NP-COMPLETE DECISION PROBLEMS FOR BINARY QUADRATICS

by

Kenneth L. Manders and Leonard Adleman

Memorandum No. ERL-M615

18 November 1976

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

NP-COMPLETE DECISION PROBLEMS FOR BINARY QUADRATICS*

Kenneth L. Manders

Group in Logic and Methodology of Science
University of California
Berkeley, California 94720

and

Leonard Adleman**

Department of Applied Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

* Research supported by National Science Foundation Grant DCR72-03725-A02.

** While participating in this research, the second author was affiliated with: Computer Science Division, Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley 94720.

Running Head: NP-Complete Binary Quadratics

Key Words: NP-complete, binary quadratics, nondeterministic Diophantine machine, computational complexity, Diophantine equation

Correspondence: Kenneth L. Manders
Group in Logic and Methodology of Science
Evans Hall
University of California
Berkeley, California 94720

Abstract

The computational complexity of deciding whether a polynomial with integer coefficients has natural-number zeroes ranges from deterministic polynomial time feasibility (for polynomials in one variable or of degree one) to undecidability (presently known to hold for polynomials in 9 or more variables). We show that for the 2-variable quadratics of the form

$$\alpha x^2 + \beta y - \gamma = 0 \; ; \quad \alpha, \beta, \gamma \in \omega$$

the problem is NP-complete. This implies NP-completeness of certain questions about the solutions x of

$$x^2 \equiv \alpha \text{ modulo } \beta \; ; \quad \alpha, \beta, x \in \omega \; .$$

It also shows that a nondeterministic Turing machine restricted to evaluating deterministically polynomials of a given form at nondeterministically constructed argument values (called a nondeterministic Diophantine machine below) can solve an NP-complete problem in polynomial time.

1. Introduction

Many number theoretical questions are formulated or can be formulated as questions about the solvability of Diophantine equations (i.e. $p(x_1 \cdots x_n) = 0$ for a multivariable polynomial $p(x_1 \cdots x_n)$ with integer coefficients) in natural numbers or integers. This becomes clear upon examination of the section on Diophantine problems in any standard number theory text; it is in fact made mathematically precise by Matijasevic's theorem [10] that all recursively enumerable sets are Diophantine (i.e. the elements of the set S correspond to the parameter values a for which an appropriate Diophantine equation $p_S(a,x_1 \cdots x_n) = 0$ has a solution). It is therefore of fundamental importance to consider algorithms for deciding the solvability of Diophantine equations, as Hilbert stressed in asking for such algorithms in the 10^{th} of his famous set of mathematical problems in 1900 [6]. One wishes to know (a) the complexity of deciding solvability for various classes of Diophantine equations, and especially (b) for which subclasses of Diophantine equations a feasible (i.e. deterministic polynomial time) algorithmic procedure to decide solvability exists.

It follows from Matijasevic's theorem that for some fixed n, the set of solvable n-variable Diophantine equations is nonrecursive. Much effort has been devoted to determining the minimum n for which this is true. The best published result is $n \leq 13$ [11]; Matijasevic has improved this to $n \leq 9$ [12]. In [11], it is conjectured that $n = 3$ may be possible, though this cannot be shown by present methods.

Even for equations with two unknowns, the decision problem is tantalizingly difficult. For example, the first major positive contribution after Hilbert's address [6] was Thue's theorem [18] (1909) that for any polynomial $f(x,y)$ irreducible over the rationals and homogeneous of degree ≥ 3,

$$f(x,y) - m = 0$$

has at most finitely many integer solutions. Regrettably, Thue's argument provided no algorithm for deciding the solvability of equations of this form, and no such algorithm was found until Alan Baker's fundamental contribution [4] (1967) which yields a nondeterministic exponential time algorithm. Baker's methods extend to various equations in two unknowns.

On the other hand, the question of solvability (in integers or natural numbers) of linear equations in two unknowns can be answered in deterministic polynomial time. Thus only the question for binary quadratics remains in the gap between tractable (binary linear) and apparently intractable (binary of degree 3) solvability questions. We show below that for binary quadratics of the form

$$\alpha x_1^2 + \beta x_2 - \gamma = 0 \ , \quad \alpha, \beta, \gamma \in \omega$$

the problem of deciding whether there are natural-number solutions is NP-complete. The reduction also shows that the problem of deciding whether, for $\alpha, \beta, \gamma \in \omega$,

$$\begin{cases} x^2 \equiv \alpha \text{ modulo } \beta \\ 0 \leq x \leq \gamma \end{cases}$$

has a solution in natural numbers is NP-complete, even if the prime factorization of the modulus β is given. If the restriction on the size of x is omitted, this problem can be solved in deterministic polynomial time using the factorization of β. Thus the results yield a fairly precise indication of the class of binary Diophantine equations of this type for which the solvability question is tractable.

The problems which we show NP-complete are distinctly number-theoretic in character; they also differ from the classically NP-complete problems (such as propositional satisfiability) with respect to the number of variables involved: Most known NP-complete problems involve an unbounded number of variables, whereas our problems involve only two variables. Because of these properties, we hope that the NP-completeness of these problems will play a role in showing the NP-completeness of further problems of a numerical nature, much as the propositional satisfiability problem has played in showing the NP-completeness of combinatorial problems [5], [7].

Our results also give information about a different type of question: What is lost in computational ability and efficiency of a nondeterministic Turing machine if we restrict it to deterministically evaluating polynomials of a fixed form at a nondeterministically constructed argument value, and accepting if and only if the polynomial evaluates to 0? This question is of interest because this computational model is number theoretically convenient; the sets accepted by such machines are exactly the Diophantine sets. In section 3 below we consider this model of computation, called a nondeterministic Diophantine machine (NDDM). We give an overview of what is known about the computational ability of NDDM's and relate the results of the present paper to this question. All known results support the conjecture that nondeterministic computation can be studied without loss of generality on NDDM's; this possibility suggests many research questions.

