Superlinear integrality gaps for the minimum majority problem

Philip M. Long
Google
plong@google.com

Abstract

The minimum majority problem is as follows: given a matrix $A \in \{-1, 1\}^{m \times n}$, minimize $\sum_{i=1}^{n} x_i$ subject to $A\mathbf{x} \geq \mathbf{1}$ and $\mathbf{x} \in (\mathbb{Z}^+)^n$. An approximation algorithm that finds a solution with value $O(\text{opt}^2 \log m)$ in $\text{poly}(m, n, \text{opt})$ time is known, which can be obtained by rounding a linear programming relaxation.

We establish integrality gaps that limit the prospects for improving on this guarantee through improved rounding and/or the application of Lovász-Schrijver (LS) or Sherali-Adams (SA) tightening of the relaxation. These gaps show that applying LS and SA relaxations cannot improve on the $O(\text{opt}^2 \log m)$ guarantee by more than a constant factor in polynomial time.

1 Introduction

This paper is about the minimum majority problem: given $A \in \{-1, 1\}^{m \times n}$, minimize $\sum_{i=1}^{n} x_i$ subject to $A\mathbf{x} \geq \mathbf{1}$ and $\mathbf{x} \in (\mathbb{Z}^+)^n$. This problem is motivated by the margin analysis of boosting [44] (see [33, 34]). It also formalizes the problem of compressing a learned ensemble [42, 8]; starting from a classifier that takes a vote over a large number of classifiers, find a small multiset of the voters that produce the same results on training data.

It is known [25] that, in $\text{poly}(m, n, \text{opt})$ time, an algorithm can find a solution with value $O(\text{opt}^2 \log m)$. This has been proved using boosting [25] and randomized rounding [33]. A simple and direct proof which borrows ideas from the analysis of boosting [19, 45, 16, 37] is provided in Section 7. Our work is motivated by the following question: is this $O(\text{opt}^2 \log m)$ algorithm the best possible?

For many discrete optimization problems, the best polynomial-time approximation algorithm known may be constructed by rounding the solution of an appropriate relaxation of the problem [39, 2]. While the "natural" LP relaxation suffices for many problems, adding carefully chosen constraints can sometimes help get better relaxations. Effective systematic methods for generating such stronger relaxations have been developed. A popular example of such a "lift-and-project" method, which recovers the best approximation algorithms in many cases, is the Lovász-Schrijver (LS) technique [36]. This technique generates a sequence of semi-definite programs with the same objective function as the original problem, but with feasible regions that approximate the convex hull of the feasible region of an integer programming problem progressively closely. If N is the size of the original problem, the rth LS relaxation has size $\text{poly}(N^r)$, which is polynomial in N if r is a constant. Arora et al. [4, 2] argued persuasively that the strength of this approach and the breadth of its applicability motivates study of its limitations. The main such mode of analysis is an *integrality gap* [35], demonstrating that a solution of the relaxed problem obtained after r rounds of the Lovász-Shrijver method has a solution that is significantly better than the best solution to the original integer programming problem.

Strong integrality gaps, sometimes for small values of r, have been obtained for several problems of central interest [35, 20, 52, 47, 11, 21, 2, 29, 31, 5, 17, 27]. Typically, these limit the prospects for improvements on the approximation ratio that can be achieved in polynomial time (and even in subexponential time) through LS relaxations: i.e., if opt is the value of the optimal solution, they limit the prospects for achieving $\text{opt} f(n)$ for some $f(n)$ (which may be a constant).

In this work, we establish integrality gaps for LS and related lift-and-project relaxations for the minimum majority problem. We first show that a constant number of levels of LS cannot yield an improvement on the $O(\text{opt}^2 \log m)$ guarantee. This is a special case of a more general result: for large enough k, $r < \frac{k}{5} - 1$, and $m \geq k^3/(r+1)^2$, there is an instance A of the minimum majority problem such that (a) the rth level Lovász-Schrijver relaxation has value k, and (b) any integer solution has value $\Omega((k^2/(r+1)^2) \log m)$. In additional to providing evidence that LS relaxations cannot improve on the $O(\text{opt}^2 \log m)$ bound in polynomial time, it also demonstrates dim prospects for substantially improving on this guarantee while using significantly less than the $O(n^{\text{opt}})$ time used by a brute-force algorithm.

