SIAM J. DISCRETE MATH.
Vol. 7, No. 2, pp. 275–283, May 1994

© 1994 Society for Industrial and Applied Mathematics
013

THE MINIMUM SATISFIABILITY PROBLEM*

RAJEEV KOHLI†, RAMESH KRISHNAMURTI‡, AND PRAKASH MIRCHANDANI§

Abstract. This paper shows that a minimization version of satisfiability is strongly NP-hard, even if each clause contains no more than two literals and/or each clause contains at most one unnegated variable. The worst-case and average-case performances of greedy and probabilistic greedy heuristics for the problem are examined, and tight upper bounds on the performance ratio in each case are developed.

Key words. minimum satisfiability, satisfiability, heuristics, probabilistic algorithms, average performance, Horn formulae

AMS subject classification. 68Q25

1. Introduction.

The satisfiability problem is perhaps one of the most well-studied problems in logic theory. Given a set V of Boolean (true/false) variables and a collection D of clauses over V, the satisfiability problem is to determine if there is a truth assignment that satisfies all clauses in D. The problem is NP-complete even when every clause in D has at most three literals (Even, Itai, and Shamir [2]). The maximum satisfiability (MAXSAT) problem is an optimization version of satisfiability that seeks a truth assignment to maximize the number of satisfied clauses (Johnson [5]). The MAXSAT problem is NP-hard even when every clause contains at most two literals (Garey, Johnson, and Stockmeyer [4]).

In this paper, we consider the following complement of the MAXSAT problem. Given a set U of Boolean variables and a collection C of clauses over U, find a truth assignment that minimizes the number of satisfied clauses. We call this the minimum satisfiability (MINSAT) problem. The existence of a truth assignment for the MINSAT problem that satisfies no clause can be trivially determined because such an assignment exists only if each variable or its negation appears in no clause. Similarly, if each clause contains one literal, the solution to the MINSAT problem is readily obtained by setting a variable true if it occurs in less clauses than its negation and setting the variable false otherwise. However, we show that, in general, the MINSAT problem is NP-hard, even if every clause contains no more than two literals. We then consider two heuristics for solving the problem. The first is a greedy heuristic similar to a procedure described by Johnson [5] for the MAXSAT problem. The second is a probabilistic greedy heuristic similar to an algorithm by Kohli and Krishnamurti [6] for the MAXSAT problem. Like the greedy heuristic, the probabilistic greedy heuristic selects a truth assignment one variable at a time. Unlike the greedy heuristic, the probabilistic greedy heuristic introduces a chance element in selecting a truth assignment, forcing a trade-off between the value of a nonoptimal solution and the probability of its selection. We characterize the worst-case and average-case performances of the two heuristics and show that, while the probabilistic greedy heuristic can select an arbitrarily bad assignment in the worst case, on average it satisfies no more than twice the optimal number of clauses, regardless of the data-generating distribution. On the other hand, if each clause contains at most s literals, the greedy heuristic satisfies no more than s times the number of clauses satisfied by the optimal assignment. However, the average performance of the greedy heuristic depends

* Received by the editors August 13, 1991; accepted for publication (in revised form) May 11, 1993.

† Graduate School of Business, Columbia University, 506 Uris Hall, New York, New York 10027.

‡ School of Computing Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.
This author was supported by Natural Sciences and Engineering Research Council of Canada grant OGP36809.

§ Joseph M. Katz School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

upon the minimum probability with which it selects an optimal assignment at any step. As this probability decreases (increases), the average performance of the greedy heuristic tends to the worst-case (optimal) solution.

In §2 we show that the MINSAT problem is NP-hard. In §3 we formally describe the greedy heuristic and analyze its worst-case and average-case performances. In §4 we analyze the average-case performance of the probabilistic greedy heuristic. We derive tight upper bounds on the performance ratio in each case. In §5 we extend our results to Horn formulae, in which each clause has no more than one unnegated variable.

2. Complexity. We show in Theorem 1 that the following decision problem, called the 2-MINSAT problem, in which each clause contains exactly two literals, is NP-complete. It follows that the MINSAT problem is NP-hard if each clause contains at least two literals.