The NP-completeness results are formulated and proved in section 2; section 3 is devoted to the discussion of NDDM's. Research problems are formulated at the ends of both sections.

2. Main Results

We recall the following (well-known) definitions: A relation R on the natural numbers is accepted in polynomial time by a (deterministic or nondeterministic) Turing machine M if and only if there is a polynomial $q(\cdot)$ such that for any $x_1,\ldots,x_n \in \omega$

$$\langle x_1 \cdots x_n \rangle \in R \Leftrightarrow \text{there is a computation of } M \text{ on input } \langle x_1 \cdots x_n \rangle$$
$$\text{which halts in an accepting state within } q(|\vec{x}|)$$
$$\text{steps, where } |x| \text{ is the length of } \vec{x} \text{ in binary,}$$
$$\vec{x} = \langle x_1 \cdots x_n \rangle .$$

P is the collection of relations on the natural numbers accepted by some deterministic Turing machine (DTM) in polynomial time; NP is the collection of relations on the natural numbers accepted by some nondeterministic Turing machine (NDTM) in polynomial time; a set in NP is NP-complete if for any set $A \in NP$ there is a deterministic polynomial time computable recursive function $f(\cdot)$ such that

$$(\forall x \in \omega)[x \in A \Leftrightarrow f(x) \in S] .$$

Theorem 1. The (problem of accepting the) set of Diophantine equations (in a standard binary encoding) of the form

$$\alpha x_1^2 + \beta x_2 - \gamma = 0 ; \quad \alpha, \beta, \gamma \in \omega$$

which have natural-number solutions x_1, x_2 is NP-complete.

Theorem 2. The (problem of accepting the) set of quadratic congruences (in a standard encoding)

$$x^2 \equiv \alpha \text{ modulo } \beta$$

with solutions $x \in \omega$ satisfying

$$0 \leq x \leq \gamma \; ; \quad \alpha, \beta, \gamma \in \omega$$

is NP-complete.

Supplement (to Theorems 1 and 2). If we consider only Diophantine equations (Theorem 1) and congruences (Theorem 2) in which all prime factors of β are less than $\log \beta$, the sets are still NP-complete.

Theorems 1 and 2 are obtained by a common argument. Let S be the set of satisfiable propositional formulas in conjunctive normal form with at most 3 literals per clause. By Cook [5] it suffices to show that there is a deterministic polynomial-time algorithm which reduces the problem of membership in S to a problem of the form(s) mentioned in the theorem, and that the problems themselves are in NP. Both problems considered are solvable by a nondeterministic "guess a solution and check whether it is correct" algorithm in polynomial time, and hence in NP. This is because, as is easily verified, there is a bound on the size of possible solutions to either problem given by a polynomial in the coefficients α, β, γ.

For proof of the Supplement, we must also show that the reduction algorithms yield only equations and congruences with β having all prime factors less than $\log \beta$. This will be evident from the algorithm.

We now give the reduction algorithm, followed by proof of correctness and analysis of computation time. The reader may wish to merely skim the algorithm initially, referring back to it when this is suggested in the proof and analysis.

The reduction algorithms for Theorems 1 and 2 are identical except for the final step, which will be given separately. The initial steps of the

algorithm in fact give a reduction of 3-satisfiability to a convenient special case of Knapsack (see [7]). The basic idea behind the algorithm is to set up means of going back and forth between the representation of a sequence as the digits of a number in some base, and as the residues of a number with respect to a system of moduli. (See comment (1) following the proof.)

2.1 The Algorithm

"On input ϕ, read ϕ and eliminate all duplicate conjuncts and those in which, for some variable x_i, both x_i and $\bar{x}_i$ occur. Count the ℓ variables occurring in the remaining formula ϕ_R. Let

$$\Sigma = \{\sigma_1,\ldots,\sigma_m\}$$

be a standard enumeration of all possible disjunctive clauses, formed from $x_1,\ldots,x_\ell$ and their complements, with at most 3 literals per clause and no variable occurring twice or both complemented and uncomplemented in a clause. Setting

$$\varepsilon_j = \begin{cases} 1 & \text{if } \sigma_j \text{ occurs in } \phi_R \\ 0 & \text{otherwise} \end{cases}, \quad j = 1,2,\ldots,m$$

compute $\tau_\phi = \sum_{j=1}^{m} \varepsilon_j \cdot 8^j$.

[Comment: τ_ϕ is the only quantity computed which depends specifically on ϕ_R, rather than just on the number ℓ of variables occurring in ϕ_R.]

Compute:

$$f_i^+ = \sum_{\substack{x_i \text{ occurs} \\ \text{in } \sigma_j}} 8^j , \quad i = 1,2,\ldots,\ell$$

$$f_i^- = \sum_{\substack{\bar{x}_i \text{ occurs} \\ \text{in } \sigma_j}} 8^j , \quad i = 1,2,\ldots,\ell$$

Set $n = 2m+\ell$ and compute c_j, $j = 0,\ldots,n$ as

$$\left.\begin{aligned} c_0 &= 1 \\ c_j &= -\tfrac{1}{2}8^k , \quad j = 2k-1 \\ c_j &= -8^k , \quad j = 2k \end{aligned}\right\} \quad j = 1,\ldots,2m$$

$$c_j = \tfrac{1}{2}(f^+_{j-2m} - f^-_{j-2m}) \qquad j = 2m+1,\ldots,2m+\ell$$

and

$$\tau = \tau_\phi + \sum_{j=0}^{n} c_j + \sum_{i=1}^{\ell} f_i^- .$$

[Comment: At this point, we have in fact obtained a knapsack problem $\sum_{j=0}^{n} c_j\alpha_j = \tau$, $\alpha_j \in \{-1,+1\}$ which is solvable if and only if ϕ is satisfiable; moreover, for any value of $\alpha_j \in \{-1,+1\}$, $|\sum_{j=0}^{n} c_j\alpha_j - \tau| < \frac{1}{2}8^{m+1}$, so the knapsack problem is equivalent to $\sum_{j=0}^{n} c_j\alpha_j \equiv \tau$ modulo 8^{m+1}, $\alpha_j \in \{-1,+1\}$. These assertions will become clear from the proof of correctness.]

