Electronic Colloquium on Computational Complexity - Reports Series 1996 - available via:
ECCC
TR96-015
FTP: ftp.eccc.uni-trier.de:/pub/eccc/
WWW: http://www.eccc.uni-trier.de/eccc/
Email: ftpmail@ftp.eccc.uni-trier.de with subject 'help eccc'

On the approximability of some NP-hard minimization problems for linear systems

Edoardo Amaldi*
School of Operations Research
and Center for Applied Mathematics
Cornell University
Ithaca, NY 14853
amaldi@cs.cornell.edu

Viggo Kann
Department of Numerical Analysis
and Computing Science
Royal Institute of Technology
S-100 44 Stockholm
viggo@nada.kth.se

Abstract

We investigate the computational complexity of two classes of combinatorial optimization problems related to linear systems and study the relationship between their approximability properties. 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 $=$, $\geq$, $>$ and $\neq$ are considered. While Min RVLS with equations was known to be NP-hard in [27], we established in [2, 6] 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 [8]. 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 bounded discrete 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. Two interesting special cases of Min RVLS and Min ULR that arise in discriminant analysis and machine learning are also discussed. In particular, we disprove a conjecture presented in [57] regarding the existence of a polynomial time algorithm to design linear classifiers (or perceptrons) that use a close-to-minimum number of features.

Keywords: Linear relations, feasible subsystems, computational complexity, approximability, Min Dominating Set-hard problems, NPO PB-complete problems, designing linear classifiers, training perceptrons, relevant feature minimization.

*E.A. is partially supported by a Postdoctoral fellowship of the Swiss National Science Foundation. Research mainly done while the author was at the Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne.

1 Introduction

The first class of problems we consider is that of finding the 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$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$: Given a linear system $A\mathbf{x}\,\mathcal{R}\,\mathbf{b}$ with rational coefficients and with a $p \times n$ matrix A, find a solution $\mathbf{x} \in \mathbb{R}^n$ which violates as few relations as possible while satisfying all the others.

Many variants of these combinatorial problems arise in various fields such as operations research [35, 31, 30], pattern recognition [59, 21, 43] and machine learning [1, 34, 46]. It is well known that feasible linear systems with equations or inequalities can be solved in polynomial time using an adequate linear programming method [39]. But for infeasible systems, least mean squared methods, which are suited to linear regression, are not appropriate when the objective is to minimize the number of unsatisfied relations.

A number of algorithms have been proposed for tackling various versions of Min ULR. The weighted variants, in which a weight is associated with each equation and the goal is to minimize the total weight of the unsatisfied relations, are often considered. Johnson and Preparata showed that the special cases of Min ULR$^{>}$ and Min ULR$^{\geq}$ with homogeneous systems are NP-hard and devised a complete enumeration method which is also applicable to the weighted and mixed variants [35]. 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 [43, 13] 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$^{\geq}$. Increasingly sophisticated models have been proposed in order to try to avoid unacceptable or trivial solutions (see [13]).

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 their number of errors. While some heuristic algorithms were devised in [26, 25, 4], Amaldi showed that solving these problems to optimality is NP-hard even when restricted to perceptrons with bipolar inputs -1 or 1 [1]. In [34] 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. With the new software and hardware advances, dealing with infeasible programs is becoming a major bottleneck in linear programming. Several methods have been proposed in order to try to locate the source of infeasibility. While the first ones looked for minimal infeasible subsystems [28, 15], the later ones aim at removing as few constraints as possible to achieve feasibility [29, 53, 51, 50, 14]. In fact, 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 [51, 50].

The second class of combinatorial 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$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$: Given a feasible system of linear relations $A\mathbf{x}\,\mathcal{R}\,\mathbf{b}$ 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$^>$ and Min RVLS$^\geq$ 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 [42, 57]. 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 [7, 41]. In [42] a genetic search strategy has been proposed for designing optimal linear classifiers with as few nonzero parameters as possible. The twofold goal of this heuristic technique is to look among the linear classifiers with a minimum number of misclassifications for one which uses a smallest subset of attributes. The problem of identifying a subset of most relevant attributes is well known in the statistical literature and is referred to as variable selection [47].

Since the late 80's there have been new substantial progresses in the study of the approximability of NP-hard optimization problems. Various classes have been defined and different reductions preserving approximability have been used to compare the approximability of optimization problems (see [36]). Moreover, the striking results which have been obtained in the area of interactive proofs triggered new advances in computational complexity theory. Strong bounds were derived on the approximability of several famous problems like maximum independent set, minimum graph coloring and minimum set cover [10, 45, 12, 11]. These results have also important consequences on the approximability of other optimization problems. For a list of the currently best approximability upper and lower bounds for NP optimization problems, see [17].

In [6] we performed a thorough study of the approximability of the complementary problems of Min ULR, named Max FLS, where one looks for minimum Feasible subsystems of Linear Systems. In particular, we showed that the basic versions with $=$, $\geq$ 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 intractability of the basic versions of Min ULR, we are interested in polynomial time algorithms that are guaranteed to provide near-optimal solutions. Such approximation algorithms are of great practical value because optimal solutions are rarely required. Although complementary pairs of problems such as Min ULR and Max FLS are equivalent to solve optimally, their approximability properties can differ enormously. This is, for instance, the case for the minimum node cover and the maximum independent set problems [27].

By delving into the interactive proof connection, Arora, Babai, Stern and Sweedyk established 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($n^{\text{polylog}\, n}$) [8] (see also [9]). Moreover, they noted that this non-approximability 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 [57, 58] 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 dominating set 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 $O(\log p)$ [57].

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 study the approximability of Min RVLS and show that our bounds for Min RVLS$^=$ directly imply those established by Arora et al. for Min ULR. In Section 4 we recall the known results about the complexity of solving the basic versions of Min ULR and determine alternative lower and upper bounds on the approximability of the inequality cases. We also discuss two important variants: the weighted ones where a different importance may be assigned to each relation as well as the constrained ones where some relations are mandatory while others are optional. Section 5 is devoted to Min RVLS and Min ULR versions where the variables are restricted to take a finite number of discrete values. In Section 6 we focus on two interesting special cases of Min RVLS and Min ULR with inequalities arising 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 [57]. The various results are summarized in Section 7. Finally, the relationship between Min ULR with real and binary variables is pointed out in the Appendix.

An earlier version of this paper appeared as a technical report [5].