2-MINSAT.

Instance. Set U of k variables, collection C of n clauses over U such that each clause $c \in C$ has $|c| = 2$ literals, positive integer $n^* \leq n$.

Question. Is there a truth assignment for U that satisfies no more than n^* clauses in C?

In Theorem 1, we transform the following 2-MAXSAT problem, which is NP-complete (Garey and Johnson [3, pp. 259–260]), to the 2-MINSAT problem.

2-MAXSAT.

Instance. Set V of h variables, collection D of l clauses over V such that each clause $d \in D$ has $|d| = 2$ literals, positive integer $l^* \leq l$.

Question. Is there a truth assignment for V that satisfies at least l^* clauses in D?

THEOREM 1. *The* 2-MINSAT *problem is* NP-*complete.*

Proof. Given a *yes* instance of the 2-MINSAT problem, we can simply count the number of satisfied clauses and verify a *yes* instance in polynomial time. Hence the 2-MINSAT problem is in NP. To show that it is NP-complete, we transform the 2-MAXSAT problem to the 2-MINSAT problem as follows.

Let $d = q_a \vee q_b$ be a clause in an instance of the 2-MAXSAT problem, where q_a and q_b denote either variables in V or their negations. For each clause $d \in D$, define a variable w_d. Let $W = \{w_d | d \in D\}$. For each clause $d = q_a \vee q_b$ of 2-MAXSAT, define a pair of clauses $c_{1d}, c_{2d} \in C$ for 2-MINSAT, where

$$c_{1d} = \bar{q}_a \vee w_d, \qquad c_{2d} = \bar{q}_b \vee \bar{w}_d.$$

Let $C = \{c_{1d}, c_{2d} | d \in D\}$. *Let* $n^* = 2l - l^*$. Thus, given an instance of the 2-MAXSAT problem defined over the set V of h variables and the set D of l clauses, we construct in polynomial time an instance of the 2-MINSAT problem defined over the set $U = V \cup W$ of $k = h + l$ variables and the set C of $n = 2l$ clauses. We now show that no less than l^* clauses can be satisfied by a truth assignment for 2-MAXSAT if and only if no more than n^* clauses can be satisfied by a truth assignment for 2-MINSAT.

Suppose there exists a truth assignment for 2-MAXSAT that satisfies $m \geq l^*$ clauses. Clause $d = q_a \vee q_b$ is not satisfied by an assignment if and only if both q_a and q_b are false, in which case, both c_{1d} and c_{2d} are satisfied. On the other hand, clause d is satisfied if and only if at least one of q_a or q_b is true. If both q_a and q_b are true, then any truth assignment for w_d satisfies exactly one of c_{1d} and c_{2d}. If q_a is true and q_b is false, then a false assignment for w_d satisfies c_{2d} but not c_{1d}. Similarly, if q_a is false and q_b is true, then a true assignment for w_d satisfies c_{1d} but not c_{2d}. Thus, if clause d is satisfied, a suitable truth assignment for w_d ensures that only one of c_{1d} or c_{2d} is satisfied. Consequently, if a truth assignment for 2-MAXSAT satisfies exactly $m \geq l^*$ clauses, then a suitable truth assignment for each $w_d \in W$ ensures that exactly $m + 2(l - m) = 2l - m \le 2l - l^* = n^*$ clauses are satisfied in the corresponding instance of 2-MINSAT. Hence every *yes* instance of the 2-MAXSAT problem corresponds to a *yes* instance of the 2-MINSAT problem.