Determine the first $n+1$ primes, $p_0,\ldots,p_n$, exceeding

$$\sqrt[n+1]{4(n+1)8^{m+1}} .$$

[This in fact never exceeds 12, so we can set $p_0 = 13$.]

Determine parameters θ_j, $j = 0,1,\ldots,n$, as: the least $\theta_j \in \omega$ such that

$$\left.\begin{aligned} \theta_j &\equiv c_j \text{ modulo } 8^{m+1} \\ \theta_j &\equiv 0 \text{ modulo } \prod_{i\neq j}^{n} p_i^{n+1} \\ \theta_j &\not\equiv 0 \text{ modulo } p_j \end{aligned}\right\}$$

Compute $H = \sum_{j=0}^{n} \theta_j$, $K = \prod_{j=0}^{n} p_j^{n+1}$ and output:

(a) for Theorem 1:

$$(K+1)^3 \cdot 2 \cdot 8^{m+1} \cdot (H^2 - x_1^2) + K(x_1^2 - \tau^2) - x_2 \cdot 2 \cdot 8^{m+1} \cdot K = 0$$

(b) for Theorem 2:

$$\begin{cases} x^2 \equiv (2 \cdot 8^{m+1} + K)^{-1} \cdot (K\tau^2 + 2 \cdot 8^{m+1} H^2) \text{ modulo } 2 \cdot 8^{m+1} \cdot K \\ 0 \leq x \leq H \end{cases}$$

where $(2 \cdot 8^{m+1} + K)^{-1}$ is the inverse of $(2 \cdot 8^{m+1} + K)$ modulo $2 \cdot 8^{m+1} \cdot K$.

2.2 Analysis of Computation Time

ℓ, and hence m and n, are bounded by a polynomial in the length of the input ϕ. Hence, by the Prime Number Theorem, the primes $p_0, \ldots, p_n$ are also bounded by such a polynomial. It follows that the sizes (numbers of digits in binary representation) of p_j^{n+1}, K, and H are bounded by a polynomial in the length of ϕ; hence the same is true of the output of the algorithm.

Moreover, we can obtain all quantities needed deterministically within polynomial time in the length of the input: The primes can be found as we have exponential time in their length to do so; i.e. we can afford to sieve for the primes. Each θ_j is of the form

$$\lambda_j \cdot \prod_{\substack{i=0 \\ i \neq j}}^{n} p_i^{n+1} \quad \text{or} \quad (\lambda_j + 8^{m+1}) \cdot \prod_{\substack{i=0 \\ i \neq j}}^{n} p_i^{n+1}$$

where

$$\lambda_j \equiv c_j \cdot \Big(\prod_{\substack{i=0 \\ i\neq j}}^{n} p_i^{n+1}\Big)^{-1} \text{ modulo } 8^{m+1} \quad , \quad 0 \leq \lambda_j < 8^{m+1}$$

and the inverse can be found in polynomial time [9] using the Euclidean Algorithm [15]. All other computations are trivially polynomial time, given the bounds on the numbers involved.

2.3 Proof of Correctness

In this section, '$|x|$' will denote the absolute value of x.

We first show that the original propositional formula ϕ is satisfiable if and only if

$$\sum_{j=0}^{n} \theta_j \alpha_j \equiv \tau \text{ modulo } 8^{m+1} \quad , \quad \alpha_j \in \{-1,+1\}, \quad j = 0,\ldots,n$$

is solvable.

Clearly, ϕ_R is satisfiable if and only if ϕ is; ϕ_R is satisfiable if and only if there is a valuation $r: \{x_1,\ldots,x_\ell\} \to \{0,1\}$ such that for each disjunctive clause $\sigma_k \in \{\sigma_1,\ldots,\sigma_m\}$

$$0 = R_k = \begin{cases} y_k - \sum\limits_{x_i \in \sigma_k} r(x_i) - \sum\limits_{\bar{x}_i \in \sigma_k} (1-r(x_i)) + 1 & \text{if } \sigma_k \in \phi_R \\ y_k - \sum\limits_{x_i \in \sigma_k} r(x_i) - \sum\limits_{\bar{x}_i \in \sigma_k} (1-r(x_i)) & \text{if } \sigma_k \notin \phi_R \end{cases}$$

is solvable by $y_k \in \{0,1,2,3\}$. (In this definition of R_k, '$\in$' abbreviates 'occurs in'.) For convenience on a technical point later in the proof, we add an artificial condition

$$0 = R_0 = \alpha_0 + 1 \quad , \quad \alpha_0 \in \{-1,+1\}$$

which does not influence the satisfiability of the system.

For any ϕ_R, any valuation $r: \{x_1,\ldots,x_\ell\} \to \{0,1\}$, and any $y_k \in \{0,1,2,3\}$, we have

$$-3 \leq R_k \leq 4 , \quad k = 1,2,\ldots,m$$
$$0 \leq R_0 \leq 2$$

and therefore

$$R_k = 0,\ k = 0,1,\ldots,m \ \Leftrightarrow \ \sum_{k=0}^{m} R_k \cdot 8^k = 0$$

and

$$\left| \sum_{k=0}^{m} R_k \cdot 8^k \right| \leq \frac{1}{2} \cdot 8^{m+1}$$

so that

$$R_k = 0,\ k = 0,1,\ldots,m \ \Leftrightarrow \ \sum_{k=0}^{m} R_k 8^k \equiv 0 \text{ modulo } 8^{m+1} .$$