Another popular lift-and-project method uses Sherali-Adams (SA) [49] relaxations which are incomparable in strength with LS relaxations[1] [32]. We next establish a bound for Sherali-Adams (SA) relaxations analogous to the one proved for LS relaxations.

As mentioned above, the minimum majority problem is motivated by applications to machine learning, but the possibility that $O(\text{opt}^2 \log m)$ might be the best possible approximation guarantee achievable in polynomial time may be of more fundamental and broader interest, since this form is qualitatively unlike other known bounds.

Other Related Work. For a comparison of various lift-and-project methods, we refer the reader to the survey by Laurent [32], and to the survey by Chlamtac and Tulsiani [13]. In recent years, several integrality gaps for SA relaxations have been shown (see e.g. [15, 46, 10, 28, 38, 26, 22, 6, 43, 7, 12]). Other kinds of hardness results for some other sparse learning problems can be found in [3, 51, 14, 53, 18].

Techniques. Our analysis uses the probabilistic method, choosing the entries of $A \in \{-1, 1\}^{m \times n}$ independently at random from a distribution that assigns slightly more probability to 1 than -1. To prove the lower bound for the integer programming solution, we need to show that any small ensemble is likely to violate one of the constraints – for the candidate solution $\mathbf{x}$, we need a lower bound on the probability that one of the rows $\mathbf{a}$ of A will satisfy $\mathbf{a} \cdot \mathbf{x} \leq 0$. For this, we need a lower bound on the tail of a sum of independent random variables. The Berry–Esseen inequality does not appear to provide enough leverage, and it was also not clear how to apply anti-concentration techniques such as [50, 40, 48, 30] to get the bounds needed here. (The challenges include the possibility of relatively large summands and their assymmetric distribution.) Instead, we argue roughly as follows. First, only the components of $\mathbf{a}$ corresponding to nonzero components of $\mathbf{x}$ matter. However, if $\mathbf{x}$ is a good solution, then there are few of those components, and, thus, typically, the probability of the projection of $-\mathbf{a}$ onto those components is not too much more than the probability of the projection of $\mathbf{a}$. It follows that the probability that $\mathbf{a} \cdot \mathbf{x} \leq 0$ cannot be too

[1] The LS relaxations analyzed here are usually denoted by LS^+, and are sometimes tighter than the "plain" LS relaxations.

much smaller than the probability that $\mathbf{a} \cdot \mathbf{x} \geq 0$, which in turn means that it cannot be very small, period.

In contrast, the fractional solution benefits from the stability conferred by averaging over *all* of the variables. This can be established using standard techniques such as Hoeffding bounds and the union bound for the linear programming relaxation. Applying a "protection lemma" of [23] demonstrates that it survives multiple rounds of LS tightening; we prove a similar protection lemma for SA relaxations. The fractional solution qualifies for these protection lemmas because it is far from violating any of the constraints.

As mentioned above, this work suggests that the unusual $O(\mathrm{opt}^2 \log m)$ guarantee may be the best possible for this problem. Based on our proof, this is in part because a fractional solution is especially good when the constraints depend on a lot of variables. In this respect, the minimum majority problem differs fundamentally from many of the problems whose approximation properties are frequently studied – many of these concern graphs and/or have sparse constraints. The minimum majority problem may be a representative of a class of problems with fundamentally different approximation properties, and as such may be an interesting subject for further study.

2 Preliminaries

Let $\mathrm{opt}(A)$ be the minimum of $\sum_{i=1}^n x_i$ subject to $A\mathbf{x} \geq \mathbf{1}$ and $\mathbf{x} \in (\mathbb{Z}^+)^n$. (If $A\mathbf{x} \geq \mathbf{1}$ is unsatisfiable, define $\mathrm{opt}(A) = \infty$.) Let $\mathrm{opt}_L(A)$ be the minimum of $\sum_{i=1}^n x_i$ subject to $A\mathbf{x} \geq \mathbf{1}$ and $\mathbf{x} \geq \mathbf{0}$.

Lemma 1 ([41]) *Let $U_1, ..., U_\ell$ be independent random variables with each U_i taking values in $[a_i, b_i]$ and let $S = \sum_{i=1}^\ell U_i$. Then $\mathbf{Pr}(S \geq \mathbf{E}(S) + \eta) \leq \exp\left(\frac{-2\eta^2}{\sum_{i=1}^\ell (b_i - a_i)^2}\right)$.*

3 An integrality gap for linear programming

As a warmup, we prove the following integrality gap theorem for the linear programming relaxation.