Now suppose that a truth assignment for the 2-MINSAT problem satisfies no more than $n^* = 2l - l^*$ clauses. Consider the pair of clauses $c_{1d} = \bar{q}_a \vee w_d$ and $c_{2d} = \bar{q}_b \vee \bar{w}_d$. Any assignment of w_d ensures that at least one of these two clauses is satisfied. Let y denote the number of pairs $c_{1d}, c_{2d}, d \in D$, where exactly one clause in the pair is satisfied by a given assignment for the 2-MINSAT problem. Without loss of generality, let $c_{1d} = \bar{q}_a \vee w_d$ be the clause that is not satisfied. Then q_a must be true, which implies that clause $d = q_a \vee q_b$ must be satisfied. Hence, at least y clauses in D must be satisfied. Since there are l pairs c_{1d}, c_{2d}, there are $l - y$ such pairs in which both clauses are satisfied. Thus, the number of clauses in C that are satisfied is $2(l - y) + y \le n^* = 2l - l^*$, which implies that $y \ge l^*$. Thus, if $n^* = 2l - l^*$ clauses in C are satisfied for the 2-MINSAT problem, then at least l^* clauses in D are satisfied for the 2-MAXSAT problem. Hence, every *yes* instance of the 2-MINSAT problem corresponds to a *yes* instance of the 2-MAXSAT problem. □

3. The greedy heuristic. Let $u_1, u_2, \ldots, u_k$ denote an arbitrary ordering of the k variables in U for the MINSAT problem. Given any ordering of the variables, the greedy heuristic sequentially selects an assignment for each variable to satisfy the smallest number of additional clauses. We begin by describing the greedy heuristic more formally below.

Initialization (*Step* 1). Let $C_1 = C$ denote the set of all clauses in an instance of the MINSAT problem. Let $C_1(u_1)$ denote the subset of clauses in C_1 that contain variable u_1. Let $C_1(\bar{u}_1)$ denote the subset of clauses in C_1 that contain variable $\bar{u}_1$. Let x_1 and y_1 denote the number of clauses in sets $C_1(u_1)$ and $C_1(\bar{u}_1)$. At the first step, the greedy heuristic selects the partial assignment u_1 (i.e., assigns u_1 to be true) if $x_1 < y_1$. Otherwise, it selects the partial assignment $\bar{u}_1$ (i.e., assigns u_1 to be false).[1] All clauses satisfied by the partial assignment are eliminated. Let C_2 denote the set of clauses not satisfied at the end of Step 1. Thus,

$$C_2 = \begin{cases} C_1 \backslash C_1(u_1) & \text{if } u_1 \text{ is selected at Step 1,} \\ C_1 \backslash C_1(\bar{u}_1) & \text{if } \bar{u}_1 \text{ is selected at Step 1.} \end{cases}$$

Recursion (*Step j*). Let C_j denote the set of clauses that are not satisfied at the end of Step $j - 1$. Let $C_j(u_j)$ denote the subset of clauses in C_j that contain u_j. Let $C_j(\bar{u}_j)$ denote the subset of clauses in C_j that contain $\bar{u}_j$. Let x_j and y_j denote the number of clauses in sets $C_j(u_j)$ and $C_j(\bar{u}_j)$. At Step j, the greedy heuristic includes u_j in the partial assignment (i.e., assigns u_j to be true) if $x_j < y_j$. Otherwise, it includes $\bar{u}_j$ in the partial assignment (i.e., assigns u_j to be false). All clauses satisfied by the partial assignment are eliminated. Let C_{j+1} denote the set of clauses not satisfied at the end of Step j. Thus,

$$C_{j+1} = \begin{cases} C_j \backslash C_j(u_j) & \text{if } u_j \text{ is selected at Step } j, \\ C_j \backslash C_j(\bar{u}_j) & \text{if } \bar{u}_j \text{ is selected at Step } j. \end{cases}$$

Termination Step. Stop if $C_{j+1} = \phi$ or if $j = k$.

Let $|c_i|$ denote the number of literals in clause c_i. Let $s = \max_i |c_i|$ denote the maximum number of literals in any clause. Let r denote the performance ratio for the greedy heuristic, i.e., the ratio of the number of clauses satisfied by the assignment selected

[1] Note that we assign u_j to be false if $x_j = y_j$. This simplifies the subsequent worst-case analysis of the heuristic, where we assume, without loss of generality, that the optimal solution is to set each variable true.

by the greedy heuristic to the number of clauses satisfied by an optimal assignment. Theorem 2 shows that the value of the performance ratio r is bounded from above by s. As there are no more than k literals, $s \le k$, and it follows trivially from Theorem 2 that $r \le k$.