In this condition, we replace the variables y_k and $r(x_i)$ by $\{-1,+1\}$-valued variables α_{2k-1}, α_{2k} and α_{2m+i} respectively:

$$y_k = -\frac{1}{2}[(1-\alpha_{2k-1}) + 2\cdot(1-\alpha_{2k})]$$
$$r(x_i) = -\frac{1}{2}(1-\alpha_{2m+i})$$

Let R'_k result from R_k by these substitutions. Then the condition

$$\sum_{k=0}^{m} R'_k 8^k \equiv 0 \text{ modulo } 8^{m+1}$$

contains only the $\{-1,+1\}$-valued variables $\alpha_0,\ldots,\alpha_{2m+\ell}$, and is solvable if and only if ϕ is satisfiable. By rearrangement of terms, using the quantities defined in the algorithm, the condition can be rewritten as

$$\sum_{j=0}^{n} c_j\alpha_j \equiv \tau \text{ modulo } 8^{m+1} , \quad \alpha_j \in \{-1,+1\}$$

which by the definition of θ_j, $j = 0,\ldots,n$, is equivalent to

$$\sum_{j=0}^{n} \theta_j\alpha_j \equiv \tau \text{ modulo } 8^{m+1} , \quad \alpha_j \in \{-1,+1\} .$$

Lemma 1. Let K and H be as in the algorithm. The general solution of the system

$$\left.\begin{aligned} 0 \leq |x| \leq H , \quad x \in \mathbb{Z} \qquad (1)\\ (H+x)(H-x) \equiv 0 \bmod K \qquad (2)\end{aligned}\right\}$$

is given by

$$x = \sum_{j=0}^{n} \alpha_j\theta_j , \quad \alpha_j \in \{-1,+1\}, \quad j = 0,1,\ldots,n .$$

Proof (of Lemma 1). It is easy to verify that all x of the given form satisfy the system. We now show that these are the only solutions.

Let x be a solution to the system (1)-(2). Then

$$(H+x)(H-x) \equiv 0 \bmod (p_j)^{n+1} , \quad j = 0,1,\ldots,n .$$

Assume (for reductio) that for some j_0,

$$p_{j_0}|(H+x) \quad \text{and} \quad p_{j_0}|(H-x) .$$

(Notation: $a|b$ means a divides b; this is equivalent to $b \equiv 0 \bmod a$.) Then $p_{j_0}|(H+x) + (H-x) = 2H$. But $p_{j_0} > 2$, p_{j_0} prime, so we must have $p_{j_0}|H$, i.e. $p_{j_0}|\sum_{j}^{n} \theta_j$. But by definition of θ_j, $p_{j_0}|\theta_j$ for all $j \neq j_0$. Hence it would have to be that $p_{j_0}|\theta_{j_0}$, contradicting the third condition

in the definition of θ_{j_0}.

Thus rejecting our assumption, we conclude that for each j, p_j^{n+1} divides exactly one of $(H+x)$ and $(H-x)$. We define

$$\alpha_j = \begin{cases} 1 & \text{if } p_j^{n+1} \mid (H-x) \\ -1 & \text{if } p_j^{n+1} \mid (H+x) \end{cases}$$

$$x' = \sum \alpha_j \theta_j$$

Then we get:

$$x' \equiv x \bmod p_j^{n+1} \quad \text{for all } j \text{ so } x' \equiv x \bmod K$$

$$\left.\begin{aligned} -H \leq x' \leq H \\ -H \leq x \leq H \end{aligned}\right\} \Rightarrow |x-x'| \leq 2H$$

But by our choice of $p_j \geq \sqrt[n+1]{4(n+1)8^{m+1}}$, and the fact that $\lambda_j = \theta_j / \prod_{\substack{i=0 \\ i\neq j}}^{n} p_i^{n+1}$ $< 2 \cdot 8^{m+1}$ (for each j), each term of H is bounded by $K/2(n+1)$. Hence $2H < K$; we now conclude that $x = x'$. Thus any solution of (1)-(2) is indeed of the form given. (End of proof.)

Using the lemma, we find that the condition

$$\sum_{j=0}^{n} \theta_j \alpha_j \equiv \tau \text{ modulo } 8^{m+1} \quad , \quad \alpha_j \text{ e } \{-1,+1\}$$

is equivalent to the system

$$\left.\begin{aligned} &\text{(i)} \quad 0 \leq |x| \leq H \ , \quad x \text{ e } \mathbb{Z} \\ &\text{(ii)} \quad x \equiv \tau \text{ modulo } 8^{m+1} \\ &\text{(iii)} \quad (H+x)(H-x) \equiv 0 \text{ modulo } K \end{aligned}\right\} \qquad \text{(I)}$$

Lemma 2. Let τ be odd, $x \in \mathbb{Z}$, $k \geq 3$.

$$(\tau+x)(\tau-x) \equiv 0 \bmod 2^{k+1} \Leftrightarrow \text{either } \tau+x \equiv 0 \text{ or } \tau-x \equiv 0 \bmod 2^{k}.$$

(The straightforward proof is left to the reader.)

We note that in our case the conditions of Lemma 2 are satisfied. Hence the system (I) is satisfiable if and only if system (II) below is satisfiable:

$$\left.\begin{aligned} &\text{(i)} \quad 0 \leq |x| \leq H, \quad x \in \mathbb{Z} \\ &\text{(ii)} \quad (\tau+x)(\tau-x) \equiv 0 \bmod 2\cdot 8^{m+1} \\ &\text{(iii)} \quad (H+x)(H-x) \equiv 0 \bmod K \end{aligned}\right\} \qquad \text{(II)}$$

For if x satisfies (I), clearly x satisfies (II); if x satisfies (II), then either $\tau+x \equiv 0 \bmod 8^{m+1}$ or $\tau-x \equiv 0 \bmod 8^{m+1}$; in the second case (I) is satisfied by x, in the first case, $-x$ satisfies conditions (i) and (iii) common to both systems, and

$$-x \equiv \tau \bmod 8^{m+1}$$

so that $-x$ satisfies system (I).

