View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

ELSEVIER

Theoretical Computer Science 209 (1998) 237–260

Theoretical Computer Science

On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems

Edoardo Amaldi[a,*,1], Viggo Kann[b]

[a] *School of Operations Research and Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, USA*

[b] *Department of Numerical Analysis and Computing Science, Royal Institute of Technology S-100 44 Stockholm, Sweden*

Received December 1995; received in revised form April 1997

Communicated by G. Ausiello

Abstract

We investigate the computational complexity of two closely related classes of combinatorial optimization problems for linear systems which arise in various fields such as machine learning, operations research and pattern recognition. In the first class (MIN ULR) one wishes, given a possibly infeasible system of linear relations, to find a solution that violates as few relations as possible while satisfying all the others. In the second class (MIN RVLS) the linear system is supposed to be feasible and one looks for a solution with as few nonzero variables as possible. For both MIN ULR and MIN RVLS the four basic types of relational operators $=$, $\geqslant$, $>$ and $\neq$ are considered. While MIN RVLS with equations was mentioned to be NP-hard in (Garey and Johnson, 1979), we established in (Amaldi; 1992; Amaldi and Kann, 1995) that MIN ULR with equalities and inequalities are NP-hard even when restricted to homogeneous systems with bipolar coefficients. The latter problems have been shown hard to approximate in (Arora et al., 1993). In this paper we determine strong bounds on the approximability of various variants of MIN RVLS and MIN ULR, including constrained ones where the variables are restricted to take binary values or where some relations are mandatory while others are optional. The various NP-hard versions turn out to have different approximability properties depending on the type of relations and the additional constraints, but none of them can be approximated within any constant factor, unless P = NP. Particular attention is devoted to two interesting special cases that occur in discriminant analysis and machine learning. In particular, we disprove a conjecture of van Horn and Martinez (1992) regarding the existence of a polynomial-time algorithm to design linear classifiers (or perceptrons) that involve a close-to-minimum number of features.
© 1998 Published by Elsevier Science B.V. All rights reserved

Keywords: Linear systems; Unsatisfied relations; Nonzero variables; Approximability bounds; Designing linear classifiers

* Corresponding author. E-mail: amaldi@cs.cornell.edu.

[1] E.A. was partially supported by a Postdoctoral fellowship of the Swiss National Science Foundation. Part of this research was done at the Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

0304-3975/98/$19.00 © 1998 Published by Elsevier Science B.V. All rights reserved
PII S0304-3975(97)00115-1

1. Introduction

The first class of problems we consider is that of finding a minimum set of relations that must be removed from a given linear system to make it feasible. The basic versions, referred to as MIN ULR for minimum Unsatisfied Linear Relations, are defined as follows.

MIN ULR$^{\mathscr{R}}$ with $\mathscr{R} \in \{=, \geqslant, >, \neq\}$: Given a linear system $A\boldsymbol{x}\,\mathscr{R}\,\boldsymbol{b}$ with rational coefficients and with a $p \times n$ matrix A, find a solution $\boldsymbol{x} \in \mathbb{R}^n$ which violates as few relations as possible while satisfying all the others.

Many variants of these combinatorial optimization problems arise in various fields such as operations research [39, 31, 30], pattern recognition [65, 21, 49] and machine learning [1, 38, 52]. It is well known that feasible systems with equalities or inequalities can be solved in polynomial time using an adequate linear programming method [44]. But least-square methods are not appropriate for infeasible systems when the objective is to minimize unsatisfied relations.

A number of algorithms have been proposed for various versions of MIN ULR, including the weighted ones in which a weight is associated with each relation and the goal is to minimize the total weight of the unsatisfied relations. Johnson and Preparata showed that the special cases of MIN ULR$^{>}$ and MIN ULR$^{\geqslant}$ with homogeneous systems are NP-hard and devised a complete enumeration method which is also applicable to the weighted and mixed variants [39]. Greer developed a tree algorithm for optimizing functions of systems of linear relations that is more efficient than complete enumeration but still exponential in the worst case [31]. This general procedure can be used to solve MIN ULR with any of the four types of relations.

During the last decade many mathematical programming formulations have been studied to design linear discriminant classifiers (see [49, 12], as well as the included references). When the goal is to determine optimal linear classifiers which misclassify the least number of points in the training set, the problem amounts to a special case of MIN ULR$^{>}$ and MIN ULR$^{\geqslant}$. Increasingly sophisticated models have been proposed in order to try to avoid unacceptable or trivial solutions (see [12]).

The same type of problem has also attracted a considerable interest in machine learning (artificial neural networks) because it arises when training perceptrons, in particular when minimizing the number of misclassifications. While some heuristic algorithms were devised in [26, 25], Amaldi showed that solving these problems to optimality is NP-hard even when restricted to perceptrons with ± 1 inputs [1]. In [38] minimizing the number of misclassifications was proved at least as hard to approximate as the hitting set problem (see [27]).

In recent years a growing attention has been paid to infeasible linear programs [30]. When formulating or modifying very large and complex models, it is hard to prevent errors and to guarantee feasibility. Infeasible programs with thousands of constraints frequently occur and cannot be repaired by simple inspection. Several methods have been proposed in order to try to locate the source of infeasibility. While earlier ones look for minimal infeasible subsystems [28, 15], the latter ones aim at removing as few constraints as possible to achieve feasibility [59, 56, 57, 13, 29]. The more practical approach in which the modeler is allowed to weight the constraints according to their importance and flexibility leads to weighted versions of MIN ULR [57, 56].

The second class of problems we consider pertains to feasible linear systems. The goal is then to minimize the number of Relevant Variables in the Linear System.

> MIN RVLS$^{\mathscr{R}}$ with $\mathscr{R} \in \{=, \geqslant, >, \neq\}$: Given a feasible system of linear relations $A\boldsymbol{x}\,\mathscr{R}\,\boldsymbol{b}$ with rational coefficients, find a solution satisfying all relations with as few nonzero variables as possible.

MIN RVLS$^{=}$ is known to be NP-hard and was referred to as *minimum weight solution to linear equations* in [27], but nothing is known about its approximability properties.

A special case of MIN RVLS with strict and nonstrict inequalities is of particular interest in discriminant analysis and machine learning. The problem occurs when, given a linearly separable set of positive and negative examples, one wants to minimize the number of attributes that are required to correctly classify all given examples [48, 64]. This objective, which is related to the concept of parsimony, is crucial because the number of nonzero parameters of a classifier has a strong impact on its performance on unknown data [6, 46]. In [48] a genetic search strategy has been proposed for designing optimal linear classifiers with as few nonzero parameters as possible.

Since the late 1980s, various complexity classes and approximation preserving reductions have been introduced and used to investigate the approximability of NP-hard optimization problems (see [43]). Using a connection with interactive proof systems, strong bounds were derived on the approximability of several famous problems like maximum independent set, minimum graph coloring and minimum set cover [9, 51, 11, 10, 36, 37]. For a list of the currently best approximability upper and lower bounds for optimization problems, see [17].

In [5] we performed a thorough study of the approximability of the complementary problems of MIN ULR, named MAX FLS, where one looks for maximum Feasible subsystems of Linear Systems. In particular, we showed that the basic versions with $=$, $\geqslant$ or $>$ relations are NP-hard even for homogeneous systems with bipolar coefficients. While MAX FLS with equations cannot be approximated within p^{ε} for some $\varepsilon > 0$ where p is the number of relations, the variants with strict or nonstrict inequalities can be approximated within 2 but not within every constant factor.

Given the NP-hardness of the basic versions of MIN ULR, we are interested in approximation algorithms that are guaranteed to provide near-optimal solutions in polynomial time. Although complementary pairs of problems such as MIN ULR and MAX FLS are equivalent to solve optimally, their approximability properties can differ enormously (e.g., the minimum node cover and the maximum independent set problems [9, 27]).

In [7] Arora et al. established that MIN ULR$^{=}$ cannot be approximated within any constant, unless $\mathrm{P} = \mathrm{NP}$, and within a factor of $2^{\log^{1-\varepsilon} n}$ for any $\varepsilon > 0$ unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{\mathrm{polylog}\, n})$ (see also [8]). Moreover, they noted that this nonapproximability result also holds for systems of inequalities and they suggested a way of extending it to the special case which occurs when minimizing the number of misclassifications of a perceptron.

In [63, 64] the variant of Min RVLS with inequalities which arises in discriminant analysis and machine learning was proved to be at least as hard to approximate as the minimum set cover problem. Furthermore, it was shown that an approximation algorithm minimizing the number of nonzero parameters within a factor of $O(\log p)$, where p is the number of examples, would require far fewer examples to achieve a given level of accuracy than any algorithm which does not minimize this quantity. Finally, it was left as an open question whether this number could be approximated within a factor of $O(\log p)$ [63].

This paper is organized as follows. Section 2 briefly mentions the facts about the approximation of minimization problems used in the sequel. In Section 3 we recall the known approximability results for the basic versions of Min ULR and determine alternative upper and lower bounds on their approximability. Two important variants of Min ULR are also studied: the weighted ones where a different importance may be assigned to each relation and the constrained ones where some relations are mandatory while others are optional. We show that the weighted versions of Min ULR are equally hard to approximate as the basic versions, and the constrained versions are about as hard to approximate as the basic versions. In Section 4 we discuss the approximability of Min RVLS and its close relationship with Min ULR. Interestingly we find that Min RVLS$^{\neq}$ is hard to approximate even though Min ULR$^{\neq}$ is trivially solvable. Section 5 is devoted to Min ULR and Min RVLS versions where the variables are restricted to take a finite number of discrete values, in particular binary values. These problems are shown to be among the hardest to approximate. In Section 6 we discuss two interesting special cases of Min ULR and Min RVLS with inequalities that have been extensively studied in discriminant analysis and machine learning. In particular, we show that no polynomial-time algorithm is guaranteed to minimize the number of nonzero parameters of a linear classifier (perceptron) within a logarithmic factor, hereby disproving a conjecture in [63]. Section 7 contains a summary of the main results and some concluding remarks.