2 Approximability of minimization problems

An NP optimization (NPO) problem over an alphabet Σ is a four-tuple $\Pi = (\mathcal{I}_\Pi, S_\Pi, f_\Pi, opt_\Pi)$ where $\mathcal{I}_\Pi \subseteq \Sigma^*$ is the set of *instances*, $S_\Pi(I) \subseteq \Sigma^*$ is the set of *feasible solutions* for instance $I \in \mathcal{I}_\Pi$, $f_\Pi : \mathcal{I}_\Pi \times \Sigma^* \to \mathbb{N}$, 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 $R_\Pi(I,x) = f_\Pi(I,x)/opt_\Pi(I)$. 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 $\mathcal{A}$ such that for every $n \in \mathbb{Z}^+$ and for all instances $I \in \mathcal{I}_\Pi$ with $|I| = n$ we have that $\mathcal{A}(I) \in S_\Pi(I)$ and $R_\Pi(I, \mathcal{A}(I)) \leq p(n)$.

Although various reductions preserving approximability within constants have been proposed (see [36]), we will use the S-reduction which is suited to relate problems that cannot be approximated within any constant.

Definition 1 [37] 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 monotonously increasing positive function and c is a positive constant.

ii) $t_1 : \mathcal{I}_\Pi \to \mathcal{I}_{\Pi'}$ and $\forall I \in \mathcal{I}_\Pi$ and $\forall x \in S_{\Pi'}(t_1(I))$, $t_2(I,x) \in S_\Pi(I)$.

iii) $\forall I \in \mathcal{I}_\Pi$ and $\forall x \in S_{\Pi'}(t_1(I))$, $R_\Pi(I, t_2(I,x)) \leq c \cdot R_{\Pi'}(t_1(I), x)$.

iv) $\forall I \in \mathcal{I}_\Pi, |t_1(I)| \leq a(|I|)$.

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 monotonously 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 poly-logarithmic 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

$$\forall I \in \mathcal{I}_\Pi \, \forall x \in S_\Pi(I), \, f_\Pi(I,x) \leq p(|I|).$$

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

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 [45] Lund and Yannakakis established, using results from interactive proofs, a lower bound on the approximability of Min Set Covering and of several other problems, such as Min Dominating Set, that are equivalent from the approximation point of view. In [12] Bellare et al. improved this result by showing, among others, that Min Set Covering 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 Covering 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 minimum dominating set 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 [32] that Min Ind Dom Set is very hard to approximate. Assuming P$\neq$NP, this problem 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 Halldórsson's 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 for NPO PB in the sense that every polynomially bounded NPO problem can be reduced to it using an approximation preserving reduction [37, 18].

The purpose of this paper is to investigate the approximability properties of the different variants of Min RVLS and Min ULR.

3 Approximability of Min RVLS

In this section we investigate the approximability of the basic Min RVLS$^\mathcal{R}$ variants with $\mathcal{R} \in \{=, \geq, >, \neq\}$ and discuss their relationship with that of the corresponding Min ULR$^\mathcal{R}$ variants.

Proposition 1 Min RVLS$^=$ *cannot be approximated within any constant factor unless* P=NP.

Proof By reduction from a variant of Min Set Covering with disjoint sets. In Min Set Covering, given a collection $\mathcal{C} = \{C_1, \ldots, C_n\}$ of subsets of a finite set S, one seeks a subcollection $\mathcal{C}' = \{C_{j_1}, \ldots, C_{j_m}\} \subseteq \mathcal{C}$ of minimum cardinality such that $\cup_{i=1}^{m} C_{j_i} = S$ with $m \leq n$. Any such $\mathcal{C}'$ is a *cover* of S. If all the sets in $\mathcal{C}'$ are pairwise disjoint, it is an *exact cover*.

The proof is based on the following result by Bellare et al. on exact covers [12]. 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 COVERING with a positive integer K such that

- if ϕ is satisfiable there exists an exact cover $\mathcal{C}'$ 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 $(S, \mathcal{C})$ of MIN SET COVERING, we can construct a system with n variables and $|S|$ equations

$$A\mathbf{x} = \mathbf{1}, \tag{1}$$

where $a_{ij} = 1$ if the ith element of S belongs to C_j and 0 otherwise. $\mathbf{1}$ denotes the $|S|$-dimensional vector with all 1 components. Clearly, the nonzero variables in any solution $\mathbf{x}$ of the above system define a set cover. Conversely, given any exact cover $\mathcal{C}'$ of cardinality K, the vector $\mathbf{x}$ given by

$$x_j = \begin{cases} 1 & \text{if } C_j \in \mathcal{C}' \\ 0 & \text{otherwise} \end{cases}$$

satisfies all equations and has K nonzero variables. Thus the minimum number of nonzero variables in a solution of $A\mathbf{x} = \mathbf{1}$ is either K or at least $\lfloor c \cdot K \rfloor$.

Note that while there is a one–to–one correspondence between the exact covers and the solutions of (1) with $0-1$ components, minimum cardinality covers do not necessarily correspond to a solution. For example, in the set covering instance associated with

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

all subcollections of cardinality 2 are minimum (nonexact) covers but the system has the unique solution $(1/2, 1/2, 1/2)$ with three nonzero variables. For infeasible systems like

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

the minimum number of nonzero variables can be considered as larger than n. $\square$

This result can be considerably strengthened.

Theorem 2 *Assuming* NP $\not\subseteq$ DTIME$(n^{\text{polylog}\, n})$, MIN RVLS$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{=, \geq, >, \neq\}$ *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 proceed by self-improvement as in [9]. The idea is to consider the reduction in Proposition 1 from MIN SET COVERING to MIN RVLS$^{=}$ with a fixed constant gap $c > 1$ between the satisfiable and unsatisfiable cases and to increase it recursively. For any particular MIN SET COVERING instance, we start with the corresponding system (1) whose coefficients are 0 or 1 and we construct the squared system

$$A'\mathbf{x} = \mathbf{1}' \tag{2}$$

obtained by replacing each 1 coefficient of A by the whole matrix A and each 0 coefficient by the $p \times n$ matrix with all 0. Thus A' is a matrix of size $p^2 \times n^2$ and $\mathbf{1}'$ is a vector with p^2 components all equal to 1. Clearly, if the corresponding instance ϕ of SAT is satisfiable there is a solution $\mathbf{x}$ of (2) with K^2 variables equal to 1 and all the other to 0. On the contrary, if ϕ is unsatisfiable any solution $\mathbf{x}$ that satisfies all equations of (2) has at least $\lfloor c^2 K^2 \rfloor$ nonzero variables.