THEOREM 2. *$r \le s$ for the greedy heuristic.*

Proof. Without loss of generality, let each variable $u_j \in U$ be true in the optimal assignment. Let $C_j^* \subseteq C$ denote the subset of clauses in C that are satisfied if variable u_j is true. Let $C^* = \bigcup_{j=1}^{k} C_j^*$ denote the subset of clauses satisfied by the optimal assignment. Let m_j denote the number of clauses in the set C_j^*. Let m denote the number of clauses in set C^*. At Step j, the greedy heuristic satisfies $z_j = \min\{x_j, y_j\}$ clauses in set C_j. Hence the total number of clauses satisfied by the greedy heuristic is $\sum_{j=1}^{k} z_j$. Now $m_j \ge x_j$, which implies that

$$\sum_{j=1}^{k} z_j \le \sum_{j=1}^{k} x_j \le \sum_{j=1}^{k} m_j.$$

Also,

$$\sum_{j=1}^{k} m_j = \sum_{c_i \in C^*} |c_i| \le ms.$$

Thus, an upper bound on the performance ratio of the greedy heuristic is

$$r = \frac{\sum_{j=1}^{k} z_j}{m} \le \frac{ms}{m} = s. \qquad \square$$

To prove that the bound in Theorem 2 is tight, consider the following problem instance in which there are s variables and $s + 1$ clauses:

$$c_1 = u_1 \vee u_2 \vee \cdots \vee u_s,$$
$$c_i = \bar{u}_{i-1} \quad \text{for } 2 \le i \le s+1.$$

Each variable u_j, $1 \le j \le s$ is true in the optimal assignment, which satisfies only one clause c_1. The greedy heuristic sets each variable u_j false, $1 \le j \le s$ and satisfies s clauses, $c_2, c_3, \ldots, c_{s+1}$. Hence, $r = s$. Note that, in this worst-case example, a reordering of the variables has no effect on the performance of the greedy heuristic. This example also suffices to show that, given the solution selected by the greedy heuristic, an interchange heuristic that seeks to maximally improve the solution value by replacing a literal by its negation does no better than the greedy heuristic alone.

In [5] Johnson suggests weighted greedy heuristics for the MAXSAT problem, the simplest of which ensures a worst-case error of $1/2^s$ when each clause has no less than s literals. An analogous weighted greedy heuristic for the MINSAT problem is as follows. At Step j, the greedy heuristic assigns

$$u_j \text{ true} \quad \text{if } \sum_{c_i \in C_j(u_j)} |c_i| > \sum_{c_i \in C_j(\bar{u}_j)} |c_i|,$$
$$u_j \text{ false} \quad \text{otherwise.}$$

The intuition behind the weighting is that a literal should be selected if the unsatisfied clauses that contain its negation are less likely to be satisfied by subsequent assignments to variables. However, this weighting scheme does not improve the worst-case performance of the greedy heuristic for the MINSAT problem. To illustrate, consider the following

$s + 1$ clauses defined over $2s - 1$ variables u_j, $1 \le j \le 2s - 1$:

$$c_1 = u_1 \vee u_2 \vee \cdots \vee u_s,$$
$$c_i = \bar{u}_{i-1} \vee \bar{u}_{s+1} \vee \bar{u}_{s+2} \vee \cdots \vee \bar{u}_{2s-1}, \qquad 2 \le i \le s + 1.$$

Each variable u_j, $1 \le j \le 2s - 1$ is true in the optimal assignment, which satisfies only one clause, c_1. At each step j of the greedy heuristic, an equal number of clauses are satisfied by setting u_j true or false. Regardless of the truth assignment for variable u_j at Step j, there are an equal number of literals in each remaining (unsatisfied) clause. Consequently, any weighting of the clauses does not affect the greedy solution, which in the worst case assigns a "false" value to each variable u_j, $1 \le j \le 2s - 1$ and satisfies the s clauses $c_2, c_3, \ldots, c_{s+1}$. Note, however, that the ordering of the variables is critical to this worst-case example. Thus, the performance ratio for this weighted greedy heuristic for the MINSAT problem is no better than s. Indeed, other weighting schemes (e.g., a scheme similar to Johnson's [5] exponential weighting of clauses) also do not improve the worst-case performance of the greedy heuristic.