Theorem 2 *There are constants $k_0, c > 0$ and a polynomial p, such that, for all $k \geq k_0$ and $m \geq k^3$, there is an $n \in \mathbb{N}$ such that $n \leq p(k, \ln m)$, and $A \in \{-1, 1\}^{m \times n}$, such that $\mathrm{opt}_L(A) \leq k$ and $\mathrm{opt}(A) \geq ck^2 \ln m$.*

3.1 Setup

The proof uses the probabilistic method, analyzing matrices A chosen at random. Setting, with foresight, $\gamma = \frac{1}{k}$, $d = \left\lfloor \frac{k^2 \ln m}{43} \right\rfloor$, $n = \lceil 4k^2 \ln m \rceil$, consider $A \in \{-1, 1\}^{m \times n}$ whose entries are drawn i.i.d., with $\mathbf{Pr}(A_{ij} = 1) = 1/2 + \gamma$, for $\gamma \leq 1/5$. (We can ensure that $\gamma \leq 1/5$ by setting $k \geq 5$.)

We set up some notation next. Let $S(n, d)$ consist of all $\mathbf{x} \in (\mathbb{Z}^+)^n$ such that $\sum_{i=1}^n x_i \leq d$ (or, equivalently, all multisets of at most d elements of $[n]$).

Say that $\mathbf{x} \in S(n, d)$ is *hit by row* i, if row i witnesses the infeasibility of $\mathbf{x}$, i.e., if $\sum_{j=1}^n A_{ij} x_j \leq 0$. Say $\mathbf{x}$ is *hit by* A if it is hit by some row of A.

3.2 The hit probability

This section analyzes the probability that a prospective solution is hit by a single row. Let Q be the distribution governing the choice of a single row of A, n i.i.d. draws from a distribution over $\{-1,1\}$ that assigns probability $1/2+\gamma$ to 1. For some realization $\mathbf{a}$ of a row of A, we will use the shorthand $Q(\mathbf{a})$ for $Q(\{\mathbf{a}\})$, the probability of generating $\mathbf{a}$.

The following will be proved in this subsection.

Lemma 3 *For all large enough k_0, for any $\mathbf{x} \in S(n,d)$, we have* $\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x} \leq 0) \;\geq\; \exp(-14\gamma^2 d)$.

Proof (of Lemma 3): First, for any realization $\mathbf{a}$, we have

$$\begin{aligned}\frac{Q(-\mathbf{a})}{Q(\mathbf{a})} &= \frac{(1/2+\gamma)^{|\{j:a_j=-1\}|}(1/2-\gamma)^{|\{j:a_j=1\}|}}{(1/2+\gamma)^{|\{j:a_j=1\}|}(1/2-\gamma)^{|\{j:a_j=-1\}|}} \\ &= \left(\frac{1-2\gamma}{1+2\gamma}\right)^{\sum_{j=1}^n a_j}. \qquad (1)\end{aligned}$$

Let us say that $\mathbf{a} \in \{-1,1\}^n$ is *balanced* if $|\sum_{i=1}^n a_i| \leq 3\gamma n$ and let B be the event that a random row is balanced. Since $\mathbf{E}_{\mathbf{a}\sim Q}\left(\sum_{i=1}^n a_i\right) = 2\gamma n$, Lemma 1 implies

$$Q(B) = \mathbf{Pr}_{\mathbf{a}\sim Q}\left(\left|\sum_{i=1}^n a_i\right| \leq 3\gamma n\right) \geq 1 - 2\exp\left(-\frac{\gamma^2 n}{2}\right). \qquad (2)$$

Furthermore, for $\mathbf{a} \in B$,

$$\frac{Q(-\mathbf{a})}{Q(\mathbf{a})} = \left(\frac{1-2\gamma}{1+2\gamma}\right)^{\sum_i a_i} \leq \left(\frac{1+2\gamma}{1-2\gamma}\right)^{3\gamma n} \leq e^{13\gamma^2 n} \qquad (3)$$

since $\gamma \leq 1/5$.