By applying this 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$^=$ instance corresponding to the considered SAT instance ϕ and β is a positive real number. The construction requires $O(n^{\mathrm{polylog}\, n})$ 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 $c = 2$ implies a total gap of $c^{2^t} = c^{\log^\beta n} = 2^{\log^{\beta/(\beta+1)} n'}$.

The theorem follows by contradiction. Suppose there exists a polynomial time algorithm that approximates MIN RVLS$^=$ 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$^=$, one could decide in $O(n^{\mathrm{polylog}\, n})$ time whether any given instance ϕ of SAT is satisfiable. But this would imply NP $\subseteq$ DTIME$(n^{\mathrm{polylog}\, n})$.

The same bound is also valid for systems with strict and nonstrict inequalities because each equation in (1) can be replaced by an appropriate pair of inequalities.

For MIN RVLS$^{\neq}$ we proceed by cost preserving reduction from MIN DOMINATING SET [27]. 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 $v_i \in V$, $1 \le i \le n$, we consider the relation

$$x_i + \sum_{j \in N(v_i)} x_j \neq 0 \tag{3}$$

where j is included in $N(v_i)$ if and only if v_j is adjacent to v_i. Thus we have a system with n relations and n variables.

It is easily verified that there exists a dominating set in G of size at most s if and only if the corresponding system (3) has a solution $\mathbf{x}$ with at most s nonzero components. Given a dominating set $V' \subseteq V$ of size s, the vector $\mathbf{x}$ defined by

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

is, of course, a solution of the corresponding system with at most s nonzero components. Conversely, given any solution $\mathbf{x}$ with at most s nonzero components, the set of nodes associated with the nonzero components of $\mathbf{x}$ is clearly a dominating set of size s. Therefore, we get a system $A\mathbf{x} \neq \mathbf{0}$ that can be self-improved as the above MIN RVLS$^=$ systems. □

As we shall see in Section 6, these non-approximability bounds also hold for an important special case of MIN RVLS with inequalities that arises in discriminant analysis and machine learning.

In homogeneous MIN RVLS$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq\}$, we are obviously not interested in the trivial solution with all zero variables. The above reduction from MIN SET COVERING can be easily extended to the case of homogeneous systems by just multiplying all right hand sides (equal to 1) by a new variable x_0. Clearly, x_0 must be nonzero otherwise the trivial solution would be optimal.

In spite of the fact that MIN RVLS and MIN ULR deal with different types of systems (feasible versus infeasible), they turn out to be closely related in terms of approximation properties.

Proposition 3 MIN ULR$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{=, \geq, >\}$ *is at least as hard to approximate as* MIN RVLS$^{\mathcal{R}}$ *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 $1 \leq i \leq n$, the equation $x_i = 0$ and by eliminating variables in this system using the set of equations in the MIN RVLS$^=$ instance.

Since each equation $x_i = 0$ can be replaced by the two complementary inequalities $x_i \geq 0$ and $x_i \leq 0$, MIN ULR$^{\geq}$ is at least as hard to approximate as MIN RVLS$^{\geq}$. Also MIN ULR$^{>}$ is at least as hard because it is equivalent to MIN ULR$^{\geq}$ for systems with integer (rational) coefficients. Indeed, any system $A\mathbf{x} \leq \mathbf{b}$ has a solution if and only if the system $A\mathbf{x} < \mathbf{b} + \varepsilon\mathbf{1}$ has a solution, where $\varepsilon = 2^{-2L}$ and L is the size (in bits) of the binary encoded input instance [49]. □

Thus, Proposition 1 and Theorem 2 immediately imply the non-approximability of MIN ULR with equations and inequalities within the same factors established by Arora et al. in [9].

In fact, there exists also a reduction from MIN ULR$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq\}$ to MIN RVLS$^{\mathcal{R}}$ with the same type of relations. However, due to its size amplification, the non-approximability bound for MIN ULR given in [9] leads to a smaller bound for MIN RVLS than that of Theorem 2.

As we will see in the proof of Theorem 5, any instance of MIN ULR$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq\}$ can be turned into an equivalent instance of homogeneous MIN ULR$^{\mathcal{R}}$. Let $\sum_{j=1}^{n} a_{ij}x_j = 0$ with $1 \leq i \leq p$ be an instance of homogeneous MIN ULR$^=$. We construct an instance of MIN RVLS$^=$ consisting of $p(n+1)$ equations of the following type:

$$\sum_{j=1}^{n} a_{ij}x_j = y_{ik}$$

where $1 \leq i \leq p$, $1 \leq k \leq n+1$ and y_{ik} are $p(n+1)$ new variables. If s equations of $A\mathbf{x} = \mathbf{0}$ are satisfied, then we will get a solution of MIN RVLS$^=$ with between $s(n+1)$ and $s(n+1)+n$ variables equal to zero. Conversely, a solution of MIN RVLS$^=$ with between $s(n+1)$ and $s(n+1)+n$ variables equal to zero, will give us a solution of MIN ULR$^=$ with s satisfied equations. Since the instance of MIN RVLS$^=$ contains $n + p(n+1)$ variables, the reduction is an S-reduction with size amplification $O(pn)$ and a constant $c = 1$. By substituting $=$ with $\geq$, this is also valid for MIN RVLS$^{\geq}$.

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 [6].

Interestingly, MAX IVLS$^{\geq}$ and MAX IVLS$^{>}$ are much harder to approximate than MAX FLS$^{\geq}$ 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^ε for some $\varepsilon > 0$ unless P=NP, where n is the number of nodes), while the latter ones can be approximated within 2 [6]. It suffices to construct, for each edge $e = [v_i, v_j]$, the inequality $x_i + x_j \geq 1$ 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.

4 Approximability of Min ULR variants

In this section we discuss lower and upper bounds on the approximability of the basic versions of Min ULR and then focus on the weighted as well as constrained variants with any type of relations. In each case, we will try to find the simplest versions of these problems that are still hard. This is achieved by restricting the set of numbers used as coefficients and in the right hand side of the system.

Homogeneous systems are considered to have the simplest right hand sides. But, as for Min RVLS$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq\}$, we are not interested in trivial solutions where all variables occurring in the satisfied relations are zero. Even if we forbid the solution $\mathbf{x} = \mathbf{0}$, there might be other trivial solutions where almost all variables are zero except a few that only occur in a few relations. In order to rule out all these trivial solutions for the equality and nonstrict inequality case, we only consider solutions where the variable(s) occurring in the largest number of satisfied relations is (are) nonzero.