We next examine the average performance of the greedy heuristic. Let P_k denote the MINSAT problem defined over the set U of k variables. Let P_{k-j+1} denote the MINSAT problem at Step j of the greedy heuristic, where problem P_{k-j+1} is defined over the subset of unassigned variables $u_j, u_{j+1}, \ldots, u_k$. Following Kohli and Krishnamurti [6], we assume that there is a probability p_j with which the truth assignment selected for a variable at step j of the greedy heuristic appears in an optimal truth assignment for problem P_{k-j+1}. We assume that the p_j are independent across the steps of the greedy heuristic. However, we do not assume either the independence of the literals across clauses, or any specific data-generating distribution.

Let r_j denote the performance ratio of the greedy heuristic for problem P_j. Let $E(r_j)$ denote the expected value of r_j. Theorem 3 characterizes the lower bound on $E(r_k)$ as a function of k and $p = \min_{1 \le j \le k} p_j$.

THEOREM 3. *$E(r_k) \le 1 - (1 - p)^k/p$ for the greedy heuristic.*

Proof. Without loss of generality, we assume that the optimal solution is to set variable u_j true for all j, $1 \le j \le k$. We prove the theorem by induction on k, the number of variables.

For $k = 1$, the greedy heuristic chooses the optimal assignment and sets u_1 true. Hence, $p = p_1 = 1$ and $E(r_k) = 1$.

Let $l \ge 1$ be an integer such that

$$E(r_k) \le \frac{1 - (1 - p)^k}{p} \quad \text{for } k = l.$$

We show below that

$$E(r_k) \le \frac{1 - (1 - p)^k}{p}$$

for $k = l + 1$.

Let $z_1 = \min\{x_1, y_1\}$. If the greedy heuristic sets u_1 true at Step 1, the value of the optimal solution to problem P_{k-1} is $m - z_1$, where m is the value of the optimal solution to problem P_k. However, if the greedy heuristic sets u_1 false at Step 1, the value of the optimal solution to problem P_{k-1} is bounded from above by m. In either case, after Step

1, the greedy heuristic solves an l variable MINSAT problem. Thus,

$$E(r_{l+1}) \leq \frac{1}{m}(z_1 + p_1 E(r_l)(m - z_1) + (1 - p_1)E(r_l)m)$$

$$= \frac{z_1}{m}(1 - p_1 E(r_l)) + E(r_l) \leq 1 + E(r_l)(1 - p_1).$$

Let $p^* = \min_{i \geq 2} p_i$. By the induction hypothesis,

$$E(r_l) \leq \frac{1 - (1 - p^*)^l}{p^*}.$$

Thus,

$$E(r_{l+1}) \leq 1 + \left(\frac{1 - (1 - p^*)^l}{p^*}\right)(1 - p_1).$$

Let $p = \min\{p^*, p_1\}$. Then

$$\frac{1 - (1 - p^*)^l}{p^*} \leq \frac{1 - (1 - p)^l}{p} \quad \text{and} \quad 1 - p_1 \leq 1 - p,$$

which implies that

$$E(r_{l+1}) \leq 1 + \left(\frac{1 - (1 - p)^l}{p}\right)(1 - p) = \frac{1 - (1 - p)^{l+1}}{p}. \qquad \square$$

Note that the bound derived in Theorem 3 approaches 1 as p approaches 1 and that it approaches k as p approaches zero. As k tends to infinity, the bound on the average performance ratio for the greedy heuristic approaches $1/p$.