An earlier version of this paper appeared as a technical report [4].

2. Approximability of minimization problems

An NP optimization (NPO) problem over an alphabet Σ is a four-tuple $\Pi = (\mathscr{I}_\Pi, S_\Pi, f_\Pi, opt_\Pi)$ where $\mathscr{I}_\Pi \subseteq \Sigma^*$ is the set of *instances*, $S_\Pi(I) \subseteq \Sigma^*$ is the set of *feasible solutions* for the instance $I \in \mathscr{I}_\Pi$, $f_\Pi : \mathscr{I}_\Pi \times \Sigma^* \to \mathbb{N}$, the *objective function*, is a polynomial-time computable function and $opt_\Pi \in \{\max, \min\}$ tells if Π is a maximization or a minimization problem. See [19] for a formal definition.

For any instance I and for any feasible solution $x \in S_\Pi(I)$ of a minimization problem, the *performance ratio* of x with respect to the optimum is denoted by $\mathscr{R}_\Pi(I,x) =$

$f_\Pi(I,x)/opt_\Pi(I)$. A problem Π can be *approximated within* $p(n)$, for a function $p:\mathbb{Z}^+ \to \mathbb{R}^+$, if there exists a polynomial-time algorithm $\mathscr{A}$ such that for every $n \in \mathbb{Z}^+$ and for all instances $I \in \mathscr{I}_\Pi$ with $|I| = n$ we have that $\mathscr{A}(I) \in S_\Pi(I)$ and $R_\Pi(I, \mathscr{A}(I)) \leqslant p(n)$.

Although various reductions preserving approximability within constants have been proposed (see [40]), we will use the S-reduction which is suited to relate problems that cannot be approximated within any constant.

Definition 1 ([41]). Given two NPO problems Π and Π', an *S-reduction with size amplification* $a(n)$ from Π to Π' is a four-tuple $t = (t_1, t_2, a(n), c)$ such that
(i) t_1, t_2 are polynomial-time computable functions, $a(n)$ is a monotonically increasing positive function and c is a positive constant.
(ii) $t_1 : \mathscr{I}_\Pi \to \mathscr{I}_{\Pi'}$ and $\forall I \in \mathscr{I}_\Pi$ and $\forall x \in S_{\Pi'}(t_1(I))$, $t_2(I,x) \in S_\Pi(I)$.
(iii) $\forall I \in \mathscr{I}_\Pi$ and $\forall x \in S_{\Pi'}(t_1(I))$, $R_\Pi(I, t_2(I,x)) \leqslant c \cdot R_{\Pi'}(t_1(I), x)$.
(iv) $\forall I \in \mathscr{I}_\Pi$, $|t_1(I)| \leqslant a(|I|)$.
The composition of S-reductions is an S-reduction. If Π S-reduces to Π' with size amplification $a(n)$ and Π' can be approximated within some monotonically increasing function $u(n)$ in the size of the input instance, then Π can be approximated within $c \cdot u(a(n))$. For constant and polylogarithmic approximable problems, the S-reduction preserves approximability within a constant for any polynomial size amplification. For n^c approximable problems, the S-reduction preserves approximability within a constant just for linear size amplification.

An NPO problem Π is *polynomially bounded* if there is a polynomial p such that

$$f_\Pi(I,x) \leqslant p(|I|). \quad \forall I \in \mathscr{I}_\Pi \forall x \in S_\Pi(I),$$

The class of all polynomially bounded NPO problems is called NPO PB. Clearly, Min ULR and Min RVLS are in NPO PB since their objective functions are bounded by the total number of relations and, respectively, the total number of variables.

The range of approximability of NP-hard optimization problems stretches from problems which can be approximated within every constant in polynomial time, i.e. that have a polynomial-time approximation scheme like the knapsack problem, to problems that cannot be approximated within $n^{1-\varepsilon}$ for every $\varepsilon > 0$, where n is the size of the input instance, unless P = NP.

In [51] Lund and Yannakakis established a lower bound on the approximability of Min Set Cover and of several closely related problems such as Min Dominating Set. In [11] Bellare et al. improved this result by showing, among others, that Min Set Cover cannot be approximated within any constant factor unless P = NP. A stronger lower bound obtained under a stronger assumption was further improved by Feige [24] who recently showed that approximating Min Set Cover within $(1-\varepsilon)\ln n$, for any $\varepsilon > 0$, would imply NP $\subseteq$ DTIME $(n^{\log\log n})$, where n is the number of elements in the ground set. Since DTIME$(T(n))$ denotes the class of problems which can be solved in time $T(n)$, the above inclusion is widely believed to be unlikely. If there is an approximation preserving reduction from Min Dominating Set to an NPO problem Π we say that Π is Min Dominating Set-*hard*, which means that it is at least as hard to approximate as the former problem.

If we require the dominating set in Min Dominating Set to be independent, we get the minimum independent dominating set problem or Min Ind Dom Set. Halldórsson established in [34] that, assuming P $\neq$ NP, Min Ind Dom Set cannot be approximated within a factor of $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the number of nodes in the graph. Inspection of the proof shows that the result is still valid if n is the input size, i.e., the sum of the number of nodes and edges in the graph. Furthermore, Kann proved that Min Ind Dom Set is complete NPO PB in the sense that every polynomially bounded NPO problem can be reduced to it using an approximation preserving reduction [41, 18].

3. Approximability of Min ULR variants

In this section we discuss lower and upper bounds on the approximability of the basic versions of Min ULR with the different types of relations and then focus on the weighted as well as constrained variants. To try to find the simplest versions of these problems that are still hard, we restrict the range of the coefficients and of the right-hand side components.

For homogeneous systems, which have the simplest right-hand sides, we are obviously not interested in trivial solutions where all variables occurring in the satisfied equalities or nonstrict inequalities are zero (see [39] for an example). Even if we forbid the solution $\boldsymbol{x} = \boldsymbol{0}$, there might be other undesirable solutions where almost all variables occurring in the set of satisfied relations are zero except a few that only occur in a few satisfied relations. In order to rule out such meaningless solutions, we only consider solutions of maximal (with respect to inclusion) feasible subsystems in which at least a small fraction $0 < f \ll 1/2$ of the variables occurring in the satisfied relations are nonzero and these nonzero variables occur in at least a fraction f of the satisfied relations. As we shall see, our results do not depend on the specific value of f, so long as it is fixed a priori.

3.1. Basic versions

In [2, 5] we proved that Min ULR$^{\mathscr{R}}$ with $\mathscr{R} \in \{=, \geqslant, >\}$ is NP-hard even when restricted to homogeneous systems with bipolar coefficients in $\{-1, 1\}$. Sankaran showed in [59] that the NP-complete problem Min Feedback Arc Set [27], in which one wishes to remove a smallest set of arcs from a directed graph to make it acyclic, reduces to Min ULR$^{\geqslant}$ with exactly one 1 and one -1 in each row of A and all right-hand sides equal to 1. Unlike the other problems, Min ULR$^{\neq}$ is trivially solvable because any such system is feasible. Indeed, for any finite set of hyperplanes associated with a set of linear relations there exists a vector $\boldsymbol{x} \in \mathbb{R}^n$ that does not belong to any of them.

Note that if the number of variables n is constant these three basic versions of MIN ULR can be solved in polynomial time using Greer's algorithm which has an $O(n \cdot p^n/2^{n-1})$ time-complexity, where p denotes the number of relations [31]. These problems are trivial when the number of relations p is constant because all subsystems can be checked in $O(n)$ time. Furthermore, they are easy when all maximal feasible subsystems contain a maximum number of relations because a greedy procedure is guaranteed to give a solution that minimizes the number of unsatisfied relations. A polynomial-time solvable special case of MIN ULR$^{\geqslant}$ involving total unimodularity is also mentioned in [59].

Before turning to lower and upper bounds on the approximability of MIN ULR, we point out a few straightforward facts.

Fact 1. MIN ULR$^{\geqslant}$ *is at least as hard to approximate as* MIN ULR$^{=}$ *but not harder than* MIN ULR$^{=}$ *with nonnegative variables.* MIN ULR$^{\geqslant}$ *and* MIN ULR$^{>}$ *with integer (rational) coefficients are equivalent.*

Minimizing the number of unsatisfied equations in an arbitrary MIN ULR$^{=}$ instance is obviously equivalent to minimizing the number of violated inequalities in the corresponding instance of MIN ULR$^{\geqslant}$ where each equation is replaced by the two complementary inequalities. Given an arbitrary instance of MIN ULR$^{\geqslant}$, replacing each variable x_i unrestricted in sign by the difference $x_i' - x_i''$ of two nonnegative variables $x_i', x_i'' \geqslant 0$ and adding a slack variable for each inequality leads to an equivalent system with p equations and $2n + p$ nonnegative variables. Finally, any system $A\boldsymbol{x} \leqslant \boldsymbol{b}$ has a solution if and only if the system $A\boldsymbol{x} < \boldsymbol{b} + \varepsilon\mathbf{1}$ has a solution, where $\varepsilon = 2^{-2L}$ and L is the size (in bits) of the binary encoded input instance [55].