Finally, system (II) is equivalent to

$$\left.\begin{aligned} &\text{(i)} \quad 0 \leq x_1 \leq H, \quad x_1 \in \mathbb{Z} \\ &\text{(ii)} \quad \lambda_1 \cdot 2\cdot 8^{m+1}(H^2-x_1^2) + \lambda_2 K(\tau^2-x_1^2) \equiv 0 \text{ modulo } 2\cdot 8^{m+1}\cdot K \\ &\text{(iii)} \quad \gcd(\lambda_1,K) = \gcd(\lambda_2, 2\cdot 8^{m+1}) = 1; \quad \lambda_1, \lambda_2 \in \mathbb{Z} \end{aligned}\right\} \qquad \text{(III)}$$

For system (II) only involves x^2 and $|x|$, so that we may assume $x \geq 0$ without loss of generality; and the congruences (ii) and (iii) of (II) are equivalent to (ii), (iii) of (III) because $2\cdot 8^{m+1}$ and K are relatively prime. The parameters λ_1, λ_2 of system (III) can be freely chosen subject

to conditions III(iii), and we will make different choices for the proofs of Theorems 1 and 2; conditions (III)(i), (ii) are satisfiable either for all λ_1, λ_2 satisfying (III)(iii) or for no such λ_1, λ_2.

We now complete the arguments separately for Theorems 1 and 2.

(a) Theorem 1. We choose $\lambda_1 = (K+1)^3$, $\lambda_2 = -1$, clearly satisfying (III)(iii). Now for $x \geq 0$

$$(K+1)^3 \cdot 2 \cdot 8^{m+1} \cdot (H^2 - x_1^2) + K(x_1^2 - \tau^2) \geq 0 \Leftrightarrow 0 \leq x_1 \leq H$$

For the first inequality can be written as

$$x_1^2 \leq H^2 + \frac{H^2 - \tau^2}{(K+1)^3 \cdot K^{-1} \cdot 2 \cdot 8^{m+1} - 1} = \text{RHS}$$

and because $\tau < H < K$, RHS satisfies $H^2 < \text{RHS} < H^2 + 1$. It follows that the equation output by the algorithm is solvable by x_1, x_2 e ω if and only if system (III) is satisfiable, i.e. if ϕ is satisfiable.

(b) Theorem 2. We choose $\lambda_1 = \lambda_2 = 1$, satisfying (III)(iii). Then (III)(ii) becomes

$$(2 \cdot 8^{m+1} + K)x^2 \equiv K\tau^2 + 2 \cdot 8^{m+1} H^2 \text{ modulo } 2 \cdot 8^{m+1} \cdot K$$

and as $2 \cdot 8^{m+1} + K$ is relatively prime to $2 \cdot 8^{m+1} \cdot K$, it has an inverse modulo $2 \cdot 8^{m+1} \cdot K$. Multiplying by the inverse, we obtain the congruence condition output by the algorithm. Thus again the conditions output by the algorithm are satisfiable if and only if system III is satisfiable.

2.4 Comments on the Reduction

(1) It is of broader interest to clarify the basic idea of the reduction algorithm from which a general method for reducing computational problems to Diophantine equations by deterministic computation can be derived. The crucial elements are contained in the system of definitions in the algorithm, and Lemma 1 of the above proof. A version of this pared to bare essentials may be obtained by picking $p_1,\ldots,p_n$ any sufficiently large primes, K as above, and for all j, θ_j minimal such that

$$\begin{cases} \theta_j \equiv 2^{j-1} & \text{modulo } 2^n \\ \theta_j \equiv 0 & \text{modulo } \prod_{\substack{i=1 \\ i\neq j}}^{n} (p_i)^{n+1} \\ \theta_j \not\equiv 0 & \text{modulo } p_j \end{cases}$$

and

$$H = \sum \theta_j \ .$$

We then obtain

Lemma 3 (Conversion Lemma). For any $\alpha \in \omega$, i.e. $\alpha = \sum_{i=0}^{n-1} \alpha_i 2^i + \bar{\alpha} 2^n$, $\alpha_i \in \{0,1\}$, $\bar{\alpha} \in \omega$. The unique $a \in \omega$ satisfying

(i) $a < K$

(ii) $a \equiv \alpha \bmod 2^n$

(iii) $a(a-H) \equiv 0 \bmod K$

is

$$a = \sum_{j=1}^{n} \alpha_{j-1} \theta_j \ ,$$

α_j coefficients of binary representation of α above.

The proof of Lemma 3 is of course analogous to that of Lemma 1 above.

The crucial idea is to provide a means of going back and forth between the representation of a sequence as the digits of a number in some base b (e.g. 2), and as the residues of a number with respect to a system of relatively prime moduli. The first allows a global description of the shifting of the sequence, the second allows global formulation (i.e., for the whole sequence in a single congruence) of a condition on the individual elements of the sequence. All known reductions of recognition of correct Turing machine computation to Diophantine equations with a number of variables independent of input size provide some means of reconciling these very different kinds of operations on the sequence studied; doing so is the fundamental problem of such reductions and the principal challenge in obtaining tight bounds on the complexity of Diophantine decision problems.

(2) If $S \in NP \cap NP^C$, i.e. $S \in NP$ and $S^C \in NP$, then the reduction algorithm described in the proof of Theorem 1 reduces the questions

$$x \in S \ ?$$
$$x \in S^C \ ?$$

to the question of the solvability of two very closely related Diophantine equations. This fact might be of help in studying the class $NP \cap NP^C$. As an example, we suggest the

Open Problem 1. Is there a complete set in $NP \cap NP^C$ (i.e. a set in $NP \cap NP^C$ to which every set in $NP \cap NP^C$ can be reduced by a deterministic polynomial time algorithm)?