To prove that the bound derived in Theorem 3 is tight, consider the following example with k variables and $n = (k + 1)N + k$ clauses. The first $(k + 1)N$ clauses are

$$c_i = \begin{cases} u_1 \vee u_2 \vee \cdots \vee u_k & \text{for } 1 \leq i \leq N \\ \bar{u}_1 & \text{for } N + 1 \leq i \leq 2N, \\ \bar{u}_2 & \text{for } 2N + 1 \leq i \leq 3N, \\ \vdots & \\ \bar{u}_k & \text{for } kN + 1 \leq i \leq (k + 1)N. \end{cases}$$

The remaining k clauses are probabilistically generated, each clause containing exactly one of the k distinct variables in negated form with probability p and in unnegated form with probability $1 - p$. Specifically, clause j, $1 \leq j \leq k$ is given by

$$c_{(k+1)N+j} = \begin{cases} \bar{u}_j & \text{with probability } p, \\ u_j & \text{with probability } 1 - p. \end{cases}$$

Each variable u_j, $1 \leq j \leq k$ is true in the optimal assignment. For $N \gg k$, the expected performance ratio of the greedy heuristic can be verified to approach from below the value

$$p + 2(1 - p)p + 3(1 - p)^2 p + \cdots + (k - 1)(1 - p)^{k-2}p + k(1 - p)^{k-1} = \frac{1 - (1 - p)^k}{p}.$$

Observe that the bound on the average performance of the greedy heuristic depends upon the value of p, which can vary, depending upon the data-generating distribution. If, as in the above example, the value of p can be made close to zero, the average performance of the greedy heuristic can be made to approach its deterministic worst-case bound. In the next section, we examine a probabilistic greedy heuristic that, regardless of the data-generating distribution, never satisfies more than twice the number of clauses satisfied by an optimal assignment for the MINSAT problem.

4. Probabilistic greedy heuristic. The proposed probabilistic greedy heuristic differs from the preceding greedy heuristic by the introduction of a probabilistic element in the choice of a truth assignment for each variable. In particular, at Step j, the probabilistic greedy heuristic sets u_j to be true with probability $q_j = y_j/(x_j + y_j)$ and sets u_j to be false with probability $1 - q_j$. Thus, the probability of setting u_j true increases as x_j/y_j decreases and is 1 only if $x_j = 0$, i.e., if no additional clauses are satisfied by setting u_j true at Step j of the heuristic. However, if the number of additional clauses satisfied is greater when u_j is true than when it is false, then u_j is set true with a smaller probability than it is set false.

The following theorem shows that, on average, the number of clauses satisfied by the probabilistic greedy heuristic is no larger than twice the number of clauses satisfied by the optimal assignment.

THEOREM 4. *$E(r_k) \leq 2$ for the probabilistic greedy heuristic.*

Proof. Without loss of generality, assume that variable u_1 is true in an optimal assignment. We prove the theorem by induction on the number of variables k.

For $k = 1$, the greedy heuristic sets u_1 true with probability $q_1 = y_1/(x_1 + y_1)$ and sets u_1 false with probability $1 - q_1$. The expected number of satisfied clauses is

$$q_1 x_1 + (1 - q_1)y_1 = \frac{y_1}{x_1 + y_1} x_1 + \frac{x_1}{x_1 + y_1} y_1 = \frac{2x_1 y_1}{x_1 + y_1}.$$

As u_1 is true in the optimal assignment, the value of the optimal solution is $m = x_1$. Thus, for $k = 1$, the value of the expected performance ratio for the probabilistic greedy heuristic is

$$E(r_k) = \frac{2x_1 y_1}{x_1(x_1 + y_1)} = \frac{2y_1}{x_1 + y_1} \leq 2.$$

Let $l \geq 1$ be an integer such that

$$E(r_k) \leq 2 \quad \text{for } k = l.$$

We show that

$$E(r_k) \leq 2 \quad \text{for } k = l + 1.$$

If the probabilistic greedy heuristic selects u_1 at Step 1, the value of the optimal solution at the second step of the greedy heuristic is $m - x_1$, where m is the optimal solution value of the k variable MINSAT problem P_k. However, if the greedy heuristic selects $\bar{u}_1$ at Step 1, the value of the optimal solution at the second step is bounded from above by m. Hence, the expected number of clauses satisfied by the probabilistic greedy heuristic is bounded from above by