In [2, 6] we proved that Min ULR$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >\}$ is NP-hard even when restricted to homogeneous systems with bipolar coefficients in $\{-1, 1\}$. Sankaran showed in [53] 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$^{\geq}$ 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 $\mathbf{x} \in \mathbb{R}^n$ 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 and n the number of variables [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$^{\geq}$ involving total unimodularity is also mentioned in [53].

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

Fact 4 Min ULR$^{\geq}$ *is at least as hard to approximate as* Min ULR$^{=}$ *but not harder than* Min ULR$^{=}$ *with nonnegative variables.* Min ULR$^{\geq}$ *and* Min ULR$^{>}$ *with integer (rational) coefficients are equivalent.*

As noted also in [8], minimizing the number of unsatisfied equations in an arbitrary Min ULR$^{=}$ instance is clearly equivalent to minimizing the number of violated inequalities in the corresponding instance of Min ULR$^{\geq}$ where each equation is replaced by the two complementary inequalities.

Given an arbitrary instance of MIN ULR$^{\geq}$, replacing each variable x_i unrestricted in sign by the difference $x_i' - x_i''$ of two nonnegative variables $x_i', x_i'' \geq 0$ and adding a slack variable for each inequality leads to an equivalent system with p equations and $2n + p$ nonnegative variables.

As mentioned in Section 1, Arora et al. showed in [8] (see also [9]) 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($n^{\text{polylog}\, n}$), where n is the number of variables. Of course, this also holds for MIN ULR with strict and nonstrict inequalities.

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

Theorem 5 MIN ULR$^{\geq}$ *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 By cost preserving reduction from MIN DOMINATING SET as for MIN RVLS$^{\neq}$ and similar to [34]. For each node $v_i \in V$, $1 \leq i \leq n$, we consider the homogeneous inequality $x_i \geq 0$ and the inhomogeneous inequality

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

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 $\mathbf{x}$ that violates at most s inequalities of the corresponding system. Given a dominating set $V' \subseteq V$ of size s, the solution $\mathbf{x}$ defined by

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

satisfies all inhomogeneous inequalities and $n - s$ homogeneous ones. Conversely, given a solution vector $\mathbf{x}$ 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 $\mathbf{x}$. 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 $x_i + \sum_{j \in N(v_i)} x_j \leq -1$ 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 $\geq$ 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 MIN ULR$^{=}$ and MIN ULR$^{\geq}$ instance into a homogeneous one[1]. Given a system consisting of p relations, we first multiply all constant right hand sides by a new variable x_0. Then we only have to ensure that this constant is nonzero (for the equality case) or positive (for the inequality case). Therefore we introduce new relations $k \cdot x_0 = k \cdot y_0$ or $k \cdot x_0 \geq 0$, for $k \in [1..p+1]$, where y_0 is a new variable. In every solution that satisfies more than p relations, at least one of the new relations must be satisfied, and since all the new relations are satisfied simultaneously, x_0 will become the variable occurring in the largest number of satisfied equations. Hence x_0 is nonzero. For the inequality case this implies that $x_0 \geq 0$.

[1]This standard reduction will be omitted from now on.

Since the reduction is cost preserving and without amplification, we have exactly the same non-approximability bounds for MIN ULR$^{\geq}$ and MIN ULR$^{>}$ as for MIN DOMINATING SET. $\square$

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 those given in [8].

Unlike for MAX FLS$^{=}$ [6], for MIN ULR$^{=}$ we can guarantee in polynomial time a performance ratio that is linear in the number of variables. This fact was mentioned without proof in [8].

Proposition 6 MIN ULR$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{=, \geq, >\}$ *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 [54]), a system $A\mathbf{x} \leq \mathbf{b}$ with p inequalities and n variables is infeasible if and only if there exists a nonnegative vector $\mathbf{y} \geq \mathbf{0}$ such that $\mathbf{y}^t A = \mathbf{0}$ and $\mathbf{y}^t \mathbf{b} < 0$. In fact, the result is still valid if the vector $\mathbf{y} \geq \mathbf{0}$ is required to have at most $n+1$ nonzero components. For infeasible systems, a polynomial time LP algorithm produces a $\mathbf{y}$ satisfying Farkas' lemma. If $\mathbf{y}$ has more than $n+1$ nonzero components, some of them can be driven to zero. Therefore it suffices to find a nontrivial solution $\mathbf{z}$ of the auxiliary system $\mathbf{z}^t[A|\mathbf{b}] = [\mathbf{0}^t|0]$ such that $\mathbf{z}$ is zero for every component where $\mathbf{y}$ is zero. This simply amounts to determine a nontrivial solution to $n+1$ homogeneous equations in more than $n+1$ variables. Subtracting a multiple of $\mathbf{z}$ from $\mathbf{y}$ leads to a new $\mathbf{y}$ with fewer nonzero components. By repeating this process, we obtain in polynomial time a $\mathbf{y}$ 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. $\square$

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

Such an upper bound can be guaranteed in polynomial time 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).

Proposition 7 *For node-arc incidence matrices,* MIN ULR$^{\geq}$ *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.*

PROOF By straightforward modification of the polynomial time reduction from MIN FEEDBACK ARC SET to MIN ULR with $\leq$ relations given in [53]. For each arc (v_i, v_j) in a given instance of MIN FEEDBACK ARC SET, we consider the nonstrict inequality $x_i - x_j \geq 1$ 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 [36]), 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] (follows from [55]). □

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, 51].

Proposition 8 *Weighted* Min ULR$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{=, \geq, >\}$ *and positive integer (rational) weights are equally hard to approximate as the corresponding basic versions.*

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 by Crescenzi, Silvestri and Trevisan [20]: For any weighted "nice subset problem" that is approximable within a polynomial $r(n)$ in the size of the input, the restricted version of the same problem where the weights are 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 to 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 $\mathcal{R}_1, \mathcal{R}_2 \in \{=, \geq, >, \neq\}$ 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 [6], enforcing some mandatory relations makes Max FLS with inequalities harder to approximate. While Max FLS$^{>}$ and Max FLS$^{\geq}$ 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^{ε} for some $\varepsilon > 0$, where n is the instance size.