As a candidate we suggest the set S

$$S = \{\langle \alpha, \beta \rangle : \alpha, \beta \in \omega;\ \exists z \in \omega,\ 1 < z < \beta \text{ and } z | \alpha\} \ .$$

S is in $NP \cap NP^c$ and the problem of accepting S is computationally deterministic polynomial time equivalent to prime factorization of numbers, and hence probably not in P (see [13] for these assertions).

3. Nondeterministic Diophantine Machines

We introduce a class of nondeterministic Turing machines with restricted, purely numerical, computational ability: Given a multivariable polynomial $p(x_1,\ldots,x_m,y_1,\ldots,y_n)$ with integer coefficients, the corresponding nondeterministic Diophantine machine (NDDM) is a nondeterministic Turing machine with the following algorithm:

"On input $a_1,\ldots,a_m \in \omega$, guess $b_1,\ldots,b_n \in \omega$. If $p(a_1,\ldots,a_m,b_1,\ldots,b_n) = 0$ then accept $\langle a_1,\ldots,a_m\rangle$; otherwise halt without accepting."

For example, if $p(x_1,y_2) = x_1 - y_2^2$, then the corresponding NDDM has the algorithm

"On input $a_1 \in \omega$; guess $b_1 \in \omega$. If $a_1 - b_1^2 = 0$, accept a_1."

It is easy to see that this NDDM accepts exactly the set of perfect squares, and in polynomial time.

The question of major interest about NDDM's is whether computational ability of NDTM's is lost in restricting the operations to those of NDDM's. For clearly, the description of NDDM's and the relations on ω which they accept is more convenient and more directly number theoretical than is the case for Turing machines. Thus much more direct applicability of number theoretic techniques to computational theory (and vice versa, see for example [1], [2]) might be possible, if in fact general Turing computation is adequately represented by NDDM's.

The development of the theory of Diophantine definability in number theory, in connection with Hilbert's 10^{th} problem, has made it possible to

answer our question; for the Diophantine relations are exactly those accepted by an NDDM. Thus we have as a corollary to Matijasevic's theorem mentioned in the introduction, that the relations accepted by NDDM's are exactly the relations accepted by Turing machines.

Our question can now be sharpened: How does the efficiency of NDDM's compare with that of NDTM's? We show in [1], [3] that the restriction to the numerical operations of NDDM's causes at most an exponential loss in efficiency: A relation accepted in time t on a NDTM is accepted in at most time 2^{10t^2} on a NDDM (for a large class of arbitrarily large or small running time functions t). No lower bounds on the loss of efficiency are known; we conjecture that essentially no efficiency is lost. This is supported by evidence in [1], [3].

For close comparison of the efficiency of NDDM's and NDTM's we compare the class D of relations accepted by NDDM's in polynomial time to NP and P. The following definitions are useful:

(i) (alternative characterization of D): For all $n \in \omega$, D^n is the set of all numerical relations R definable by a formula of the form:

$$\langle x_1, x_2, \ldots, x_m \rangle \in R \Leftrightarrow \exists y_1, \ldots, y_n \leq 2^{q(|(x_1+x_2+\cdots+x_m)|)} [P(x_1, x_2, \ldots, x_m, y_1, \ldots, y_n) = 0]$$

where q and p are polynomials, and, as throughout this section, '$|x|$' denotes the length of x in binary.

Then we have:

$$D = \bigcup_{i \in \omega} D^i .$$

(ii) For any m-ary numerical relation R and any ℓ-ary numerical relation S: R *is D-reducible to* S (notation: $R \leq_D S$) if and only if R is

definable by a formula of the form:

$$\langle x_1,\ldots,x_m\rangle \in R \Leftrightarrow \exists y_1,\ldots,y_\ell,y_{\ell+1},\ldots,y_n \leq 2^{q(|x_1+\cdots+x_m|)} [P(x_1,\ldots,x_m,y_1,\ldots,y_n) = 0 \;\&\; \langle y_1,\ldots,y_\ell\rangle \in S]$$

where q and p are polynomials.

(iii) For any numerical relation R in NP: R is D-complete if and only if every other numerical relation in NP is D-reducible to R. Clearly if R is D-complete then $R \in D \Leftrightarrow NP = D$.

There seems to be considerable symmetry between D and P as subclasses of NP. For the complete problems in NP with respect to $\leq_D$ and $\leq_P$ (deterministic polynomial time many-one reducibility), this can be illustrated using the results of the previous section.

Theorem 3.

(a) There is an NP-complete problem in D (in fact even in D^2).

(b) There is a D-complete problem in P.

Corollary

(a) $D \subseteq P \Rightarrow P = NP$

(b) $P \subseteq D \Rightarrow D = NP$

Proof. (a) Let $p(x_1,x_2,x_3,y_1,y_2) = x_1y_1^2 + x_2y_2 - x_3$. Then the NDDM corresponding to $p(x_1x_2x_3,y_1y_2)$ accepts the relation S:

$$S = \{\langle\alpha,\beta,\gamma\rangle: \alpha,\beta,\gamma \in \omega,\ \exists y_1y_2 \in \omega[\alpha y_1^2 + \beta y_2 - \gamma = 0]\}$$

which is NP-complete as it is in deterministic polynomial-time 1-1 correspondence

with the set of Diophantine equations asserted to be NP-complete in Theorem 1; and the NDDM runs in polynomial time.

(b) We must find a relation $S \in P$ and for any NDTM M accepting a set in NP, a D-reducing Diophantine equation $p_M(\cdots) = 0$. We consider a nondeterministic simulation of an arbitrary NDTM M (when $M \in \omega$ will serve as the index of the NDTM) which will accept in nondeterministic polynomial time if M does. The computation involved in this simulation will be "divided" into a deterministic part and a nondeterministic part; the deterministic part will give us the definition of a relation S which will be in P; the nondeterministic part will give us the definition of the "reducing" NDDM, for it will be a Diophantine relation.