As previously mentioned, Arora et al. showed in [7] (see also [8]) that MIN ULR$^{=}$ cannot be approximated within any constant, unless P=NP, and within a factor of $2^{\log^{1-\varepsilon} n}$, for any $\varepsilon > 0$, unless NP $\subseteq$ DTIME$(n^{\mathrm{polylog}\, n})$, where n is the number of variables. Of course, this also holds for MIN ULR with strict and nonstrict inequalities.

The following nonapproximability result for MIN ULR with inequalities is more likely to be true but the bound is not as strong.

Theorem 2. MIN ULR$^{\geqslant}$ *and* MIN ULR$^{>}$ *are* MIN DOMINATING SET*-hard even when restricted to homogeneous systems with ternary coefficients in* $\{-1, 0, 1\}$. *They cannot be approximated within any constant, unless* P = NP, *and within* $(1 - \varepsilon)\ln n$, *for any* $\varepsilon > 0$, *unless* NP $\subseteq$ DTIME$(n^{\log\log n})$, *where n is the number of variables.*

Proof. We proceed by cost preserving reduction from MIN DOMINATING SET [27] similar to [38]. Given an undirected graph $G = (V, E)$, one seeks a minimum cardinality set $V' \subseteq V$ that dominates all nodes of G, i.e. for all $v \in V \setminus V'$ there exists $v' \in V'$ such that $[v, v'] \in E$. Let $G = (V, E)$ be an arbitrary instance of MIN DOMINATING SET. For each node $v_i \in V$, $1 \leqslant i \leqslant n$, we consider the homogeneous inequality $x_i \geqslant 0$ and the inhomogeneous inequality

$$x_i + \sum_{j \in N(v_i)} x_j \leqslant -1$$

where $N(v_i)$ is the set of indices of the nodes adjacent to v_i. Thus we have a system with $2n$ inequalities and n variables.

It is easily verified that there exists a dominating set in G of size at most s if and only if there exists a solution $\boldsymbol{x}$ that violates at most s inequalities of the corresponding system. Given a dominating set $V' \subseteq V$ of size s, the solution $\boldsymbol{x}$ defined by

$$x_i = \begin{cases} -1 & \text{if } v_i \in V', \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

satisfies all inhomogeneous inequalities and $n - s$ homogeneous ones. Conversely, given a solution vector $\boldsymbol{x}$ that violates s inequalities, we can always satisfy every inhomogeneous inequality that is not already satisfied by making one variable x_i in the inequality negative enough. This operation yields a solution that satisfies at least as many inequalities as $\boldsymbol{x}$. Consider the set of nodes $V' \subseteq V$ containing all nodes v_i such that $x_i \neq 0$. V' is clearly a dominating set of size s, because $x_i + \sum_{j \in N(v_i)} x_j \leqslant -1$ only when at least one of the variables is negative, which corresponds to the case where at least one of the nodes is in the dominating set.

By replacing $\geqslant$ by $>$ and -1 in the right-hand side of the second type of inequalities by 0, this reduction can be adapted to homogeneous MIN ULR$^{>}$.

There is a standard way to transform any inhomogeneous instance of MIN ULR$^{\geqslant}$ into a homogeneous one. Given an arbitrary inhomogeneous instance with p inequalities in n variables and given any value of the fraction f in the meaningful solution criterion, we first multiply all constant right-hand sides by a new variable x_0. The resulting homogeneous system is obviously equivalent to the original one if x_0 is restricted to be strictly positive. To enforce this constraint, it suffices to add the $3L$ inequalities $x_0 \geqslant x_{0l}$, $x_0 \leqslant x_{0l}$ and $x_{0l} \geqslant 0$ involving the new variables x_{0l} with $l \in [1..L]$, where L is a large enough integer such that $p/(3L+p) < f$ and $(3L)/(3L+n) \geqslant f$. These $3L$ new inequalities can be clearly satisfied by assigning to x_0 and to all x_{0l} the same nonnegative value. Due to the choice of L, this value cannot be zero because, if $x_0 = x_{0l} = 0$ for all $l \in [1..L]$, at most a fraction $p/(3L+p) < f$ of the inequalities of the homogeneous system are satisfied. Moreover, when $x_0 = x_{0l} > 0$ for all l at least a fraction f of all variables are nonzero. Thus the homogeneous system admits a solution that violates at most s inequalities with $s \leqslant p$ if and only if the original inhomogeneous system admits such a solution.

Since the reduction is cost preserving and without amplification, we have exactly the same nonapproximability bounds for MIN ULR$^{\geqslant}$ and MIN ULR$^{>}$ as for MIN DOMINATING SET. □

Clearly, for large n and small $\varepsilon > 0$, a factor of $2^{\log^{1-\varepsilon} n}$ is larger than $\ln n$, but NP $\subseteq$ DTIME($n^{\mathrm{polylog}\, n}$) is more likely to be true than NP $\subseteq$ DTIME($n^{\log \log n}$). Furthermore, the above proof is much simpler than that given in [7].

Unlike for MAX FLS$^{=}$ [5], for MIN ULR$^{=}$ we can guarantee in polynomial time a performance ratio that is linear in the number of variables. This fact is mentioned without proof in [7, 8].

Proposition 3. MIN ULR$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{=, \geqslant, >\}$ *is approximable within* $n+1$, *where* n *is the number of variables.*

Proof. When applied to linear systems, Helly's theorem (see [16]) implies that, for any infeasible system of inequalities or equations in n variables, all minimal infeasible subsystems contain at most $n+1$ relations. Such a *Helly obstruction* can be found using any polynomial time method for linear programming (LP) [22]. According to Farkas' lemma (see [60]), a system $A\boldsymbol{x} \leqslant \boldsymbol{b}$ with p inequalities and n variables is infeasible if and only if there exists a nonnegative vector $\boldsymbol{y} \geqslant \boldsymbol{0}$ such that $\boldsymbol{y}^{\mathrm{t}}A = \boldsymbol{0}$ and $\boldsymbol{y}^{\mathrm{t}}\boldsymbol{b} < 0$. In fact, the result is still valid if the vector $\boldsymbol{y} \geqslant \boldsymbol{0}$ is required to have at most $n+1$ nonzero components. For infeasible systems, a polynomial-time LP algorithm produces a $\boldsymbol{y}$ satisfying Farkas' lemma. If $\boldsymbol{y}$ has more than $n+1$ nonzero components, some of them can be driven to zero. Therefore it suffices to find a nontrivial solution $\boldsymbol{z}$ of the auxiliary system $\boldsymbol{z}^{\mathrm{t}}[A|\boldsymbol{b}] = [\boldsymbol{0}^{\mathrm{t}}|0]$ such that $\boldsymbol{z}$ is zero for every component where $\boldsymbol{y}$ is zero. This simply amounts to determining a nontrivial solution to $n+1$ homogeneous equations in more than $n+1$ variables. Subtracting a multiple of $\boldsymbol{z}$ from $\boldsymbol{y}$ leads to a new $\boldsymbol{y}$ with fewer nonzero components. By repeating this process, we obtain in polynomial time a $\boldsymbol{y}$ with at most $n+1$ nonzero components that correspond to the inequalities in a Helly obstruction.

Thus, starting with an infeasible system, we can identify an obstruction and delete it iteratively until the resulting system is feasible, that is at most $p/(n+1)$ times. Clearly, we remove at most $n+1$ times more inequalities than needed because at each step we delete at most $n+1$ relations corresponding to a Helly obstruction while a single one may suffice. □

The question of whether it is NP-hard to guarantee a polylogarithmic performance ratio in n is still open in the general case.

The answer is negative for a particular class of inequality systems with totally unimodular matrices. More precisely, we consider node-arc incidence matrices of directed graphs, i.e. which contain exactly one 1 and one -1 in each row (all other components being 0). For this type of matrices, MIN ULR$^{\geqslant}$ with all second hand sides equal to 1 and homogeneous MIN ULR$^{>}$ cannot be approximated within every constant, unless P = NP, but are approximable within a factor of $O(\log n \log\log n)$, where n is the number of variables. This follows using a straightforward modification of the polynomial-time reduction from MIN FEEDBACK ARC SET to MIN ULR with $\leqslant$ relations given in [59]. For each arc (v_i, v_j) in a given instance of MIN FEEDBACK ARC SET, we consider the nonstrict inequality $x_i - x_j \geqslant 1$ or, respectively, the strict inequality $x_i - x_j > 0$. In fact, it is readily verified that the two special cases of MIN ULR with inequalities are equivalent to MIN FEEDBACK ARC SET. Since MIN FEEDBACK ARC SET is

APX-hard (see for example [40]), it cannot be approximated within every constant unless P = NP. However, it is known to be approximable within $O(\log n \log\log n)$, where n is the number of nodes in the graph [23].

3.2. Weighted and constrained versions

In many practical situations, all relations do not have the same importance. This can be taken into account by assigning a weight to each one of them and by looking for a solution that minimizes the total weight of the unsatisfied relations [31, 56].

Proposition 4. *Weighted* MIN ULR$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{=, \geqslant, >\}$ *and positive integer (rational) weights is equally hard to approximate as the corresponding basic version.*

Proof. Basic MIN ULR$^{\mathcal{R}}$ is clearly a special case of weighted MIN ULR$^{\mathcal{R}}$ where all weights are equal to one.

For proving the other direction, we first use the following result from [20]: For any "nice subset problem" with polynomially bounded weights that is approximable within a polynomial $r(n)$ in the size of the input, the unrestricted version of the same problem where the weights are not polynomially bounded is approximable within $r(n) + 1/n$.