$$q_1(x_1 + E(r_l)(m - x_1)) + (1 - q_1)(y_1 + E(r_l)m).$$

As $E(r_l) \leq 2$ by the induction hypothesis, the value of the above expression is no greater than

$$q_1(x_1 + 2(m - x_1)) + (1 - q_1)(y_1 + 2m).$$

Thus, an upper bound on the expected performance ratio for the probabilistic greedy heuristic is

$$E(r_{l+1}) \le \frac{1}{m}\,(q_1(x_1 + 2(m - x_1)) + (1 - q_1)(y_1 + 2m))$$

$$= \frac{1}{m}\,(2m - q_1(x_1 + y_1) + y_1) = 2. \qquad \square$$

To prove the above bound is tight, consider the following example with $k = 2$ variables and $n = N + 1$ clauses. Let

$$c_1 = u_1 \vee u_2,$$
$$c_2 = \bar{u}_1,$$
$$c_i = \bar{u}_2, \qquad 3 \le i \le N + 1.$$

The optimal assignment sets both u_1 and u_2 true and satisfies one clause, c_1. The probabilistic greedy heuristic sets u_1 true or false with equal probability ($= \frac{1}{2}$) at its first step. If it sets u_1 true, then it obtains the optimal solution, setting u_2 true with probability 1 at its second step. Otherwise, at the second step, it

(i) sets u_2 true with probability $(N - 1)/N$, satisfying 2 clauses, c_1 and c_2, and

(ii) sets u_2 false with probability $1/N$, satisfying N clauses, $c_2, c_3, c_4, \ldots, c_{N+1}$.

Hence the expected performance ratio (= expected number of satisfied clauses) for the probabilistic greedy heuristic is

$$\frac{1}{2}\cdot 1 + \frac{1}{2}\left(\frac{N-1}{N}\cdot 2 + \frac{1}{N}\cdot N\right) = 2 - \frac{1}{N}.$$

As N tends to infinity, the value of this expression approaches from below the bound derived in Theorem 4. Note that $k = 2$ in this example and that the optimal clause c_1 contains $s = 2$ variables. Thus, unlike the worst-case and average performance bound for the (deterministic) greedy heuristic, the bound on the average performance of the probabilistic greedy heuristic does not depend on k or s.

5. Horn clauses. An important special case of the satisfiability problem occurs when each clause contains no more than one unnegated variable. Such clauses are called *Horn clauses*. The satisfiability problem defined over a set of Horn clauses can be solved in linear time (Dowling and Gallier [1]).

We show below that the MINSAT problem continues to be NP-hard even if it is restricted to a set of Horn clauses. In particular, we transform the 2-MINSAT problem to a MINSAT problem in Horn clauses.

Let the 2-MINSAT problem be defined over the set V of h variables and the set D of l clauses. We transform this to a MINSAT problem in Horn clauses defined over a set U of $k = h + l$ variables and a set C of $n = 3l$ clauses. Let $d = q_a \vee q_b$ be a clause in an instance of the 2-MINSAT problem, where q_a and q_b denote either variables in V or their negations. For each clause $d \in D$, define a variable w_d. Let $W = \{w_d \mid d \in D\}$. The set U is defined to be $V \cup W$. For each clause $d = q_a \vee q_b$ of 2-MINSAT, define three Horn clauses c_{1d}, $c_{2\mathrm{d}}$, and $c_{3d} \in C$, where

$$c_{1d} = q_a \vee \bar{w}_d, \quad c_{2d} = q_b \vee \bar{w}_d, \quad c_{3d} = w_d.$$

If clause d is satisfied by a truth assignment, then the same truth assignment for variables q_a and q_b, and a suitable truth assignment for variable w_d, satisfies two (the minimum that must be satisfied) of the above three clauses. If clause d is not satisfied by a truth assignment, the same truth assignment for variables q_a and q_b, and a suitable truth assignment for variable w_d satisfies one (the minimum that must be satisfied) of the above three clauses. Based on the above observations and using an argument similar to that in Theorem 1, we can show that no more than l^* clauses can be satisfied by some truth assignment for 2-MINSAT if and only if no more than $l + l^*$ clauses can be satisfied by a truth assignment for the MINSAT problem in Horn clauses.