The Simulation Algorithm:

"For NDTM M, on input $x \in \omega$: Guess a time $t \in \omega$ (such that M on input x may halt within z steps); By Cook's algorithm [5], compute a propositional formula in disjunctive form with at most 3 literals per clause, satisfiable if and only if M on input x halts within t steps; By the reduction algorithm in the proof of Theorem 1, compute the appropriate $\alpha, \beta, \gamma \in \omega$; Guess x_1, x_2 (such that $\alpha x_1^2 + \beta x_2 - \gamma = 0$ may hold); If $\alpha x_1^2 + \beta x_2 - \gamma = 0$, accept x; Halt."

$S = \{\langle M,x,t,\alpha,\beta,\gamma\rangle$: $M,x,t,\alpha,\beta,\gamma \in \omega$, and α,β,γ are as computed in the simulation algorithm from $M,x,t\}$

$$p_M(x_1,y_1,\ldots,y_8) = (y_4y_7^2 + y_5y_8 - y_6)^2 + (y_1-M)^2 + (y_2-x_1)^2$$

Note that

$$p_M(\cdots) = 0 \Leftrightarrow \begin{cases} y_4y_7^2 + y_5y_8 - y_6 = 0 \\ y_1 = M \\ y_2 = x_1 \end{cases}$$

Now assume that M is a NDTM accepting a set $S_M \in NP$, i.e. M has a running-time function which is a polynomial in the length of the input. Then the Simulation Algorithm will accept S_M in polynomial time and needs only to guess numbers bounded by

$$2^{q_M(|x|)}$$

for some polynomial $q_M(\cdot)$ which could be determined from the running time of M by an analysis of the reduction algorithm used in proving Theorem 1 above. Thus we have, for all $x_1 \in \omega$:

$$x_1 \in S_M \Leftrightarrow \exists y_1,\ldots,y_8 \in \omega;\ y_1,\ldots,y_8 \leq 2^{q_M(|x_1|)} \text{ such that}$$

$$\begin{cases} p_M(x_1,y_1,\ldots,y_8) = 0 \\ \langle y_1,\ldots,y_6\rangle \in S \end{cases}$$

so that M is indeed D-reducible (by p_M) to $S \in P$. As M was arbitrary and S does not depend on M, S is D-complete. (End of proof.)

Theorem 3 answers an open problem in [2]: "Find a set $A \in D$ such that for all $B \subseteq \omega$, if $B \in D$, then B is polynomial reducible to A." That any such set would be NP-complete (as Theorem 3 implies) was rather unexpected. One would have expected the D^i, $i \in \omega$ to be a hierarchy of progressively harder problems (in the sense of P-reducibility); moreover, it was suspected that number-theoretic problems obviously in NP would be less than NP-complete: The deep structure of number theory should allow development of nontrivial and efficient algorithms for such problems. Theorem 3 indicates that all this is wrong.

The argument for Theorem 3(b) suggests that D and P are in a sense complementary subclasses of NP; it shows how all nondeterministic polynomial-

time computation can be decomposed into deterministic polynomial-time computation and NDDM-polynomial time computation, by a fixed DTM and an essentially fixed NDDM.

There are various natural subdivisions of the class D corresponding to characteristics of the defining polynomials of NDDM's computing sets in D: number of variables, degree in the variables, and the magnitude of the time bound. These suggest many problems of classification of sets and questions about whether the characteristics are significant for complexity theory. We now list several of the possibilities.

$D^{2(K)}$ will denote the subclass of D^2 where the relevant polynomial $P(x_1,\ldots,x_m,y_1,y_2)$ is of degree $\leq K$ in the variables y_1, y_2. Clearly, the set S in the proof of Theorem 3(a) is in $D^{2(2)}$. It then easily follows from Theorem 3(a) that

Theorem 4. The following are equivalent.

(a) $\bar{S} \in NP$

(b) $(D^{2(2)})^c \subseteq NP$

(c) $(NP)^c = NP$

where A^c denotes the set of complements of sets in A and $\bar{A}$ denotes the complement of A.

Theorem 4 is, in many respects, as strong as possible: We can show ([3]) that

(i) $(D^{2(1)})^c \subseteq P \subseteq NP$

(ii) $(D')^c \subseteq D \cap P \subseteq NP$

It is very interesting to consider the extent of $D^{2(2)}$.

Consider the following sets:

$$S_1 = \{\alpha \mid \exists x,y \in \omega: y^2 - \alpha x^2 = 1\}$$

$$S_2 = \{\alpha \mid \alpha \text{ composite}\}$$

$$S_3 = \{\langle\alpha,\beta\rangle \mid \exists z \in \omega: 1 < z < \beta \text{ and } z \text{ divides } \alpha\}$$

$$S_4 = \{\langle\alpha,\beta,\gamma\rangle \mid \exists y,z \in \omega: \alpha y^2 + \beta z - \gamma = 0\}$$

All of these are in $D^{2(2)}$. But also:

S_1 is in P (see [15]).

S_2 is in P, if the Extended Riemann Hypothesis is true [see [13], [14]).

S_3: As noted at the end of section 2, $S_3 \in NP \cap NP^c$; probably $S_3 \notin P$.

S_4 is NP-complete.

These examples illustrate that $D^{2(2)}$ is a microcosm of the principal subclasses of NP. This suggests that a more detailed determination of the extent of $D^{2(2)}$ would be valuable.

Open Problem 2.

(a) Show that $D^{2(2)} \neq NP$.

(b) Show that $P \not\subseteq D^{2(2)}$.

(Obviously, the second implies the first.)

(c) Does every degree in NP with respect to $\leq_p$ (deterministic polynomial time reducibility) have a representative in $D^{2(2)}$?

Open Problem 3. What is the relationship of $D \cap D^c$ to P? It is known [3] that $(D^1)^c \subset D$, so that $D^1 \subseteq D \cap D^c \cap P$.