We have

$$\begin{aligned}\frac{1}{\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x} \leq 0)} &\leq \frac{\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x} \leq 0) + \mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x} \geq 0)}{\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x} \leq 0)} \\ &= 1 + \frac{\mathbf{Pr}_{\mathbf{a}\sim Q}(-\mathbf{a}\cdot\mathbf{x} \leq 0)}{\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x} \leq 0)}.\end{aligned}$$

Furthermore

$$\begin{aligned}
\frac{\mathbf{Pr}_{\mathbf{a}\sim Q}(-\mathbf{a}\cdot\mathbf{x}\le 0)}{\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x}\le 0)} &= \frac{\sum_{\mathbf{a}:(-\mathbf{a})\cdot\mathbf{x}\le 0} Q(\mathbf{a})}{\sum_{\mathbf{a}:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a})} \\
&= \frac{\sum_{\mathbf{a}:\mathbf{a}\cdot\mathbf{x}\le 0} Q(-\mathbf{a})}{\sum_{\mathbf{a}:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a})} \\
&= \frac{\sum_{\mathbf{a}\in B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(-\mathbf{a}) + \sum_{\mathbf{a}\notin B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(-\mathbf{a})}{\sum_{\mathbf{a}\in B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a}) + \sum_{\mathbf{a}\notin B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a})} \\
&\le \frac{\sum_{\mathbf{a}\in B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a})e^{13\gamma^2 n} + \sum_{\mathbf{a}\notin B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(-\mathbf{a})}{\sum_{\mathbf{a}\in B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a}) + \sum_{\mathbf{a}\notin B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a})} \\
&\qquad \text{(by (3))} \\
&\le e^{13\gamma^2 n} + \frac{\sum_{\mathbf{a}\notin B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(-\mathbf{a})}{\sum_{\mathbf{a}\in B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a}) + \sum_{\mathbf{a}\notin B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(\mathbf{a})} \\
&= e^{13\gamma^2 n} + \frac{\sum_{\mathbf{a}\notin B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(-\mathbf{a})}{\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x}\le 0)}.
\end{aligned}$$

Finally,

$$\sum_{\mathbf{a}\notin B:\mathbf{a}\cdot\mathbf{x}\le 0} Q(-\mathbf{a}) \le \sum_{\mathbf{a}\notin B} Q(-\mathbf{a}) = \sum_{\mathbf{a}\notin B} Q(\mathbf{a}) \le 2e^{-\gamma^2 n/2}$$

by (2).

Tracing back, we get

$$\frac{1}{\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x}\le 0)} \le 1 + e^{13\gamma^2 n} + \frac{2e^{-\gamma^2 n/2}}{\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x}\le 0)}.$$

Solving for $\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x}\le 0)$ yields

$$\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x}\le 0) \ge \frac{1-2e^{-\gamma^2 n/2}}{1+e^{13\gamma^2 n}}.$$

Note that the truth or falsehood of $\mathbf{a}\cdot\mathbf{x}\le 0$ is determined by $\{a_i : i\in[n], x_i\ne 0\}$. Since $\mathbf{x}\in S(n,d)$, when proving this lemma, ignoring the irrelevant components of $\mathbf{a}$, we may assume without loss of generality that $n\le d$. Thus,

$$\mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x}\le 0) \ge \frac{1-2e^{-\gamma^2 d/2}}{1+e^{13\gamma^2 d}} \ge e^{-14\gamma^2 d}$$

for all large enough k_0, since $m\ge k^3$ and $d=\Omega(k^2\ln m)$, completing the proof. ∎

3.3 Analyzing opt_L and opt

Proof (of Theorem 2): Armed with Lemma 3, we are ready to analyze $\mathrm{opt}_L(A)$ and $\mathrm{opt}(A)$ for a random A. Let $\mathcal{Q}$ be the distribution governing the random choice of A, and, as before, let Q be the distribution of any row of A.

First, we claim that $\mathbf{Pr}_{A\sim\mathcal{Q}}(\mathrm{opt}_L(A) \le k) \ge 2/3$. To see this, consider the solution $x_1 = ... = x_n = k/n$. Then $\sum_{i=1}^n x_i = k$, and, for any row $\mathbf{a}$ of A, $\mathbf{E}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x}) = 2\gamma k = 2$. Applying Lemma 1 and a union bound,

$$\mathbf{Pr}_{A\sim\mathcal{Q}}(\neg A\mathbf{x} \ge \mathbf{1}) \le m \exp\left(\frac{-2}{n(2k/n)^2}\right) = m\exp\left(\frac{-n}{2k^2}\right) \le m\exp\left(-2\ln m\right) \le 1/3$$

for large enough k (since $m \ge k^3$), proving that $\mathbf{Pr}_{A\sim\mathcal{Q}}(\mathrm{opt}_L(A) \le k) \ge 2/3$.