Since it is easily verified that MIN ULR with equalities or inequalities are nice subset problems, only instances with polynomially bounded weights need to be considered. Thus, it suffices to show that any such instance can be associated with an equivalent unweighted one. This is simply achieved by making for each relation a number of copies equal to the corresponding weight. The number of relations will still be polynomial since the weights are polynomially bounded. □

Interesting special cases of weighted MIN ULR include the constrained versions where some relations are mandatory while the others are optional (see [31] for an example from the field of linear numeric editing). C MIN ULR$^{\mathcal{R}_1;\mathcal{R}_2}$ with $\mathcal{R}_1, \mathcal{R}_2 \in \{=, \geqslant, >, \neq\}$ denotes the variant where the mandatory relations are of type $\mathcal{R}_1$ and the optional ones of type $\mathcal{R}_2$. When $\mathcal{R}_1 = \mathcal{R}_2$ the problem can be seen as a weighted MIN ULR$^{\mathcal{R}_1}$ problem in which the weight of every mandatory relation is larger than the total weight of all optional ones. In this case, the constrained versions of MIN ULR are equally hard to approximate as the corresponding basic versions.

It is worth noting that no such relation exists between constrained and unweighted versions of the complementary problems MAX FLS. As we proved in [5], enforcing some mandatory relations makes MAX FLS with inequalities harder to approximate. While MAX FLS$^{\geqslant}$ and MAX FLS$^{>}$ can be approximated within a factor 2, the constrained variants are at least as hard as the maximum independent set problem and hence cannot be approximated within a factor of $n^{1-\varepsilon}$ for any $\varepsilon > 0$ unless NP $\subseteq$ co-RP, where n is the number of nodes [36].

Any instance of a constrained problem C MIN ULR$^{=;\mathcal{R}}$ with $\mathcal{R} \in \{=, \geqslant, >, \neq\}$ can be transformed into an equivalent instance of MIN ULR$^{\mathcal{R}}$. Indeed, by applying Gaussian elimination to the mandatory equations, each variable is expressed in terms of other possibly free variables and it then suffices to substitute the variables in the optional relations accordingly. Since C MaxFLS$^{\mathscr{R},\neq}$ with $\mathscr{R} \in \{=, \geqslant, >, \neq\}$ is solvable in polynomial time [5], all the problems C Min ULR$^{\mathscr{R};\neq}$ are solvable in polynomial time.

Fact 5. C Min ULR$^{\geqslant;=}$ *is equally hard to approximate as* Min ULR$^{\geqslant}$.

According to Fact 1, Min ULR$^{\geqslant}$ can be reduced to Min ULR$^{=}$ with nonnegative variables. The latter problem is a particular case of C Min ULR$^{\geqslant;=}$ where the mandatory inequalities are just nonnegativity constraints. Conversely, C Min ULR$^{\geqslant;=}$ can be reduced to C Min ULR$^{\geqslant;\geqslant}$ by substituting each optional equation by two complementary inequalites. We can then use the standard reduction from C Min ULR$^{\geqslant;\geqslant}$ to Min ULR$^{\geqslant}$. Similarly we can show that C Min ULR$^{>;=}$ is at least as hard to approximate as Min ULR$^{>}$.

Proposition 6. C Min ULR$^{\neq;\neq}$ *is* Min Dominating Set-*hard even when restricted to homogeneous systems with binary coefficients.*

Proof. By reduction from Min Dominating Set as in Theorem 2. For each node $v_i \in V$, $1 \leqslant i \leqslant n$, of an arbitrary instance $G=(V,E)$, we consider the optional equation $x_i = 0$ and the mandatory relation

$$x_i + \sum_{j \in N(v_i)} x_j \neq 0. \tag{2}$$

Then there exists a dominating set in G of size at most s if and only if there exists a solution that violates at most s optional equations. □

4. Approximability of Min RVLS

In this section we discuss the approximability of the basic Min RVLS$^{\mathscr{R}}$ variants with $\mathscr{R} \in \{=, \geqslant, >, \neq\}$. Although Min RVLS and Min ULR deal with different types of systems (feasible versus infeasible), we shall see that they are very closely related.

Theorem 7. *Assuming* NP $\not\subseteq$ DTIME($n^{\mathrm{polylog}\, n}$), Min RVLS$^{\mathscr{R}}$ *with* $\mathscr{R} \in \{=, \geqslant, >, \neq\}$ *is not approximable within a factor of* $2^{\log^{1-\varepsilon} n}$, *for any* $\varepsilon > 0$, *where n is the number of variables.*

Proof. For Min RVLS$^{=}$, we show that there exists a simple cost preserving reduction from Min ULR$^{=}$ and use the fact that the latter problem cannot be approximated within a factor of $2^{\log^{1-\varepsilon} n}$ [8].

Let $(A, \boldsymbol{b})$ be an arbitrary instance of Min ULR$^{=}$ and $\boldsymbol{s} \in \mathbb{R}^p$ be the vector of slack variables. Looking for a solution of $A\boldsymbol{x} = \boldsymbol{b}$ with as few unsatisfied equations as possible is equivalent to looking for $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{s} \in \mathbb{R}^p$ such that

$$A\boldsymbol{x} + \boldsymbol{s} = \boldsymbol{b} \tag{3}$$

with as few nonzero slacks s_i, $1 \leqslant i \leqslant p$, as possible. Obviously $\boldsymbol{s} \neq \boldsymbol{0}$ whenever the system is infeasible. Now consider a matrix D whose rows form a basis for the subspace orthogonal to the column space of A, i.e. D satisfies $DA = 0$ and has the largest rank among all such matrices, namely $rank(D) = p - rank(A)$. Since (3) is equivalent to $DA\boldsymbol{x} + D\boldsymbol{s} = D\boldsymbol{b}$, the instance of Min ULR$^=$ is equivalent to finding a solution $\boldsymbol{s} \in \mathbb{R}^p$ of the feasible system

$$D\boldsymbol{s} = D\boldsymbol{b} \tag{4}$$

with a minimum number of nonzero components. Note that, if $A\boldsymbol{x} = \boldsymbol{b}$ is infeasible, $D\boldsymbol{b} \neq \boldsymbol{0}$. Clearly, there exists an $\boldsymbol{s}$ satisfying (4) with k nonzero components if and only if there exists an $\boldsymbol{x}$ satisfying all but k equations of $A\boldsymbol{x} = \boldsymbol{b}$. The simple connection is given by (3).

The same bound is also valid for systems with strict and nonstrict inequalities because, according to Fact 1, each equation can be replaced by an appropriate pair of inequalities.

For Min RVLS$^{\neq}$, we first proceed by cost preserving reduction from Min Dominating Set as in Theorem 2. For each node $v_i \in V$, $1 \leqslant i \leqslant n$, of an arbitrary instance $G = (V, E)$, we consider the relation

$$x_i + \sum_{j \in N(v_i)} x_j \neq 0 \tag{5}$$

where $N(v_i)$ is the set of indices of the nodes adjacent to v_i. Thus we have a system with n relations and n variables.

Clearly, there exists a dominating set in G of size at most s if and only if the corresponding system (5) has a solution $\boldsymbol{x}$ with at most s nonzero components. Given the lower bounds for Min Dominating Set, Min RVLS$^{\neq}$ cannot be approximated within any constant factor $c > 1$ unless P = NP. In other words, assuming P $\neq$ NP, any polynomial time algorithm would provide a solution with more than $\lfloor cs \rfloor$ nonzero variables for some "bad" Min RVLS$^{\neq}$ instances corresponding to Min Dominating Set instances with a dominating set of size s.

To obtain the $2^{\log^{1-\varepsilon} n}$ factor, we proceed by self-improvement as in [8]. The idea is to show that any gap $c > 1$ can be increased recursively. For an arbitrary Min Dominating Set instance, we start with the corresponding system (5) ($A\boldsymbol{x} \neq \boldsymbol{0}$) whose coefficients are 0 or 1 and we construct the squared system

$$A'\boldsymbol{x} \neq \boldsymbol{0}' \tag{6}$$

obtained by replacing each element of A equal to 1 by the whole matrix A and each element equal to 0 by the $p \times n$ matrix with all zeroes. Thus A' is a matrix of size $p^2 \times n^2$ and $\boldsymbol{0}'$ is a vector with p^2 components all equal to 0. Clearly, if the original Min Dominating Set instance has a dominating set $V' \subseteq V$ of size s, there is a solution

$\boldsymbol{x}$ of (6) with s^2 nonzero variables. However, any polynomial-time algorithm would provide solutions with more than $\lfloor c^2 s^2 \rfloor$ nonzero variables for its "bad" instances.

By applying the above construction t times recursively, we obtain a system with p^{2^t} equations and n^{2^t} variables. Let $t = \log(\log^\beta n)$, where n is the number of variables in the MIN RVLS$^{\neq}$ instance corresponding to the considered MIN DOMINATING SET instance and β is a positive real number. The construction requires $O(n^{\text{polylog}\, n})$ time because the system has $p' = p^{\log^\beta n}$ equations and $n' = n^{\log^\beta n} = 2^{\log^{\beta+1} n}$ variables. Since $\log n' = \log^{\beta+1} n$, an initial gap of, say, $c = 2$ implies a total gap of $c^{2^t} = c^{\log^\beta n} = 2^{\log^{\beta/(\beta+1)} n'}$.