Any instance of a constrained problem C Min ULR$^{=;\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$ can be transformed into an equivalent instance of Min ULR$^{\mathcal{R}}$ by eliminating variables in the set of optional relations using the set of mandatory ones. Since Min ULR$^{\neq}$ is trivial, all the problems C Min ULR$^{\mathcal{R};\neq}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$ are solvable in polynomial time.

Proposition 9 C Min ULR$^{\geq;=}$ *is equally hard to approximate as* Min ULR$^{\geq}$.

Proof According to Fact 4, Min ULR$^{\geq}$ can be reduced to Min ULR$^{=}$ with nonnegative variables. The latter problem is a particular case of C Min ULR$^{\geq;=}$ where the mandatory inequalities are just nonnegativity constraints. Conversely, C Min ULR$^{\geq;=}$ can be reduced to C Min ULR$^{\geq;\geq}$ by substituting each optional equation by two complementary inequalities. We can then use the standard reduction from C MIN ULR$^{\geq;\geq}$ to MIN ULR$^{\geq}$. $\square$

Similarly we can show that C MIN ULR$^{>;=}$ is at least as hard to approximate as MIN ULR$^{>}$.

Proposition 10 *C MIN ULR$^{\neq;=}$ is MIN DOMINATING SET-hard even when restricted to homogeneous systems with binary coefficients. Therefore it cannot be approximated within any constant, unless P=NP, and within $(1-\varepsilon)\ln n$, for any $\varepsilon > 0$, unless NP $\subseteq$ DTIME($n^{\log\log n}$).*

PROOF The proof is by reduction from MIN DOMINATING SET. Let $G = (V, E)$ be an arbitrary instance of MIN DOMINATING SET. For each node $v_i \in V$, $1 \leq i \leq n$, we consider the optional equation $x_i = 0$ and the mandatory relation

$$x_i + \sum_{j \in N(v_i)} x_j \neq 0. \tag{4}$$

Clearly, 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 relations of the corresponding linear system. Given a dominating set $V' \subseteq V$ of size s, the solution $\mathbf{x}$ defined by

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

satisfies all mandatory relations and all but s of the optional ones. Conversely, given a solution vector $\mathbf{x}$ that satisfies all mandatory relations and violates s of the optional relations, we consider the set of nodes $V' \subseteq V$ containing all nodes v_i such that $x_i \neq 0$. V' is a dominating set of size s because $x_i + \sum_{j \in N(v_i)} x_j \neq 0$ only when at least one of the variables is nonzero, which corresponds to the case where at least one of the nodes is in the dominating set. $\square$

5 Hardness of variants with bounded discrete variables

In this section we determine the approximability of MIN RVLS$^{\mathcal{R}}$ and MIN ULR$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$ when the variables are restricted to take a finite number of discrete values. In particular, we consider the cases with binary variables in $\{0, 1\}$ and bipolar ones in $\{-1, 1\}$. The corresponding variants are referred to as BIN MIN RVLS, BIP MIN RVLS, BIN MIN ULR and BIP MIN ULR, respectively.

Theorem 11 *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 $\mathcal{R}_1, \mathcal{R}_2 \in \{=, \geq, >, \neq\}$. 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$^{\geq;\geq}$ 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]. Let $G = (V, E)$ be an arbitrary instance of MIN IND DOM SET. For each node $v_i \in V$, $1 \leq i \leq n$, we consider the optional inequality

$$x_i \leq 0 \tag{5}$$

and the mandatory one

$$x_i + \sum_{j \in N(v_i)} x_j \geq 1 \tag{6}$$

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

$$x_i + x_j \leq 1. \tag{7}$$

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

It is readily verified that there exists an independent dominating set in G of size at most s if and only if there exists a solution $\mathbf{x} \in \{0,1\}^n$ that violates s relations of the corresponding linear system. Given an independent dominating set $V' \subseteq V$ of size s, the solution $\mathbf{x}$ defined by

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

satisfies all mandatory inequalities and $n - s$ of the optional ones. Conversely, given a solution vector $\mathbf{x}$ that violates s of the relations, we consider the set of nodes $V' \subseteq V$ containing all nodes v_i such that $x_i = 1$. V' is clearly of size s, independent (because of the mandatory relations (7)) and dominating (because of the mandatory relations (6)).

The theorem 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$^{\mathcal{R}_1;\mathcal{R}_2}$, we use the same reduction as above but the right hand side of the three types of relations must be substituted according to the following table.

	operator $\geq$	operator $>$	operator $\neq$	operator $=$
type (5)	≤ 0	< 1	$\neq 1$	$= 0$
type (6)	≥ 1	> 0	$\neq 0$	$= \lvert N(v_i)\rvert - \sum_{j=1}^{\lvert N(v_i)\rvert-1} y_{ij}$
type (7)	≤ 1	< 2	$\neq 2$	$= 1 - z_{ij}$

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 $|E|$.

To get the hardness results 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 non-approximability bound than for Min Ind Dom Set. □

It is worth noting that Bin Min ULR$^{\mathcal{R}_1}$ and C Bin Min ULR$^{\mathcal{R}_1;\mathcal{R}_2}$ with $\mathcal{R}_1, \mathcal{R}_2 \in \{=, \geq, >, \neq\}$ 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 introduce new mandatory relations $x_0 \geq 0$, $x_0 > 0$, $x_0 \neq 0$, or $x_0 = y_0$ with a new variable y_0, depending on the type of relations. In the case of nonstrict inequalities and equalities, we add (as in the proof of Theorem 5) a large enough number of copies of the relations to make x_0 the variable occurring the most frequently in the satisfied relations.

Also BIP MIN ULR$^{\mathcal{R}_1}$ and C BIP MIN ULR$^{\mathcal{R}_1;\mathcal{R}_2}$ are NPO PB-complete for every combination of $\mathcal{R}_1, \mathcal{R}_2 \in \{=, \geq, >, \neq\}$. Since for any relation with binary variables $x_i \in \{0,1\}$ one can construct an equivalent relation with bipolar variables $y_i \in \{-1,1\}$ using the substitution $y_i = 2x_i - 1$, we have exactly the same non-approximability bounds as in Theorem 11.

Similar bounds also hold for MIN RVLS with binary or bipolar variables.