Now we want a lower bound on the probability that $\mathrm{opt}(A) > d$. If, for all $\mathbf{x} \in S(n,d)$, $\mathbf{x}$ is hit by some row of A, then $\mathrm{opt}(A) > d$. Thus, if we denote the set of rows of A by $\mathcal{A}$, we may apply Lemma 3 to get

$$\begin{aligned}
\mathbf{Pr}_{A\sim\mathcal{Q}}(\mathrm{opt}(A) \le d) &= \mathbf{Pr}(\exists\mathbf{x} \in S(n,d),\ \forall\mathbf{a} \in \mathcal{A}, \mathbf{a}\cdot\mathbf{x} > 0) \\
&\le \sum_{\mathbf{x}\in S(n,d)} \prod_{\mathbf{a}\in\mathcal{A}} (1 - \mathbf{Pr}_{\mathbf{a}\sim Q}(\mathbf{a}\cdot\mathbf{x} \le 0)) \\
&\le \sum_{\mathbf{x}\in S(n,d)} \prod_{\mathbf{a}\in\mathcal{A}} \left(1 - \exp(-14\gamma^2 d)\right) \qquad \text{(by Lemma 3)} \\
&= \binom{d+n}{d} \left(1 - \exp(-14\gamma^2 d)\right)^m \\
&\le \exp\left(d\left(1 + \ln\left(\frac{n}{d}\right)\right) - m\exp(-14\gamma^2 d)\right).
\end{aligned}$$

Substituting the values of d, n, and γ, recalling $m \ge k^3$, for all large enough values of k_0, we have

$$\mathbf{Pr}_{A\sim\mathcal{Q}}(\mathrm{opt}(A) \le d) \le \exp\left(k^2 \ln m - m\exp\left(-\frac{14\ln m}{43}\right)\right),$$

and, once again since $m \ge k^3$, for a large enough value of k, this implies

$$\mathbf{Pr}_{A\sim\mathcal{Q}}(\mathrm{opt}(A) \le d) \le 1/3.$$

Combining this with $\mathbf{Pr}_{A\sim\mathcal{Q}}(\mathrm{opt}_L(A) \le k) \ge 2/3$, this completes the proof ∎

4 A 0/1 integer programming formulation

Systematic methods for constructing more refined relaxations of integer programming problems often are designed for the case in which the variables are constrained to 0/1 values. If algorithms are allowed time polynomial in opt, as well as n and m, then they can make use of integer programs that use $\mathrm{poly}(n, m, \mathrm{opt})$ variables. A natural formulation of the minimum majority problem as a 0/1 integer programming problem satisfies this requirement. Define the q-bounded minimum majority problem to be the minimum majority problem under the additional restriction that each variable is included the solution with multiplicity at most q. We may formulate the q-bounded minimum majority problem as follows, where U is an $n \times q$ matrix:

$$\min \sum_{j=1}^n \sum_{\ell=1}^q U_{j,\ell}, \text{ s.t. } \forall i,\ \ \sum_{j=1}^n A_{ij} \sum_{\ell=1}^q U_{j\ell} \ge 1,\ \ \forall j,\ell,\ U_{j\ell} \in \{0,1\}.$$

When $q \geq \text{opt}$, this is equivalent to the original minimum majority problem. An (approximation) algorithm may then solve the original minimum majority problem by "guessing" a suitable upper bound q. One may view $U_{j\ell}$ as an indicator function for $x_j \geq \ell$, so that $x_j = \sum_{\ell=1}^{q} U_{j\ell}$.

Another alternative uses a binary encoding:

$$\min \sum_{j=1}^{n} \sum_{\ell=1}^{q} 2^{\ell-1} U_{j,\ell}, \text{ s.t. } \forall i, \ \sum_{j=1}^{n} A_{ij} \sum_{\ell=1}^{q} 2^{\ell-1} U_{j\ell} \geq 1, \ \forall j, \ell, \ U_{j\ell} \in \{0,1\}.$$

This is equivalent to the original minimum majority problem when $q \geq \log_2 \text{opt}$, and, again, an approximation algorithm could "guess" a suitable value of q.

