{1/\sqrt{\lambda_j}} \Big\{\gamma_1(j)\cos\Big(\frac{2\theta-\alpha\pi/4}{\alpha+2}\Big) + \gamma_2(j)\cos\Big(\frac{2\theta+\alpha\pi/4}{\alpha+2}\Big)\Big\}^2 d\theta.\end{aligned} \quad \text{(D.48)}$$

Note that the integral is with respect to a periodic function, we have the upper bound

$$\int_1^{1/\sqrt{\lambda_j}} \Big\{\gamma_1(j)\cos\Big(\frac{2\theta-\alpha\pi/4}{\alpha+2}\Big) + \gamma_2(j)\cos\Big(\frac{2\theta+\alpha\pi/4}{\alpha+2}\Big)\Big\}^2 d\theta \leq c_3(\alpha)\cdot\lambda_j^{-\frac{1}{2}},$$

and the lower bound

$$\begin{aligned}&\int_1^{1/\sqrt{\lambda_j}} \Big\{\gamma_1(j)\cos\Big(\frac{2\theta-\alpha\pi/4}{\alpha+2}\Big) + \gamma_2(j)\cos\Big(\frac{2\theta+\alpha\pi/4}{\alpha+2}\Big)\Big\}^2 d\theta \\ &\geq \Big\{\frac{1}{\pi(\alpha+2)\sqrt{\lambda_j}} - 2\Big\} \\ &\quad \times \int_0^{\pi(\alpha+2)} \Big\{\gamma_1(j)\cos\Big(\frac{2\theta-\alpha\pi/4}{\alpha+2}\Big) + \gamma_2(j)\cos\Big(\frac{2\theta+\alpha\pi/4}{\alpha+2}\Big)\Big\}^2 d\theta\end{aligned}$$

For any pair (γ_1, γ_2) such that $\gamma_1^2 + \gamma_2^2 = 1$, we have

$$\begin{aligned}
&\int_0^{2\pi} \Big\{ \gamma_1 \cos\Big(\theta - \frac{\alpha\pi/4}{\alpha+2}\Big) + \gamma_2 \cos\Big(\theta + \frac{\alpha\pi/4}{\alpha+2}\Big) \Big\}^2 d\theta \\
&= \pi - 2\gamma_1\gamma_2 \int_0^{2\pi} \cos\Big(\theta - \frac{\alpha\pi/4}{\alpha+2}\Big) \cos\Big(\theta + \frac{\alpha\pi/4}{\alpha+2}\Big) d\theta \\
&= \pi - 2\pi\gamma_1\gamma_2 \cos\big(\frac{\alpha\pi}{2\alpha+4}\big) \\
&\geq \pi\Big\{1 - \cos\big(\frac{\alpha\pi}{2\alpha+4}\big)\Big\},
\end{aligned}$$

which is a positive constant depending only on α, and independent of γ_1 and γ_2.

Substituting these bounds back to equation (D.48) yields

$$c_4'(\alpha)\lambda_j^{\frac{\alpha}{4\alpha+8}} \leq \|\widetilde{\phi}_j\|_{\mathbb{L}^2(\pi)} \leq c_3'(\alpha)\lambda_j^{\frac{\alpha}{4\alpha+8}}$$

Combining with equation (D.47) completes the proof of equation (D.45b).

D.5.2 Proof of Corollary 5.2

We prove the claims about the averaged functional τ^* and the one-point functional $\mathscr{L}_\omega(\mu^*, \delta_{x_0})$ separately in the following two subsections.

D.5.2.1 Bounds on the minimax risk for the averaged functional

By Theorems 5.1 and 5.3, we have $\mathscr{M}_n\big(\mathbb{B}_{\mathbb{H}}(1)\big) \asymp n^{-1}\big\{V_{\xi^*}^2(\mu^*) + V_{\sigma,n}^2(\xi^*, \pi, g; \mathcal{F})(\mathcal{F}; n)\big\}$. Taking the measure $\omega(\cdot \mid x) = \delta_x$ for any $x \in \mathbb{X}$, it can be seen that

$$V_{\xi^*}^2(\mu^*) = \mathrm{var}_{X\sim\xi}\Big(\mu^*(X, T(X))\Big).$$

By the generalized Morrey's embedding theorem (e.g., see §5.6.3 in Evans [58]), for any smoothness index $s > (d_x + d_a)/2$, we have

$$\sup_{(x,a)\in\mathcal{S}\times\mathbb{A}} |\mu^*(x,a)| \leq c'\|\mu^*\|_{\mathbb{H}} \leq c',$$

for a constant c' depending on the triple (d_x, d_a, s). So we have $V_{\xi^*}^2(\mu^*) \precsim 1$ in the worst case.

Since the conditional variance function is a constant, we have

$$V_{\sigma,n}^2(\xi^*, \pi, g; \mathcal{F}) = \sum_{j,k\geq 1} \frac{\bar{u}_{j,k}^2}{1 + \frac{1}{n}\frac{1}{\lambda_{j,k}}},$$

where we define the projection coefficients

$$\bar{u}_{j,k} = \int_{\mathbb{X}} \phi_j(x)\psi_k(T(x))dx = \langle \phi_j,\, \psi_k \circ T\rangle_{\mathbb{L}^2(\mathbb{X})}.$$

By the eigenvalue decay condition (5.35), we have $\lambda_{j,k}^{-1} \asymp j^{2s/d_x} + k^{2s/d_a}$, which implies that

$$n^{-1}V^2_{\sigma,n}(\xi^*,\pi,g;\mathcal{F})(\mathcal{F};n) \asymp V_{\bar{\sigma},n}(\xi^*,\pi,g;\mathbb{B}_{\mathscr{H}}(R)) = \sum_{j,k\geq 1} \frac{\left|\langle \phi_j,\, \psi_k \circ T\rangle_{\mathbb{L}^2(\mathbb{X})}\right|^2}{n + j^{2s/d_x} + k^{2s/d_a}}, \tag{D.49}$$

which proves the instance-dependent bound.