Proposition 12 BIN MIN RVLS$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{=, \geq, >, \neq\}$ *is* NPO PB-*complete. Assuming* P$\neq$NP, BIN MIN RVLS$^{=}$ *and* BIN MIN RVLS$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{\geq, >, \neq\}$ *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 11 for proving the NPO PB-hardness of C BIN MIN ULR$^{\mathcal{R};\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$. The BIN MIN RVLS$^{\mathcal{R}}$ instance is simply composed of the mandatory relations constructed in the C BIN MIN ULR$^{\mathcal{R};\mathcal{R}}$ instance. The number of violated optional relations in the proof exactly corresponds to the number of nonzero variables. Therefore, BIN MIN RVLS$^{\geq}$, BIN MIN RVLS$^{>}$, and BIN MIN RVLS$^{\neq}$ are NPO PB-hard and not approximabile within $n^{1-\varepsilon}$.

For BIN MIN RVLS$^{=}$, we still 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 variables, 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 $\mathbf{x}$ of the resulting instance we will have, for each variable x_i, that $x_i = x_{i1} = \ldots = x_{iN}$. 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 $\mathbf{x}$ are equal to 1, the size of the independent 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 non-approximability bound $n^{0.5-\varepsilon}$, where n is the number of variables. $\square$

The same lower bounds are valid for the BIP MIN RVLS variants.

Note that BIN MIN RVLS$^{\geq}$ is equivalent to MIN POLYNOMIALLY BOUNDED 0-1 PROGRAMMING, which was shown to be NPO PB-complete in [37]. Moreover, the corresponding maximization problem BIN MAX IVLS$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$ 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 [38].

6 Two special cases from discriminant analysis and machine learning

In this section we discuss two interesting special cases of MIN RVLS and MIN ULR with inequalities which arise in discriminant analysis and machine learning, more precisely, when designing two-class linear classifiers [21] and when training perceptrons [48].

Given a set of vectors $T = \{\mathbf{a}^k\}_{1 \leq k \leq p} \subset \mathbb{R}^n$ labeled as positive or negative examples, we look for a hyperplane H, specified by a normal vector $\mathbf{w} \in \mathbb{R}^n$ 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 $\mathbf{a}^k$ if $\mathbf{a}^k\mathbf{w} > w_0$ or $\mathbf{a}^k\mathbf{w} \leq w_0$ depending on whether $\mathbf{a}^k$ 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 *perceptron* and its parameters w_j, $1 \leq j \leq n$ as its *weights* [33].

In the general situation where T is nonlinearly separable, a natural objective is to minimize the number of vectors $\mathbf{a}^k$ that are misclassified (see [43, 25] and the included references). This problem is referred to as Min Misclassifications. Note that we have studied in [6] the approximability of the complementary problem where one looks for a hyperplane which is consistent with as many $\mathbf{a}^k \in T$ as possible.

In [8] a way of extending the non-approximability bounds for Min ULR$^=$ to the symmetric version of Min Misclassifications where we ask $\mathbf{a}^k\mathbf{w} < w_0$ 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 [8], 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 $\mathbf{aw} \geq 0$ is replaced by $\mathbf{aw} + \delta > 0$. 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 as in Proposition 1. 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 13 Suppose we have a system $\mathbf{a}^k\mathbf{w} > b^k$ with $1 \leq k \leq p$ 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 $\mathbf{a}^k\mathbf{w} > w_0$ type or $\mathbf{a}^k\mathbf{w} < w_0$ 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 [8], we just apply this technique to the system consisting of $\mathbf{aw} + \delta > 1/(2L)$ 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 NP $\subseteq$ DTIME($n^{\text{polylog}\, n}$).

A special case of Min RVLS$^>$ and Min RVLS$^\geq$ 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 to minimize the number of parameters w_j, $1 \leq j \leq n$, that are required to correctly classify all examples in T [42, 44, 58]. 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 [7, 41]). The problem of identifying a subset of most relevant features is well known in the discriminant analysis literature under the name of variable selection [47, 42].

In practice, when the classifier (perceptron) at hand cannot satisfactorily classify the training set based on the original n_o features, new features derived from the original ones are added. For instance, the $O(n_o^d)$ higher-order products of the original features are frequently included for several values of $d \geq 2$ [33]. Although such a procedure may lead to a total number of features $n \gg n_o$, often only a small fraction of them is needed. From the generalization point of view, it is thus important to try to remove the superfluous ones.

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

While Lin and Vitter showed that Min Relevant Features with binary inputs is NP-hard [44], van Horn and Martinez established that the symmetric variant with strict inequalities is at least as hard to approximate as Min Dominating Set [57, 58]. 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 sense [56] 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.

By extending the non-approximability bound for Min RVLS$^=$ derived in Theorem 2, we provide strong evidence that no such approximation algorithm exists.

Theorem 14 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 We adapt the reduction from Min Set Covering used in the proof of Theorem 2 to the case of linear classifiers (perceptrons) with real parameters (weights).

To show that Min Relevant Features cannot be approximated within any constant factor unless $\mathrm{P} = \mathrm{NP}$, we proceed like in the proof of Proposition 1.

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

$$A\mathbf{w} > w_0\, \mathbf{1}, \tag{8}$$

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 $w_0 \geq 0$. Hence we have a training set with $|S|$ positive and one negative 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 $\mathcal{C}'$ of cardinality K, the parameter vector $\mathbf{w}$ given by

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

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 recursively by self-improvement like in the proof of Theorem 2. Since in the reduction the size of the examples n is polynomially related to the number of examples p, the same non-approximability bound is also valid with respect to n. $\square$

The consequences of this result on the hardness of designing compact feedforward networks are discussed in detail in [3]. From an artificial neural network perspective, Theorem 14 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 [40] a stronger non-approximability bound but under a stronger cryptographic assumption. In particular, they showed that if trapdoor functions[2] 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 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 non-approximability 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 is to devise efficient heuristics with good average behavior.

7 Conclusions

The various versions of Min ULR$^{\mathcal{R}}$ and Min RVLS$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$ that we have considered are obtained by placing constraints on the coefficients, on the variables or by assigning a possibly different importance to each relation.

Table 1 summarizes the non-approximability results that hold for Min ULR variants unless P=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 required the variables occurring most frequently in the satisfied relations to be 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$(n^{\text{polylog}\, n})$ [8, 9]. 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. Moreover, we got for free the constant factor result.

[2] A trapdoor function $\mathcal{T}$ is a one-to-one function such that $\mathcal{T}$ and its inverse are easy to evaluate but, given $\mathcal{T}$, the inverse $\mathcal{T}^{-1}$ cannot be constructed in polynomial time [52].