Both of these fall within the class of algorithms that perform

$$\min \sum_{j=1}^{n} \sum_{\ell=1}^{q} \beta_\ell U_{j,\ell}, \text{ s.t. } \forall i, \ \sum_{j=1}^{n} A_{ij} \sum_{\ell=1}^{q} \beta_\ell U_{j\ell} \geq 1, \ \forall j, \ell, \ U_{j\ell} \in \{0,1\}$$

for $\beta_1 = 1$ and $\beta_2, ..., \beta_q \in \mathbb{N}$. The natural linear programming relaxation of such an integer program replaces each constraint that $U_{j\ell} \in \{0,1\}$ with $U_{j\ell} \in [0,1]$. We will prove integrality gaps for all such programs.

5 Gaps for Lovász-Schrijver relaxations

In this section, we establish integrality gaps for Lovász-Schrijver (LS) relaxations of 0/1 integer programming formulations of the minimum majority problem.

5.1 Definition of LS relaxations and statement of gap theorem

Lovász-Schrijver relaxations are obtained through rounds, which progressively tighten the relaxation. Here is the definition of one round, following [23, 2]. (For intuition, please see [36, 13].) Before the round, the feasible region is $P = \{\mathbf{x} : A\mathbf{x} \geq \mathbf{b}, \mathbf{x} \in [0,1]^n\}$.

The first step is to embed P into $\mathbb{R}^{n+1}$, to produce a cone K that may be regarded as equivalent: $K = \{(x_0, \mathbf{x}) : A\mathbf{x} \geq x_0\mathbf{b}, \mathbf{x} \in [0, x_0]^n\}$. When analyzing K, we will number indices from 0 as usual; for each $i \in \{0, ..., n\}$, let $\mathbf{e}_i$ be the element of $\{0,1\}^{n+1}$ with a 1 only in position i.

Next is the "lifting" step, which defines a subset $M(K)$ of $\mathbb{R}^{(n+1)\times(n+1)}$, which might be thought of as constraints on products of pairs of variables. A symmetric Y is in $M(K)$ if

- $Y\mathbf{e}_0 = \text{diag}(Y)$, and
- $Y\mathbf{e}_i, Y(\mathbf{e}_0 - \mathbf{e}_i) \in K$ for all $i \in \{1, ..., n\}$.

The set $M_+(K)$ is obtained from $M(K)$ by adding the constraint that Y is positive semi-definite.

Next is the "project" step, producing $N_+(K) = \{\text{diag}(Y) : Y \in M_+(K)\}$. The extra variable added when K was defined is still present in $N_+(K)$. To get $N_+(P)$, it is removed, by setting $N_+(P) = \{(x_1, ..., x_n) : (1, x_1, ..., x_n) \in N_+(K)\}$.

This process can be iterated: $N_+^0(P) = P$ and $N_+^r(P) = N_+(N_+^{r-1}(P))$.

Definition 4 *For $q \in \mathbb{N}$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_q)$, let $\text{opt}_{LS,r,\boldsymbol{\beta}}(A)$ be the value of the optimal solution to minimizing $\sum_{j=1}^{n} \sum_{\ell=1}^{q} \beta_\ell U_{j\ell}$ subject to membership in $N_+^r(P)$.*

Note that, for $q \geq \text{opt}(A)$, we have $\text{opt}_L(A) = \text{opt}_{LS,0,\boldsymbol{\beta}}(A)$. Here is the gap theorem for LS relaxations.

Theorem 5 *For all $r \in \mathbb{N}$, there are constants $k_0, c > 0$ and a polynomial p, such that, for all $k \in \mathbb{Z}^+$ such that $k > \max\{k_0, 5(r+1)\}$ and $m \geq k^3/(r+1)^2$, there is an $n \in \mathbb{N}$ with $n \leq p(k, \ln m)$, and $A \in \{-1,1\}^{m \times n}$, such that, for any $q \geq 1$, for any $\beta \in \mathbb{N}^q$, $\text{opt}_{LS,r,\boldsymbol{\beta}}(A) \leq k$, but $\text{opt}(A) \geq c\left(\frac{k}{r+1}\right)^2 \ln m$.*