By Parseval's identity, for each $k \geq 1$, we have

$$\sum_{j=1}^{\infty} \langle \phi_j,\, \psi_k \circ T\rangle^2_{\mathbb{L}^2(\mathbb{X})} = \|\psi_k \circ T\|^2_{\mathbb{L}^2(\mathbb{X})} \leq \|\psi_k\|^2_\infty = 1.$$

Substituting into the instance-dependent bound (D.49), we have the worst-case instantiation

$$\begin{aligned}\frac{1}{n}V^2_{\sigma,n}(\xi^*,\pi,g;\mathcal{F})(\mathcal{F};n) &\leq \sum_{k=1}^{n^{\frac{d_a}{2s}}}\sum_{j=1}^{\infty} \frac{\left|\langle \phi_j,\, \psi_k \circ T\rangle_{\mathbb{L}^2(\mathbb{X})}\right|^2}{n} + \sum_{k=n^{\frac{d_a}{2s}}}^{\infty}\sum_{j=1}^{\infty} \frac{\left|\langle \phi_j,\, \psi_k \circ T\rangle_{\mathbb{L}^2(\mathbb{X})}\right|^2}{k^{2s/d_a}}\\ &\leq \Big\{1 + \frac{2s}{2s-d_a}\Big\} n^{\frac{d_a}{2s}-1}.\end{aligned}$$

Thus, we obtain the worst-case upper bound $\sup_T \mathscr{M}_n\big(\mathbb{B}_{\mathbb{H}}(1)\big) \lesssim n^{\frac{d_a}{2s}-1}$.

On the other hand, for any $a_0 \in \mathbb{A}$, taking the target functional $T_{a_0}(x) \equiv a_0$ for any $x \in \mathbb{X}$, we have

$$n^{-1}V^2_{\sigma,n}(\xi^*,\pi,g;\mathcal{F}) = \sum_{j,k\geq 1} \frac{|\psi_k(a_0)|^2 \left|\langle \phi_j,\, \mathbf{1}\rangle_{\mathbb{L}^2(\mathbb{X})}\right|^2}{n + j^{2s/d_x} + k^{2s/d_a}} = \sum_{k\geq 1} \frac{|\psi_k(a_0)|^2}{n+1+k^{2s/d_a}}$$

For the Fourier basis ψ_k, we have $|\psi_k(a_0)|$ for any $a_0 \in \mathbb{A}$, which leads to the lower bound

$$\sup_T \mathscr{M}_n\big(\mathbb{B}_{\mathbb{H}}(1)\big) \gtrsim \sum_{k\geq 1} \frac{1}{n+1+k^{2s/d_a}} \gtrsim n^{\frac{d_a}{2s}-1}.$$

D.5.2.2 Minimax bounds for the one-point functional

For any $x_0 \in \mathbb{X}$, Theorem 5.1(b) and equation (5.32) imply that $\mathscr{M}_n\big(x_0;\mathcal{F}\big) \asymp n^{-1}V^2_{\sigma,n}(\delta_{x_0},\pi,g;\mathcal{F})$. By the variational representation of $V^2_{\sigma,n}(\delta_{x_0},\pi,g;\mathcal{F})$, it can be seen that

$$V^2_{\sigma,n}(\delta_{x_0},\pi,g;\mathcal{F}) \asymp \sum_{j,k\geq 1} \frac{\left|\phi_j(x_0)\psi_k(T(x_0))\right|^2}{1+n^{-1}\lambda_{j,k}} = \sum_{j,k\geq 1} \frac{1}{1+n^{-1}\lambda_{j,k}},$$

where the last equation follows from the fact that the complex Fourier bases ϕ_j and ψ_k take value at unit circle.

Substituting with the eigenvalue decay condition (5.35), we obtain that

$$n^{-1}V_{\sigma,n}^2(\delta_{x_0},\pi,g;\mathcal{F}) \asymp \sum_{j,k\geq 1}\frac{1}{n+j^{2s/d_x}+k^{2s/d_a}} =: S_n.$$

It remains to study the summation S_n. On the one hand, we note that

$$S_n \geq \sum_{j=1}^{n^{\frac{d_x}{2s}}}\sum_{k=1}^{n^{\frac{d_a}{2s}}}\frac{1}{n+j^{2s/d_x}+k^{2s/d_a}} \geq \frac{1}{3}n^{\frac{d_x+d_a}{2s}-1}.$$

On the other hand, we have the upper bound

$$S_n \leq c_{d_x,s}\sum_{j=1}^{\infty}\left(n+j^{2s/d_x}\right)^{\frac{d_a}{2s}-1} \leq c_{d_x,d_a,s}n^{\frac{d_x+d_a}{2s}-1}.$$

Therefore, we conclude that $\mathscr{M}_n(x_0;\mathcal{F}) \asymp n^{\frac{d_x+d_a}{2s}-1}$ for any $x_0 \in \mathbb{X}$ and deterministic policy T.