	Real variables	Binary variables
MIN ULR$^{=}$	not within any constant [8]	
MIN ULR$^{\geq}$		
MIN ULR$^{\neq}$	trivial	
C MIN ULR$^{\geq;\geq}$	as hard as MIN ULR$^{\geq}$	
C MIN ULR$^{\geq;>}$	at least as hard as MIN ULR$^{>}$	NPO PB-complete
C MIN ULR$^{\geq;=}$	at least as hard as MIN ULR$^{\geq}$	
C MIN ULR$^{\neq;\geq}$		
C MIN ULR$^{\neq;=}$	MIN DOMINATING SET-hard	
C MIN ULR$^{\mathcal{R};\neq}$	polynomial time	

Table 1: Main approximability results for MIN ULR variants with $\mathcal{R} \in \{=, \geq, >, \neq\}$ that hold assuming $\mathrm{P} \neq \mathrm{NP}$. The constrained versions C MIN ULR$^{=;\mathcal{R}}$ with real variables and mandatory equations are equivalent to the corresponding MIN ULR$^{\mathcal{R}}$. The table is still valid when nonstrict inequalities are substituted by strict inequalities and vice versa. The only difference is that C MIN ULR$^{\geq;=}$ is as hard as MIN ULR$^{\geq}$ while C MIN ULR$^{>;=}$ is only at least as hard as MIN ULR$^{>}$.

The weighted and constrained variants of MIN ULR turn out to be equally hard to approximate as the unweighted ones. Restricting the variables to binary (or bipolar) 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 non-approximability 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, 6]). 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$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >, \neq\}$, we have shown that they cannot be approximated within a constant factor and within $2^{\log^{1-\varepsilon} n}$ under the usual assumptions. These results, which are of interest in their own right, directly imply the bounds established in [9] for the basic versions of MIN ULR while the converse is not true. When the variables are restricted to take binary or bipolar 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 compact perceptrons or linear classifiers, is not approximable within a logarithmic factor as conjectured in [57], unless all problems in NP are solvable in quasi-polynomial time.

Acknowledgments

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 [55].

Appendix

There exists an interesting relationship between MIN ULR with real and binary variables. If the number of violated relations is measured with respect to a given threshold, the basic variants with real variables are at least as hard as those with binary variables.

The threshold variant of MIN ULR, named THRESHOLD MIN ULR, with any type of relations is defined as follows. The input consists of a linear system of relations together with a *threshold* T, which is a lower bound on the minimum number of unsatisfiable relations of the system. Given a solution vector $\mathbf{x}$ that violates s relations, the objective function is defined as $s - T$. Since at least T relations are violated, this value is always nonnegative.

Proposition 15 THRESHOLD MIN ULR$^{\mathcal{R}}$ *with* $\mathcal{R} \in \{=, \geq, >\}$ *is* NPO PB*-complete. Assuming* P$\neq$NP*, it cannot be approximated within* $n^{1-\varepsilon}$ *for any* $\varepsilon > 0$*, where* n *is the number of variables, and within* $p^{0.5-\varepsilon}$ *for any* $\varepsilon > 0$*, where* p *is the number of relations.*

PROOF By reduction from BIN MIN ULR$^{\mathcal{R}}$ with $\mathcal{R} \in \{=, \geq, >\}$. Given an input instance of BIN MIN ULR$^{\mathcal{R}}$ with n binary variables and p relations, we construct an instance of THRESHOLD MIN ULR$^{\mathcal{R}}$ by extending the system with relations which ensure that the variables just take zero or one values.

In the equality case we add, for every variable x_i, $p + 1$ copies of the equations $x_i = 0$ and $x_i = 1$. If all variables in a solution vector take values in $\{0, 1\}$, exactly $n(p + 1)$ of the added equations will be satisfied and $n(p + 1)$ violated. For each variable that is neither zero nor one, $p + 1$ additional equations will be violated, which are more than the number of equations in the original system. If we choose $n(p + 1)$ as the threshold we will get an S-reduction from BIN MIN ULR$^{=}$ to THRESHOLD MIN ULR$^{=}$ without variable amplification and with a relation amplification of $O(np)$. Since in the proof of Theorem 11 the number of relations is about the same as the number of variables, THRESHOLD MIN ULR$^{=}$ cannot be approximated within $p^{0.5-\varepsilon}$ for any $\varepsilon > 0$, where p is the number of relations.

For THRESHOLD MIN ULR$^{\geq}$ we use the same construction, but instead of the two equations $x_i = 0$ and $x_i = 1$ we include the four inequalities $x_i \leq 0$, $x_i \geq 0$, $x_i \leq 1$, and $x_i \geq 1$. If all components of a solution vector are either zero or one, just one of these four inequalities will be violated. Otherwise two of the inequalities will be violated. The threshold and the rest of the proof are the same as in the equality case.

For THRESHOLD MIN ULR$^{>}$ we just use the above-mentioned equivalence between MIN ULR$^{\geq}$ and MIN ULR$^{>}$. $\square$

References

[1] E. Amaldi. On the complexity of training perceptrons. In T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, *Artificial Neural Networks*, pages 55–60, Amsterdam, 1991. Elsevier science publishers B.V.

[2] E. Amaldi. Complexity of problems related to training perceptrons. In M. Chaleyat-Maurel, M. Cottrell, and J.-C. Fort, editors, *Actes du congrès "Aspects théoriques des réseaux de neurones", Congrès Européen de Mathématiques.* Paris, July 1992.

[3] E. Amaldi. *From finding maximum feasible subsystems of linear systems to feedforward neural network design.* PhD thesis, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1994.

[4] E. Amaldi and C. G. Diderich. On the probabilistic and thermal perceptron training algorithms. Technical Report ORWP-8-94, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1994.

[5] E. Amaldi and 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.

[6] E. Amaldi and V. Kann. The complexity and approximability of finding maximum feasible subsystems of linear relations. *Theoretical Computer Science*, 147:181–210, 1995.

[7] D. Angluin and C. H. Smith. Inductive inference: theory and methods. *ACM Computing Surveys*, 15:237–269, 1983.

[8] S. Arora, L. Babai, J. Stern, and Z. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. In *Proc. of 34th Ann. IEEE Symp. on Foundations of Comput. Sci.*, pages 724–733. IEEE Computer Society, 1993.

[9] S. Arora, L. Babai, J. Stern, and Z. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. Manuscript, 1994.

[10] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. In *Proc. of 33rd Ann. IEEE Symp. on Foundations of Comput. Sci.*, pages 14–23. IEEE Computer Society, 1992.