The bound follows by contradiction. Suppose there exists a polynomial-time algorithm that approximates MIN RVLS$^{\neq}$ instances with n variables within a factor of $2^{\log^{1-\varepsilon} n}$ for any $\varepsilon > 1/(\beta+1)$. By applying it to the resulting instance of MIN RVLS$^{\neq}$, one could approximate within a factor of 2 and in $O(n^{\text{polylog}\, n})$ time any given instance of MIN DOMINATING SET. But this would imply $\text{NP} \subseteq \text{DTIME}(n^{\text{polylog}\, n})$. □

As we shall see in Section 6, this nonapproximability bound also holds for a special case of MIN RVLS with inequalities that arises in discriminant analysis and machine learning. The same is true for homogeneous MIN RVLS$^{\mathscr{R}}$ with $\mathscr{R} \in \{=, \geqslant\}$. Indeed, the reduction for MIN ULR$^{=}$ given in [8] (cf. also the proof of Theorem 12) can be easily extended to the case of homogeneous systems in which the trivial solution with all zero variables is discarded.

Note that the *shortest codeword problem* in coding theory (see MS7 entry in [27]) is the same problem as MIN RVLS$^{=}$ over $GF(2)$. By similar methods as above this problem can be shown to have the same nonapproximability bound as ordinary MIN RVLS$^{=}$. This result can also be shown directly by using the recent structural results by Khanna et al. [47].

In fact, not only MIN ULR$^{=}$ is a special case of MIN RVLS$^{=}$ but we also have:

Proposition 8. MIN ULR$^{\mathscr{R}}$ *with* $\mathscr{R} \in \{=, \geqslant, >\}$ *is at least as hard to approximate as* MIN RVLS$^{\mathscr{R}}$ *with the same type of relations.*

Proof. For any instance of MIN RVLS$^{=}$, one can construct an equivalent instance of MIN ULR$^{=}$ by considering, for each variable x_i with $1 \leqslant i \leqslant n$, the equation $x_i = 0$. By applying Gaussian elimination to the original MIN RVLS$^{=}$ instance, each variable x_i is expressed in terms of some possibly free variables. It then suffices to replace each variable in $\boldsymbol{x} = \boldsymbol{0}$ by the corresponding expression.

Since each equation $x_i = 0$ can be replaced by the two complementary inequalities $x_i \geqslant 0$ and $x_i \leqslant 0$, MIN ULR$^{\geqslant}$ is at least as hard to approximate as MIN RVLS$^{\geqslant}$. Also MIN ULR$^{>}$ is at least as hard because it is equivalent to MIN ULR$^{\geqslant}$ for systems with integer (rational) coefficients. □

The same reduction implies that the complementary maximization problem MAX IVLS$^{=}$ (maximum number of Irrelevant Variables in Linear Systems) restricted to homogeneous systems is equally hard to approximate as homogeneous MAX FLS$^=$, i.e. not approximable within p^ε for some $\varepsilon > 0$ unless P = NP [5].

Interestingly, MAX IVLS$^\geqslant$ and MAX IVLS$^>$ are much harder to approximate than MAX FLS$^\geqslant$ and MAX FLS$^>$, respectively. It is easy to show that the former problems are harder than the maximum independent set problem (which is not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$ unless NP $\subseteq$ co-RP, where n is the number of nodes [36]), while the latter ones can be approximated within 2 [5]. It suffices to construct, for each edge $e = [v_i, v_j]$, the inequality $x_i + x_j \geqslant 1$ or $x_i + x_j > 0$ and to observe that there is a correspondence between the independent sets of cardinality at least s and the solutions with at least s zero components.

5. Hardness of variants with bounded discrete variables

In this section we consider the MIN ULR and MIN RVLS variants in which the variables are restricted to take a finite number of discrete values. See [14, 33] for the problem of analyzing mixed-integer and integer linear programs. Since systems with bounded discrete variables can be reduced to systems with binary variables in $\{0, 1\}$, we study the latter class of problems that is referred to as BIN MIN ULR.

Theorem 9. BIN MIN ULR$^{\mathcal{R}_1}$ *and* C BIN MIN ULR$^{\mathcal{R}_1;\mathcal{R}_2}$ *are* NPO PB-*complete for every combination of* $\mathcal{R}_1, \mathcal{R}_2 \in \{=, \geqslant, >, \neq\}$. *Assuming* P $\neq$ NP, C BIN MIN ULR$^{\mathcal{R}_1;\mathcal{R}_2}$ *and* BIN MIN ULR$^{\mathcal{R}_1}$ *cannot be approximated within* $s^{1-\varepsilon}$ *and, respectively, within* $s^{0.5-\varepsilon}$ *for any* $\varepsilon > 0$, *where* s *is the sum of the number of variables and relations.*

Proof. We show the result for C BIN MIN ULR$^{\geqslant;\geqslant}$ and then extend it to the other variants. We proceed by reduction from MIN IND DOM SET in which, given an undirected graph $G = (V, E)$, one seeks a minimum cardinality independent set $V' \subseteq V$ that dominates all nodes of G [27]. For each node $v_i \in V$, $1 \leqslant i \leqslant n$, of an arbitrary instance $G = (V, E)$, we consider the optional inequality

$$x_i \leqslant 0 \tag{7}$$

and the mandatory one

$$x_i + \sum_{j \in N(v_i)} x_j \geqslant 1 \tag{8}$$

where $N(v_i)$ is defined as above. Furthermore, we construct for each edge $[v_i, v_j] \in E$ the mandatory inequality

$$x_i + x_j \leqslant 1. \tag{9}$$

Thus we have a system with n variables, n optional inequalities and $n + |E|$ mandatory ones.

It is easily verified that there exists an independent dominating set in G of size at most s if and only if there exists a solution $x \in \{0, 1\}^n$ that violates s optional relations

Table 1

	operator $\geqslant$	operator $>$	operator $\neq$	operator $=$
type (7)	$\leqslant 0$	<1	$\neq 1$	$=0$
type (8)	$\geqslant 1$	>0	$\neq 0$	$= \lvert N(v_i)\rvert - \sum_{j=1}^{\lvert N(v_i)\rvert - 1} y_{ij}$
type (9)	$\leqslant 1$	<2	$\neq 2$	$= 1 - z_{ij}$

of the corresponding system. The mandatory relations (8) enforce the dominance constraint while the relations (9) enforce the independence constraint. The result follows because Min Ind Dom Set is NPO PB-complete and cannot be approximated within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the sum of the number of nodes and edges in the graph.

For the other constrained problems C Bin Min ULR$^{\mathscr{R}_1;\mathscr{R}_2}$, we use the same reduction as above but the right-hand side of the three types of relations must be substituted according to Table 1. In the case of mandatory equations we need to introduce $2|E| - n$ additional slack variables y_{ij} and $|E|$ additional slack variables z_{ij}. Thus the total number of variables will be $3|E|$, that is, still a linear number in n and $|E|$.

For the unconstrained problems, we add $|V| + 1$ copies of each mandatory relation so that they are more valuable than the optional ones. Since such a reduction has a quadratic size amplification, we get a weaker nonapproximability bound than for Min Ind Dom Set. □

It is worth noting that Bin Min ULR$^{\mathscr{R}_1}$ and C Bin Min ULR$^{\mathscr{R}_1;\mathscr{R}_2}$ with $\mathscr{R}_1, \mathscr{R}_2 \in \{=, \geqslant, >, \neq\}$ remain NPO PB-complete for homogeneous systems. In the above reduction, we multiply each nonzero constant in the right-hand side of a relation by a new variable x_0. In order to prevent x_0 from being zero we add the new mandatory relations $x_0 > 0$, $x_0 \neq 0$, $x_0 = x_{01}$, or $x_0 \leqslant x_{01}$ and $x_{01} \geqslant 0$ involving a new variable x_{01}, depending on the type of relations. In the case of nonstrict inequalities and equalities, we add (as in the proof of Theorem 2) a large enough number of copies of those relations.

Similar bounds also hold for Min RVLS with binary variables that is referred to as Bin Min RVLS.

Proposition 10. Bin Min RVLS$^{\mathscr{R}}$ *with* $\mathscr{R} \in \{=, \geqslant, >, \neq\}$ *is* NPO PB-*complete. Assuming* P $\neq$ NP, Bin Min RVLS$^{=}$ *and* Bin Min RVLS$^{\mathscr{R}}$ *with* $\mathscr{R} \in \{\geqslant, >, \neq\}$ *are not approximable within* $n^{0.5-\varepsilon}$ *and, respectively, within* $n^{1-\varepsilon}$ *for any* $\varepsilon > 0$, *where n is the number of variables.*

Proof. The reduction is very similar to the one used in Theorem 9 for C Bin Min ULR$^{\mathscr{R};\mathscr{R}}$ with $\mathscr{R} \in \{=, \geqslant, >, \neq\}$. The Bin Min RVLS$^{\mathscr{R}}$ instance is simply composed of the mandatory relations (8) and (9). Since the number of violated optional relations exactly corresponds to the number of nonzero variables, Bin Min RVLS$^{\mathcal{R}}$ with $\mathcal{R} \in \{\geqslant, >, \neq\}$ are NPO PB-hard and not approximable within $n^{1-\varepsilon}$.

For Bin Min RVLS$^{=}$, we have to deal with the slack variables y_{ij} and z_{ij} that have been added. Suppose there is a total number of N slack variables. In order to make the x variables more valuable than all the N slack ones, we introduce, for each variable x_i, N new variables $x_{i1}, \ldots, x_{iN}$ and the N additional equations $x_i - x_{ij} = 0$ for $j \in [1..N]$. In any solution $\boldsymbol{x}$ of the resulting instance we will have, for each variable x_i, that $x_i = x_{i1} = \cdots = x_{iN}$. Consider the set of nodes $V' \subseteq V$ containing all nodes v_i such that $x_i = 1$. V' is clearly independent and dominating. If t variables in $\boldsymbol{x}$ are equal to 1, the size of this set will be $\lfloor t/(N+1) \rfloor$.