Open Problem 4. What is the structure of NP under $\leq_D$ (D-reducibility)? If $D \neq NP$, are there D-(reducibility) degrees between 0 (the degree of sets in D) and the degree of the D-complete set of Theorem 3? See, for the corresponding assertion about P-reducibility, Ladner [8].

Open Problem 5. $D = NP$?

Open Problem 6. It can be shown by methods of algebraic topology that the set of composite numbers S_2, and its complement Pr, the set of primes, are not in D^1 ([17]). Hence $D^{2(2)} \neq D^1$.

(a) Can this or a different argument be generalized to show:

$$\text{for all } i \in \omega\text{: } D^i \neq D^{i+1} \text{ ?}$$

(b) In the case of the composites and the primes, it follows [17] from

$$\text{composites} \in D^2 \setminus D^1$$
$$\text{primes} = (\text{composites})^c$$

that primes $\notin D^1$. Is it always true for $S \subseteq \omega$ that

$$S \in D^{i+1} \setminus D^i \Rightarrow S^c \notin D^i \text{ ?}$$

Open Problem 7. (Further classification of the set Pr of prime numbers). By Pratt [16], Pr $\in$ NP; hence in $NP \cap NP^c$. We can now ask:

(a) Is Pr $\in$ D ?

(b) Is Pr D-complete?

Open Problem 8. (A different subdivision of D): For $k \geq 1$, let D(k) be the set of all numerical relations definable by a formula of the form

$$\langle x_1,\ldots,x_m\rangle \in R \Leftrightarrow \exists y_1\cdots y_n \leq 2^{c(|x_1+\cdots+x_m|)^k} : P(x_1\cdots x_m, y_1\cdots y_n) = 0$$

where $c > 0$ and P is a polynomial. If $k > \ell$, then $D(\ell) \subseteq D(k)$; $D(1)$ is just the class where the definition can be chosen as

$$\exists y_1\cdots y_n \leq q(x_1\cdots x_m): P(x_1\cdots x_m, y_1\cdots y_n) = 0 ,$$

p, q polynomials.

(a) The relation $x = y^z$ is in $D(2)$ [2]. Is it in $D(1)$?

(b) Are any of the inclusions $D(\ell) \subseteq D(k)$, for $k > \ell$, strict? An affirmative answer would follow from $D = NP$, by use of diagonalization over nondeterministic Turing machines running in time n^k. But this question could be independent of 'NP = D'.

Acknowledgment

The authors thank Bill Sakoda for proofreading a draft of this article.

References

1. L. Adleman, Number theoretic aspects of computational complexity, Ph.D. Dissertation, University of California, Berkeley, November 1976.
2. L. Adleman and K. Manders, The computational complexity of decision procedures for polynomials, *16th Annual IEEE Symp. on Foundations of Computer Science* (1975), 169-177.
3. L. Adleman and K. Manders, "Diophantine complexity, *17th Annual IEEE Symp. on Foundations of Computer Science* (1976).
4. A. Baker, Contributions to the theory of Diophantine equations. I. On the representation of integers by binary forms, *Philos. Trans. Roy. Soc. London Ser. A* 263 (1967/68), 173-191.
5. S.A. Cook, The complexity of theorem-proving procedures, *Conf. Rec. 3rd ACM Symp. on Theory of Computing* (1971), 151-158.
6. D. Hilbert, Mathematische Probleme: Vortrag gehalten auf dem internationalen Mathematiker-Kongress zu Paris, 1900. *Nachrichten Akad. Wiss. Göttingen, Math.-Phys. Kl.* (1900) 253-297.
7. R.M. Karp, Reducibility among combinatorial problems, *in:* "Complexity of Computer Computation," eds. R.N. Miller and J.W. Thatcher, Plenum Press, 1972, 85-104.
8. R. Ladner, On the structure of polynomial time reducibility, *J. ACM* 22, 1 (Jan. 1975).
9. M. Lamé, Note sur la limite du nombre des divisions dans la recherche du plus grand common diviseur..., *Comptes Rendus, Académie des Sciences, Paris* 19 (1844), 867-869.
10. Y. Matijasevic, Enumerable sets are Diophantine (Russian), *Dokl. Akad. Nauk SSSR* 191 (1970), 279-282.

11. Y. Matijasevic and J. Robinson, Reduction of an arbitrary Diophantine equation to one in 13 unknowns, *Acta Arithemtica* 27 (1975) 521-553.

12. Y. Matijasevic, Private communication.

13. G.L. Miller, Ph.D. Dissertation, University of California, Berkeley, 1975.

14. G.L. Miller, Riemann's hypothesis and tests for primality, *7th ACM Symp. on Theory of Computing* (1975), 234-239.

15. I. Niven and H. Zuckerman, "An Introduction to the Theory of Numbers," John Wiley and Sons, Inc., 1972.

16. V. Pratt, Succinct certificates for primes, to appear.

17. D. Sato, Private communication.

18. A. Thue, Uber Annäherungs werte algebraischer Zahlen. *J. reine angewandte Math.* 135 (1909), 284-305.

11. Y. Matijasevic and J. Robinson, Reduction of an arbitrary Diophantine equation to one in 13 unknowns, Acta Arithmetica 27 (1975) 521-553.

12. Yu. Matijasevic, Private communication.

13. G.L. Miller, Ph.D. Dissertation, University of California, Berkeley, 1975.

14. G.L. Miller, Riemann's hypothesis and tests for primality, 7th Ann. ACM Symp. on Theory of Computing (1975), 234-239.

15. I. Niven and H. Zuckerman, "An Introduction to the Theory of Numbers", John Wiley and Sons, Inc., 1972.

16. V. Pratt, Succinct certificates for primes, to appear.

17. D. Sato, Private communication.

18. A. Thue, Über Annäherungswerte algebraischer Zahlen, J. reine angew. Math. 135 (1909), 284-305.