[11] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs and non-approximability – towards tight results. In *Proc. of 36th Ann. IEEE Symp. on Foundations of Comput. Sci.*, pages 422–431. IEEE Computer Society, 1995.

[12] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient probabilistically checkable proofs and applications to approximation. In *Proc. Twenty fifth Ann. ACM Symp. on Theory of Comp.*, pages 294–304. ACM, 1993.

[13] K. P. Bennett and O. L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1:23–34, 1992.

[14] J. Chinneck. An effective polynomial-time heuristic for the minimum-cardinality IIS set-covering problem. Technical Report SCE-95-02, Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada, 1995. To appear in Annals of Mathematics and Artificial Intelligence.

[15] J. Chinneck and E. Dravnieks. Locating minimal infeasible constraint sets in linear programs. *ORSA Journal on Computing*, 3:157–168, 1991.

[16] V. Chvatal. *Linear programming.* Freeman and Co., New York, 1983.

[17] P. Crescenzi and 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, and L. Trevisan. Structure in approximation classes. In *Proc. of 1st International Conf. on Computing and Combinatorics*, pages 539–548. Springer-Verlag, 1995. Lecture Notes in Comput. Sci. 959.

[19] P. Crescenzi and A. Panconesi. Completeness in approximation classes. *Inform. and Comput.*, 93(2):241–262, 1991.

[20] P. Crescenzi, R. Silvestri, and L. Trevisan. To weight or not to weight: Where is the question? Manuscript, 1995.

[21] R. O. Duda and P. E. Hart. *Pattern classification and scene analysis.* Wiley, New York, 1973.

[22] J. Edmonds. Redundency and Helly for linear programming. Lecture notes, April 1994.

[23] G. Even, J. Naor, B. Schieber, and M. Sudan. Approximating minimum feedback sets and multi-cuts in directed graphs. In *Proc. of 4th Int. Conf. on Integer Prog. and Combinatorial Optimization*, pages 14–28. Springer-Verlag, 1995. Lecture Notes in Comput. Sci. 920.

[24] U. Feige. A threshold of ln(n) for approximating set cover. Manuscript, 1995.

[25] M. R. Frean. A thermal perceptron learning rule. *Neural Computation*, 4:946–957, 1992.

[26] S. I. Gallant. Perceptron-based learning algorithms. *IEEE Trans. on Neural Networks*, 1:179–191, 1990.

[27] M. R. Garey and 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 and J. Ryan. Identifying minimally infeasible subsystems of inequalities. *ORSA Journal on Computing*, 3:61–63, 1990.

[29] H. J. Greenberg. Consistency, redundancy, and implied equalities in linear systems. Manuscript, November 1993.

[30] H. J. Greenberg and F. H. Murphy. Approaches to diagnosing infeasible linear programs. *ORSA Journal on Computing*, 3:253–261, 1991.

[31] R. Greer. *Trees and Hills: Methodology for Maximizing Functions of Systems of Linear Relations*, volume 22 of *Annals of Discrete Mathematics.* Elsevier science publishing company, Amsterdam, 1984.

[32] M. M. Halldórsson. Approximating the minimum maximal independence number. *Inform. Process. Lett.*, 46:169–172, 1993.

[33] M. Hassoun. *Fundamentals of artificial neural networks.* MIT Press, Cambridge, MA, 1995.

[34] K-U. Höffgen, H-U. Simon, and K. van Horn. Robust trainability of single neurons. *J. Comput. System Sci.*, 50:114–125, 1995.

[35] D. S. Johnson and F. P. Preparata. The densest hemisphere problem. *Theoretical Computer Science*, 6:93–107, 1978.

[36] V. Kann. *On the Approximability of NP-complete Optimization Problems.* PhD thesis, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, 1992.

[37] V. Kann. Polynomially bounded minimization problems that are hard to approximate. *Nordic J. Computing*, 1:317–331, 1994.

[38] V. Kann. Strong lower bounds on the approximability of some NPO PB-complete maximization problems. In *Proc. of 20th Ann. Mathematical Foundations of Comput. Sci.*, page to appear. Springer-Verlag, 1995. Lecture Notes in Comput. Sci. 969.

[39] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[40] M. Kearns and L. Valiant. Cryptographic limitations on learning boolean formulae and finite automata. *Journal of the ACM*, 41:67–95, 1994.

[41] M. Kearns and U. Vazirani. *An introduction to Computational Learning Theory.* MIT Press, Cambridge, MA, 1994.

[42] G. J. Koehler. Linear discriminant functions determined by genetic search. *ORSA Journal on Computing*, 3:345–357, 1991.

[43] G. J. Koehler and S. S. Erenguc. Minimizing misclassifications in linear discriminant analysis. *Decision Sciences*, 21:63–85, 1990.

[44] J.-H. Lin and J. S. Vitter. Complexity results on learning by neural nets. *Machine Learning*, 6:211–230, 1991.

[45] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41:960–981, 1994.

[46] M. Marchand and M. Golea. On learning simple neural concepts: from halfspace intersections to neural decision lists. *Network: Computation in Neural Systems*, 4:67–85, 1993.

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

[48] M. L. Minsky and S. Papert. *Perceptrons: An introduction to computational Geometry.* MIT Press, Cambridge, MA, 1988. Expanded edition.

[49] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization, Algorithms and Complexity.* Prentice Hall, 1982.

[50] M. Parker. *A set covering approach to infeasibility analysis of linear programming problems and related issues.* PhD thesis, Department of Mathematics, University of Colorado at Denver, 1995.

[51] M. Parker and J. Ryan. Finding the minimum weight IIS cover of an infeasible system of linear inequalities. Manuscript, 1995.

[52] R. Rivest. Cryptography. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume A: Algorithms and complexity, pages 717–755. Elsevier science publishing company, 1990.

[53] J. Sankaran. A note on resolving infeasibility in linear programs by constraint relaxation. *Oper. Res. Letters*, 13:19–20, 1993.

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

[55] P. Seymour. Packing directed circuits fractionally. *Combinatorica*, 15:281–288, 1995.

[56] L. G. Valiant. A theory of the learnable. *Communications of ACM*, 27:1134–1142, 1984.

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

[58] K. van Horn and T. Martinez. The minimum feature set problem. *IEEE Trans. on Neural Networks*, 7:491–494, 1994.

[59] R. E. Warmack and R. C. Gonzalez. An algorithm for optimal solution of linear inequalities and its application to pattern recognition. *IEEE Trans. on Computers*, 22:1065–1075, 1973.