Conversely, an independent dominating set containing s nodes corresponds to a solution of the Bin Min RVLS$^{=}$ instance with between $s(N+1)$ and $s(N+1)+N$ variables equal to 1. Thus the reduction is an S-reduction with size amplification $O(nN)$ and we get the nonapproximability bound $n^{0.5-\varepsilon}$, where n is the number of variables. □

Note that Bin Min RVLS$^{\geqslant}$ is equivalent to Min Polynomially Bounded 0–1 Programming, which was shown to be NPO PB-complete in [41]. Moreover, the corresponding maximization problem Bin Max IVLS$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geqslant, >, \neq\}$ is NPO PB-complete and cannot be approximated within $s^{1/3-\varepsilon}$ for any $\varepsilon > 0$, where s is the sum of the number of variables and relations, unless P = NP [42].

6. Special cases from discriminant analysis and machine learning

In this section we discuss two interesting special cases of Min ULR and Min RVLS with inequalities which arise in discriminant analysis and machine learning, more precisely, when designing two-class linear classifiers [21] and when training perceptrons [54].

Given a set of vectors $T = \{\boldsymbol{a}^k\}_{1 \leqslant k \leqslant p} \subset \mathbb{R}^n$ labeled as positive or negative examples, we look for a hyperplane H, specified by a normal vector $\boldsymbol{w} \in \mathbb{R}^n$ and a bias $w_0 \in \mathbb{R}$, such that all the positive vectors lie on the positive side of H while all the negative ones lie on the negative side. A hyperplane H is said to be *consistent* with an example $\boldsymbol{a}^k$ if $\boldsymbol{a}^k \boldsymbol{w} > w_0$ or $\boldsymbol{a}^k \boldsymbol{w} \leqslant w_0$ depending on whether $\boldsymbol{a}^k$ is positive or negative. In other words, we seek a discriminant hyperplane separating the examples in the first class from those in the second class. In the artificial neural network literature, such a linear threshold unit is known as a *preceptron* and its parameters $w_j, 1 \leqslant j \leqslant n$ as its *weights* [35].

In the general situation where T is nonlinearly separable, a natural objective is to minimize the number of vectors $\boldsymbol{a}^k$ that are misclassified (see [49, 25] and the included references). This problem is referred to as Min Misclassifications. Note that we have studied in [5] the approximability of the complementary problem where one looks for a hyperplane which is consistent with as many $\boldsymbol{a}^k \in T$ as possible.

In [7] a way of extending the nonapproximability bounds for MIN ULR$^=$ to the symmetric version of MIN MISCLASSIFICATIONS where we ask $\boldsymbol{a}^k \boldsymbol{w} < w_0$ for negative examples is suggested. Although the argument used does not suffice to complete the proof, it can easily be fixed.

The problem is related to the fact that starting with any instance of MIN ULR$^=$ we must construct a system with strict inequalities with a particular variable playing the role of the bias w_0. As mentioned in [7], one can easily associate to any considered instance of MIN ULR$^=$ an equivalent inhomogeneous instance of MIN ULR$^>$. It suffices to replace each equation by two nonstrict inequalities, and then to add a new slack variable δ so as to turn each nonstrict inequality into a strict one. More precisely, every relation $\boldsymbol{a}\boldsymbol{w} \geqslant 0$ is replaced by $\boldsymbol{a}\boldsymbol{w} + \delta > 0$. Now, in order to make sure that the two systems are equivalent we must have $\delta < 1/L$ with $L = \lfloor c \cdot K \rfloor$, where c and K are constants as in the proof of Theorem 12. This can of course be guaranteed by introducing a large enough number of copies of this strict inequality, but then the resulting system is not an instance of symmetric MIN MISCLASSIFICATIONS. Indeed, if δ is considered as the threshold the inequalities ensuring $\delta < 1/L$ are not homogeneous.

Fortunately, there exists a simple and general technique to construct, for any instance of inhomogeneous MIN ULR$^>$, an equivalent instance of symmetric MIN MISCLASSIFICATIONS.

Observation 11. Suppose we have a system $\boldsymbol{a}^k \boldsymbol{w} > b^k$ with $1 \leqslant k \leqslant p$ where all b^k are nonzero. Multiply each inequality by an appropriate constant so that all right-hand sides are equal to 1. By replacing all right-hand sides constants 1 by a variable w_0, we get a system with either $\boldsymbol{a}^k \boldsymbol{w} > w_0$ type or $\boldsymbol{a}^k \boldsymbol{w} < w_0$ type inequalities. Clearly, any solution of this new system such that $w_0 > 0$ gives a solution of the original system. Thus by adding a large enough number of copies of $w_0 > 0$ the two problems are guaranteed to be equivalent.

In order to complete the reduction in [7], we just apply this technique to the system consisting of $\boldsymbol{a}\boldsymbol{w} + \delta > 1/(2L)$ inequalities and a large enough number of copies of $\delta < 1/L$.

It is worth noting that the same argument can be used to show that (nonsymmetric) MIN MISCLASSIFICATIONS cannot be approximated within $2^{\log^{1-\varepsilon} n}$, for any $\varepsilon > 0$, unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{\mathrm{polylog}\, n})$.

A special case of MIN RVLS with inequalities is also of particular interest in discriminant analysis and machine learning. The problem occurs when, given a linearly separable training set T, we want to minimize the number of parameters $w_j, 1 \leqslant j \leqslant n$, that are required to correctly classify all examples in T [48, 50, 64]. This objective plays a crucial role because it has been shown theoretically and experimentally that the number of nonzero parameters has a strong impact on the performance of the classifier (perceptron) for unseen data. According to Occam's principle, among all models that account for a given set of data, the simplest ones – with the smallest number of free parameters – are more likely to exhibit good generalization (see for instance

[6, 46]). The problem of identifying a subset of most relevant features is well known in the statistical discriminant analysis literature under the name of variable selection [53].

In practice, when a linear classifier (perceptron) cannot satisfactorily classify the training set based on the original n_0 features, new features derived from the original ones are added. For instance, the $O(n_0^d)$ higher-order products of the original features may be included for several values of $d \geqslant 2$ [35]. Other simple functions of the original features such as radial basis functions are also frequently introduced [35]. Since any training set can be correctly classified given enough additional features, the objective is to minimize the overall number of features that are actually used.

Min Relevant Features: Given a training set $T = \{a^k\}_{1 \leqslant k \leqslant p} \subset \mathbb{R}^n$ containing p labeled examples, find a hyperplane defined by $(w, w_0) \in \mathbb{R}^{n+1}$ that is consistent with T and has as few nonzero parameters $w_j, 1 \leqslant j \leqslant n$, as possible.

While Lin and Vitter showed that Min Relevant Features with binary inputs is NP-hard [50], van Horn and Martinez established that the symmetric variant with strict inequalities is at least as hard to approximate as Min Set Cover [63, 64]. Furthermore, they showed that an approximation algorithm that also minimizes the number of nonzero parameters within a factor of $O(\log p)$ would require far fewer examples to achieve a given level of accuracy than any algorithm that does not minimize the number of relevant features. More precisely, for such an Occam algorithm the number of training examples needed to learn in Valiant's Probably Approximately Correct (PAC) sense [62] would be almost linear in the minimum number of nonzero parameters s. If $s \ll n$ this is much less than the $O(n)$ examples required by a simplistic training procedure without feature minimization.

The following result provides strong evidence that no such approximation algorithm exists.

Theorem 12. Min Relevant Features *cannot be approximated within any constant, unless* $\mathrm{P} = \mathrm{NP}$, *and within a factor of* $2^{\log^{1-\varepsilon} p}$, *for any* $\varepsilon > 0$, *unless* $\mathrm{NP} \subseteq \mathrm{DTIME}(p^{\mathrm{polylog}\, p})$, *where p is the number of examples.*

Proof. To show nonapproximability within any constant factor, we adapt the reduction from Min Set Cover used for Min $\mathrm{ULR}^=$ in [8]. In Min Set Cover, given a collection $\mathscr{C} = \{C_1, \ldots, C_n\}$ of subsets of a finite set S, one seeks a sub-collection $\mathscr{C}' = \{C_{j_1}, \ldots, C_{j_m}\} \subseteq \mathscr{C}$ of minimum cardinality such that $\bigcup_{i=1}^m C_{j_i} = S$ with $m \leqslant n$. Any such $\mathscr{C}'$ is a *cover* of S. If all the sets in $\mathscr{C}'$ are pairwise disjoint, it is an *exact cover*.

According to [11], for every $c > 1$, there exists a polynomial-time reduction that transforms any instance ϕ of the satisfiability problem Sat (see [27]) into an instance of Min Set Cover with a positive integer K such that

- if ϕ is satisfiable there exists an exact cover $\mathscr{C}'$ of size K,
- if ϕ is unsatisfiable no set cover has size less than $\lfloor c \cdot K \rfloor$.

By construction, the size of the ground set S and the number of subsets are polynomially related.

For any such instance $(S, \mathscr{C})$ of Min Set Cover, we construct $|S|$ positive examples with n components corresponding to the system

$$A\mathbf{w} > w_0 \mathbf{1}, \tag{10}$$

where $a_{ij} = 1$ if the ith element of S belongs to C_j and 0 otherwise. Furthermore, we include the negative example $\mathbf{0}$ ensuring that $w_0 \geqslant 0$. Hence we have a training set with $p = |S| + 1$ examples.