We can verify that the worst-case and average-case bounds for the greedy heuristic and the probabilistic greedy heuristic remain unchanged for Horn clauses and that these bounds continue to be tight. Finally, we note that the MAXSAT problem in Horn clauses is also NP-hard (see the Appendix). Thus, while the satisfiability of Horn clauses can be assessed in linear time, the identification of assignments that maximize or minimize the number of Horn clauses satisfied are NP-hard problems.

Appendix. Complexity of MAXSAT for Horn clauses. We show below that the MAXSAT problem comprised of only Horn clauses is NP-hard. In particular, we transform the 2-MAXSAT problem to a MAXSAT problem in Horn clauses.

Let the 2-MAXSAT problem be defined over the set V of h variables and the set D of l clauses. We transform this problem to a MAXSAT problem in Horn clauses, defined over a set U of $k = h + 2l$ variables and a set C of $n = 5l$ clauses. Let $d = q_a \vee q_b$ be a clause in an instance of the 2-MAXSAT problem, where q_a and q_b denote either variables in V or their negations. For each clause $d \in D$, define two new variables w_{1d}, w_{2d}. Let $W = \{w_{1d}, w_{2d} | d \in D\}$. The set U is defined to be $V \cup W$. For each clause $d = q_a \vee q_b$ of 2-MAXSAT, define five Horn clauses $c_{1d}, c_{2d}, \ldots, c_{5d} \in C$, where

$$c_{1d} = q_a \vee \bar{w}_{1d},$$

$$c_{2d} = q_b \vee \bar{w}_{2d},$$

$$c_{3d} = \bar{w}_{1d} \vee \bar{w}_{2d},$$

$$c_{4d} = w_{1d},$$

$$c_{5d} = w_{2d}.$$

If clause d is satisfied by a truth assignment, the same truth assignment for variables q_a and q_b and a suitable truth assignment for variables w_{1d} and w_{2d} satisfy four (the maximum that can be satisfied) of the above five clauses in MAXSAT. If clause d is not satisfied by a truth assignment, the same truth assignment for variables q_a and q_b and a suitable truth assignment for variables w_{1d} and w_{2d} satisfy three (the maximum that can be satisfied) of the above five clauses. Based on the above observations and using an argument similar to that in Theorem 1, we can now show that no less than l^* clauses can be satisfied by some truth assignment for 2-MAXSAT if and only if no less than $3l + l^*$ clauses can be satisfied by a truth assignment for the MAXSAT problem in Horn clauses.

REFERENCES

[1] W. F. DOWLING AND J. H. GALLIER, *Linear-time algorithms for testing the satisfiability of propositional Horn formulae*, J. Logic Programming, 3 (1984), pp. 267–284.

[2] S. EVEN, A. ITAI, AND A. SHAMIR, *On the complexity of timetable and multicommodity flow problems*, SIAM J. Comput., 5 (1976), pp. 691–703.

[3] M. R. GAREY AND D. S. JOHNSON, *Computers and Intractibility: A Guide to the Theory of* NP-*Completeness*, W. H. Freeman, San Francisco, CA, 1979.

[4] M. R. GAREY, D. S. JOHNSON, AND L. STOCKMEYER, *Some simplified* NP-*complete graph problems*, Theoret. Comput. Sci., 1 (1976), pp. 237–267.

[5] D. S. JOHNSON, *Approximation algorithms for combinatorial problems*, J. Comput. System Sci., 9 (1974), pp. 256–278.

[6] R. KOHLI AND R. KRISHNAMURTI, *Average performance of heuristics for satisfiability*, SIAM J. Discrete Math., 2 (1989), pp. 508–523.