Clearly, the nonzero components of any parameter vector $\mathbf{w} \in \mathbb{R}^n$ that correctly classifies all examples define a cover. Conversely, given any cover $\mathscr{C}'$ of cardinality K, the parameter vector $\mathbf{w}$ given by

$$w_j = \begin{cases} 1 & \text{if } C_j \in \mathscr{C}', \\ 0 & \text{otherwise,} \end{cases}$$

together with the bias $w_0 = 0$ correctly classifies all examples and has K nonzero components. Therefore the minimum number of nonzero parameters is either K or at least $\lfloor c \cdot K \rfloor$ depending on whether the corresponding Sat instance is satisfiable or unsatisfiable.

The constant gap between the satisfiable and unsatisfiable cases can then be increased by self-improvement like in the proof of Theorem 7. Since in the reduction the number of examples p is polynomially related to the size of the examples n, the same nonapproximability bound is also valid with respect to p. □

Note that, while this paper was being reviewed, Grigni et al. [32] addressed the parameterized complexity of designing linear classifiers with a number of nonzero parameters smaller or equal to a given bound.

The consequences of Theorem 12 on the hardness of designing compact feedforward networks are discussed in detail in [3]. From an artificial neural network perspective, Theorem 12 shows that designing close-to-minimum size networks in terms of nonzero weights is very hard even for linearly separable training sets that are performable by the simplest type of networks, namely perceptrons. Clearly, the general problem for multilayer networks is at least as hard as Min Relevant Features. Since our result holds for perceptrons, i.e. single units, the problem of designing compact networks does not become easier even if we know in advance the number of units in each layer of a minimum size network and we only need to find an appropriate set of values for the weights.

It is worth noting that Kearns and Valiant established in [45] a stronger nonapproximability bound but under a stronger cryptographic assumption. In particular, they showed that if trapdoor functions [1] exist it is intractable to find a feedforward network with a bounded number of layers that performs a given training set and that is at most polynomially larger than the minimum possible one. The size is there measured in terms

[1] A trapdoor function $\mathscr{T}$ is a one-to-one function such that $\mathscr{T}$ and its inverse are easy to evaluate but, given $\mathscr{T}$, the inverse $\mathscr{T}^{-1}$ cannot be constructed in polynomial time [58].

of the number of bits needed to describe the network. Although their result indicates that even approximating minimum networks within polynomial ratios is intractable, it leaves open the possibility that this strong nonapproximability bound depends on the fact that intricate networks with a large number of hidden layers may be considered. Indeed, the target functions that Kearns and Valiant proved hard to learn are the very special inverses of trapdoor functions.

From a practical point of view, our lower bound implies that the best we can do even for the simplest type of networks and of tasks is to devise efficient heuristics with good average-case behavior.

7. Conclusions

The various versions of MIN ULR$^{\mathscr{R}}$ and MIN RVLS$^{\mathscr{R}}$ with $\mathscr{R} \in \{=, \geqslant, >, \neq\}$ that we have considered are obtained by restricting the range of the variables and of the coefficients or by assigning a weight to each relation.

Table 2 summarizes the nonapproximability results that hold for MIN ULR variants unless $\mathrm{P} = \mathrm{NP}$. The results are valid for inhomogeneous systems with integer coefficients and no pairs of identical relations, and some of them are still valid for homogeneous systems with ternary, and even binary, coefficients. In order to avoid trivial solutions in the equality and nonstrict inequality cases, we require that at least a small

Table 2
Main approximability results for MIN ULR variants with $\mathscr{R} \in \{=, \geqslant, >, \neq\}$ that hold assuming $\mathrm{P} \neq \mathrm{NP}$

<table>
<tr><th></th><th>Real variables</th><th>Binary variables</th></tr>
<tr><td>MIN ULR$^{=}$
MIN ULR$^{\geqslant}$</td><td>Not within any constant [7]</td><td rowspan="8">NPO PB-compete</td></tr>
<tr><td>MIN ULR$^{\neq}$</td><td>Trivial</td></tr>
<tr><td>C MIN ULR$^{\geqslant;\geqslant}$</td><td>As hard as MIN ULR$^{\geqslant}$</td></tr>
<tr><td>C MIN ULR$^{\geqslant;>}$</td><td>At least as hard as MIN ULR$^{>}$</td></tr>
<tr><td>C MIN ULR$^{\geqslant;=}$</td><td rowspan="2">At least as hard as MIN ULR$^{\geqslant}$</td></tr>
<tr><td>C MIN ULR$^{\neq;\geqslant}$</td></tr>
<tr><td>C MIN ULR$^{\neq;=}$</td><td>MIN DOMINATING SET-hard</td></tr>
<tr><td>C MIN ULR$^{\mathscr{R};\neq}$</td><td>Polynomial time</td></tr>
</table>

Note: The constrained versions C MIN ULR$^{=;\mathscr{R}}$ with real variables and mandatory equations are equivalent to the corresponding MIN ULR$^{\mathscr{R}}$. The table is still valid when nonstrict inequalities are substituted by strict inequalities and vice versa. The only difference is that C MIN ULR$^{\geqslant;=}$ is as hard as MIN ULR$^{\geqslant}$ while C MIN ULR$^{\geqslant;=}$ is only at least as hard as MIN ULR$^{>}$.

fraction of the variables occurring in at least a small fraction of the satisfied relations are nonzero.

Arora et al. showed that MIN ULR with equalities or inequalities is not approximable within any constant, unless P = NP, and within a factor of $2^{\log^{1-\varepsilon} n}$, for any $\varepsilon > 0$, unless NP $\subseteq$ DTIME($n^{\mathrm{polylog}\, n}$) [7, 8]. Using a simple reduction from MIN DOMINATING SET, we have obtained a weaker but more likely logarithmic lower bound for MIN ULR with strict and nonstrict inequalities.

The weighted and constrained variants of MIN ULR turn out to be equally hard and, respectively, about as hard to approximate as the unweighted ones. Restricting the variables to binary values makes all versions of MIN ULR NPO PB-complete. Although the basic version of MIN ULR$^{\neq}$ is trivial, various constrained variants are hard to approximate. The nonapproximability bounds such as $n^{1-\varepsilon}$ for any $\varepsilon > 0$ makes the existence of any nontrivial approximation algorithm extremely unlikely.

It is worth noting that the overall situation for MIN ULR differs considerably from that for the complementary class of problems MAX FLS (see [3, 5]). Unlike for MAX FLS, MIN ULR with equations and (nonstrict) inequalities are equivalent to approximate. Moreover, while all basic versions of MIN ULR can be approximated within a factor of $n + 1$, MAX FLS$^{=}$ cannot be approximated within p^{ε} for some $\varepsilon > 0$, where p is the number of equations.

As to MIN RVLS$^{\mathscr{R}}$ with $\mathscr{R} \in \{=, \geqslant, >, \neq\}$, we have shown that they cannot be approximated within a constant factor and within $2^{\log^{1-\varepsilon} n}$ under the usual assumption. Note that, in spite of the close relationship between MIN RVLS and MIN ULR, MIN RVLS$^{\neq}$ is hard to approximate while MIN ULR$^{\neq}$ is trivially solvable. When the variables are restricted to take binary values, MIN RVLS turns out to be NPO PB-complete for any type of relational operator.

Finally, we have shown that the interesting special case MIN RELEVANT FEATURES, arising when designing linear classifiers and compact perceptrons, is not approximable within a logarithmic factor as conjectured in [63], unless all problems in NP are solvable in quasi-polynomial time.

Acknowledgements

The authors thank Peter Jonsson for a discussion on the relationship between MIN ULR and the problem of finding optimum acyclic subgraphs of directed graphs as well as David Shmoys for pointing out [61]. They are also grateful to the anonymous referees for helpful comments, especially regarding the first part of the proof of Theorem 7.

References

[1] E. Amaldi, On the complexity of training preceptrons, in: T. Kohonen, K. Mäkisara, O. Simula, J. Kangas (Eds.), Artificial Neural Networks, Elsevier, Amsterdam, 1991, pp. 55–60.

[2] E. Amaldi, Complexity of problems related to training perceptrons, in: M. Chaleyat-Maurel, M. Cottrell, J.-C. Fort (Eds.), Actes du congrès "Aspects théoriques des réseaux de neurones", Congrès Européen de Mathématiques, Paris, 1992.
[3] E. Amaldi, From finding maximum feasible subsystems of linear systems to feedforward neural network design, Ph.D. Thesis, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1994.
[4] E. Amaldi, V. Kann, On the approximability of removing the smallest number of relations from linear systems to achieve feasibility, Technical Report ORWP-6-94, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1994.
[5] E. Amaldi, V. Kann, The complexity and approximability of finding maximum feasible subsystems of linear relations, Theoret. Comput. Sci. 147 (1995) 181–210.
[6] D. Angluin, C.H. Smith, Inductive inference: Theory and methods, ACM Comput. Surveys 15 (1983) 237–269.
[7] S. Arora, L. Babai, J. Stern, Z. Sweedyk, The hardness of approximate optima in lattices, codes, and systems of linear equations, in: Proc. 34th Ann. IEEE Symp. on Foundations of Comput. Sci., IEEE Comput. Soc. Press, Silver Spring, MD, 1993, pp. 724–733.
[8] S. Arora, L. Babai, J. Stern, Z. Sweedyk, The hardness of approximate optima in lattices, codes, and systems of linear equations, J. Comput. System Sci. (1997) 317–331.
[9] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy, Proof verification and hardness of approximation problems, in: Proc. 33rd Ann. IEEE Symp. on Foundations of Comput. Sci., IEEE Comput. Soc. Press, Silver Spring, MD, 1992, pp. 14–23.
[10] M. Bellare, S. Goldwasser, C. Lund, A. Russell, Efficient probabilistically checkable proofs and applications to approximation, in: Proc. 25th Ann. ACM Symp. on Theory of Comp., ACM, New York, 1993, pp. 294–304.
[11] M. Bellare, O. Goldreich, M. Sudan, Free bits, PCPs and non-approximability – towards tight results, in: Proc. 36th Ann. IEEE Symp. on Foundations of Comput. Sci., IEEE Comput. Soc. Press, Silver Spring, MD, 1995, pp. 422–431.
[12] K.P. Bennett, O.L. Mangasarian, Robust linear programming discrimination of two linearly inseparable sets, Optimization Meth. Software 1 (1992) 23–34.
[13] J. Chinneck, An effective polynomial-time heuristic for the minimum-cardinality IIS set-covering problem, Ann. Math. Artificial Intelligence 17 (1996) 127–144.
[14] J. Chinneck, Feasibility and viability, in: T. Gal, H.J. Greenberg (Eds.) Advances in sensitivity analysis and parametric programming, Kluwer Academic Publishers, Dordrecht, 1997.
[15] J. Chinneck, E. Dravnieks, Locating minimal infeasible constraint sets in linear programs, ORSA J. Comput. 3 (1991) 157–168.
[16] V. Chvatal, Linear programming, Freeman, New York, 1983.
[17] P. Crescenzi, V. Kann, A compendium of NP optimization problems, Technical Report SI/RR-95/02, Dipartimento di Scienze dell'Informazione, Università di Roma "La Sapienza", 1995. The list is updated continuously. The latest version is available by anonymous ftp from `nada.kth.se` as `Theory/Viggo-Kann/compendium.ps.Z`.
[18] P. Crescenzi, V. Kann, R. Silvestri, L. Trevisan, Structure in approximation classes, in: Proc. 1st Internat. Conf. on Computing and Combinatorics, Lecture Notes in Computer Science, vol. 959, Springer, Berlin, 1995, pp. 539–548.
[19] P. Crescenzi, A. Panconesi, Completeness in approximation classes, Inform. and Comput. 93 (2) (1991) 241–262.
[20] P. Crescenzi, R. Silvestri, L. Trevisan, To weight or not to weight: Where is the question? in: Proc. 4th Israel Symp. on Theory of Computing and Systems, IEEE Comput. Soc. Press, Silver Spring, MD, 1996, pp. 68–77.
[21] R.O. Duda, P.E. Hart, Pattern classification and scene analysis, Wiley, New York, 1973.
[22] J. Edmonds, Redundancy and Helly for linear programming, Lecture Notes, 1994.
[23] G. Even, J. Naor, B. Schieber, M. Sudan, Approximating minimum feedback sets and multi-cuts in directed graphs, in: Proc. 4th Internat. Conf. on Integer Prog. and Combinatorial Optimization, Lecture Notes in Computer Science, vol. 920, Springer, Berlin, 1995, pp. 14–28.
[24] U. Feige, A threshold of $\ln n$ for approximating set cover, in: Proc. 28th Ann. ACM Symp. on Theory of Comp., ACM, New York, 1996, pp. 314–318.
[25] M.R. Frean, A thermal perceptron learning rule, Neural Computation 4 (1992) 946–957.

[26] S.I. Gallant, Perceptron-based learning algorithms, IEEE Trans. Neural Networks 1 (1990) 179–191.
[27] M.R. Garey, D.S. Johnson, Computers and Intractability: A guide to the theory of NP-completeness, W.H. Freeman and Company, San Francisco, 1979.
[28] J. Gleeson, J. Ryan, Identifying minimally infeasible subsystems of inequalities, ORSA J. Computing 3 (1990) 61–63.
[29] H.J. Greenberg, Consistency, redundancy, and implied equalities in linear systems, Ann. Math. Artificial Intelligence 17 (1996) 37–83.
[30] H.J. Greenberg, F.H. Murphy, Approaches to diagnosing infeasible linear programs, ORSA J. Computing 3 (1991) 253–261.
[31] R. Greer, Trees and hills: Methodology for maximizing functions of systems of linear relations, volume 22 of Annals of Discrete Mathematics, Elsevier, Amsterdam, 1984.
[32] M. Grigni, V. Mirelli, C.H. Papadimitriou, On the difficulty of designing good classifiers, in: Proc. 2nd Internat. Conf. on Computing and Combinatorics, Lecture Notes in Computer Science, vol. 1090, Springer, Berlin, 1996, pp. 273–279.
[33] O. Guieu, J. Chinneck, Analyzing infeasible mixed-integer and integer linear programs, Technical Report Technical Report SCE-96-05, Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada, 1996.
[34] M.M. Halldóroson, Approximating the minimum maximal independence number, Inform. Process. Lett. 46 (1993) 169–172.
[35] M. Hassoun, Fundamentals of artificial neural networks, MIT Press, Cambridge, MA, 1995.
[36] J. Håstad, Clique is hard to approximate within $n^{1-\varepsilon}$, in: Proc. 37th Ann. IEEE Symp. on Foundations of Comput. Sci., IEEE Comput. Soc. Press, Silver Spring, MD, 1996, pp. 627–636.
[37] J. Håstad, Some optimal inapproximability results, in: Proc. 29th Ann. ACM Symp. on Theory of Comp, ACM, New York, 1997, pp. 1–10.
[38] K-U. Höffgen, H-U. Simon, K. van Horn, Robust trainability of single neurons, J. Comput. System Sci. 50 (1995) 114–125.
[39] D.S. Johnson, F.P. Preparata, The densest hemisphere problem, Theoret. Comput. Sci. 6 (1978) 93–107.
[40] V. Kann, On the Approximability of NP-complete Optimization Problems, Ph.D. Thesis, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, 1992.
[41] V. Kann, Polynomially bounded minimization problems that are hard to approximate, Nordic J. Computing 1 (1994) 317–331.
[42] V. Kann, Strong lower bounds on the approximability of some NPO PB-complete maximization problems, in: Proc. 20th Ann. Mathematical Foundations of Comput. Sci., Lecture Notes in Computer Science, vol. 969. Springer, Berlin, 1995, pp. 227–236.
[43] V. Kann, A. Panconesi, Hardness of approximation, in: M. Dell'Amico, F. Maffioli, S. Martello (Eds.), Annotated Bibliographies in Combinatorial Optimization, Ch. 2, Wiley, New York, 1997.
[44] N. Karmarkar, A new polynomial time algorithm for linear programming, Combinatorica 4 (1984) 373–395.
[45] M. Kearns, L. Valiant, Cryptographic limitations on learning boolean formulae and finite automata, J. ACM 41 (1994) 67–95.
[46] M. Kearns, U. Vazirani, An introduction to Computational Learning Theory. MIT Press, Cambridge, MA, 1994.
[47] S. Khanna, M. Sudan, L. Trevisan, Constraint satisfaction: The approximability of minimization problems, in: Proc. 12th Annual IEEE Conf. Comput. Complexity. IEEE Comput. Soc., Press, Silver Spring, MD, 1997, pp. 282–296.
[48] G.J. Koehler, Linear discriminant functions determined by genetic search, ORSA J. Computing 3 (1991) 345–357.
[49] G.J. Koehler, S.S. Erenguc, Minimizing misclassifications in linear discriminant analysis, Decision Sci. 21 (1990) 63–85.
[50] J.-H. Lin, J.S. Vitter, Complexity results on learning by neural nets, Mach. Learning 6 (1991) 211–230.
[51] C. Lund, M. Yannakakis, On the hardness of approximating minimization problems, J. ACM 41 (1994) 960–981.
[51A] O.L. Mangasarian, Misclassification minimization, J. Global Optimization 5 (1994) 309–323.
[52] M. Marchand, M. Golea, On learning simple neural concepts: from halfspace intersections to neural decision lists, Network: Computation in Neural Systems 4 (1993) 67–85.

[53] G.J. McLachlan, Discriminant analysis and statistical pattern recognition. Wiley, New York, 1992.

[54] M.L. Minsky, S. Papert, Perceptrons: An introduction to computational Geometry, MIT Press Cambridge, MA, 1988, expanded edition.

[55] C.H. Papadimitriou, K. Steiglitz, Combinatorial optimization, Algorithms and Complexity, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[56] M. Parker, A set covering approach to infeasibility analysis of linear programming problems and related issues, Ph.D. Thesis, Department of Mathematics, University of Colorado, Denver, 1995.

[57] M. Parker, J. Ryan, Finding the minimum weight IIS cover of an infeasible system of linear inequalities, Ann. Math. Artificial Intelligence 17 (1996) 107–126.

[58] R. Rivest, Cryptography, in: J. van Leeuwen (Ed.), Handbook of Theoretical Computer Science, Vol. A: Algorithms and complexity, Elsevier, Amsterdam, 1990, pp. 717–755.

[59] J. Sankaran, A note on resolving infeasibility in linear programs by constraint relaxation, Oper. Res. Lett. 13 (1993) 19–20.

[60] A. Schrijver, Theory of linear and integer programming, Interscience series in discrete mathematics and optimization, Wiley, New York, 1986.

[61] P. Seymour, Packing directed circuits fractionally, Combinatorica 15 (1995) 281–288.

[62] L.G. Valiant, A theory of the learnable, Communications of ACM 27 (1984) 1134–1142.

[63] K. van Horn, T. Martinez, The minimum feature set problem, Technical Report CS-92-8, Computer Science Department, Brigham Young University, Provo, 1992.

[64] K. van Horn, T. Martinez, The minimum feature set problem, IEEE Trans. Neural Networks 7 (1994) 491–494.

[65] R.E. Warmack, R.C. Gonzalez, An algorithm for optimal solution of linear inequalities and its application to pattern recognition, IEEE Trans. Comput. 22 (1973) 1065–1075.