5.2 Setup

As in the proof of Theorem 2, this proof analyzes matrices A chosen at random. Let $\gamma = \frac{r+1}{k}$, $d = \left\lfloor \frac{k^2 \ln m}{43(r+1)^2} \right\rfloor$, $n = \left\lceil \frac{4k^2 \ln m}{(r+1)^2} \right\rceil$, consider $A \in \{-1,1\}^{m \times n}$ whose entries are drawn i.i.d., with $\mathbf{Pr}(A_{ij} = 1) = 1/2 + \gamma$. (Note that, since $k > 5(r+1)$, we have $\gamma < 1/5$.)

Let $S(n, d, \boldsymbol{\beta})$ consist of all $U \in \{0,1\}^{n \times q}$ such that $\sum_{i=1}^n \sum_{j=1}^q \beta_j U_{ij} \leq d$.

Lemma 6 *If k_0 is large enough,*

$$\mathbf{Pr}(\text{opt}(A) \leq d) \leq 1/3.$$

Proof: By construction, for any $U \in S(n, d, \boldsymbol{\beta})$, there is in $\mathbf{x} \in (\mathbb{Z}^+)^n$ such that $A\mathbf{x} \geq \mathbf{1}$ iff U is feasible.

Noting that, as before, $d = \left\lfloor \frac{\ln m}{\gamma^2 43} \right\rfloor$, $n = \left\lceil \frac{4 \ln m}{\gamma^2} \right\rceil$, and $\gamma \leq 1/5$, exactly the same analysis as in the proof of Theorem 2 then implies that

$$\mathbf{Pr}(\text{opt}(A) \leq d) \leq \exp\left(d\left(1 + \ln\left(\frac{n}{d}\right)\right) - m\exp(-14\gamma^2 d)\right).$$

Using the values of d, n and γ from this proof, we get

$$\mathbf{Pr}(\text{opt}(A) \leq d) \leq \exp\left(\frac{k^2 \ln m}{r^2} - m\exp\left(-\frac{14 \ln m}{43}\right)\right),$$

which, as before, is at most $1/3$ for large enough k, since $m \geq k^3/(r+1)^2$. ∎

5.3 A protection lemma

Recall that the LS technique works by tightening the constraints so that $N_+^r(P)$ is an increasingly accurate approximation to the feasible region for the 0/1 integer program associated with A. So, to establish an integrality gap, we would like to show that a fractional solution that is much better than the best integer solution can survive many LS lift-project rounds.

The first step is a "protection lemma", which describes conditions under which a fractional solution survives one round.

To state this protection lemma, the following definition will be helpful. For $\mathbf{x} \in [0,1]^n$, $i \in \{1, ..., n\}$ and $b \in \{0,1\}$, let $\text{round}(\mathbf{x}, i, b)$ be the element of $[0,1]^n$ obtained by replacing x_i with b.

We will use Lemma 9 of [9], attributed there to [23] (see also the third paragraph after Lemma 2.1 in [1]).

Lemma 7 ([23]) *If $\mathbf{x} \in P$ and, for any $i \in \{1, ..., n\}$, $\text{round}(\mathbf{x}, i, 0) \in P$ and $\text{round}(\mathbf{x}, i, 1) \in P$, then $\mathbf{x} \in N_+(P)$.*

For $I \subseteq \{1, ..., n\}$ and $\mathbf{b} \in \{0,1\}^{|I|}$, define $\text{round}(\mathbf{x}, I, \mathbf{b})$ analogously to $\text{round}(\mathbf{x}, i, b)$. As observed in [2], Lemma 7 has the following corollary, which can be proved by induction.

Lemma 8 ([2]) *If* $\mathbf{x} \in P$ *and, for any* $I \in \{1, ..., n\}$ *with* $|I| = r$ *and any* $\mathbf{b} \in \{0,1\}^r$, $\text{round}(\mathbf{x}, I, \mathbf{b}) \in P$, *then* $\mathbf{x} \in N_+^r(P)$.

5.4 Bounding $\text{opt}_{LS,r,\boldsymbol{\beta}}$

Lemma 9 $\mathbf{Pr}(\text{opt}_{LS,r,\boldsymbol{\beta}}(A) \le k) \ge 2/3$.

Proof: Consider the solution U with $U_{11} = ... = U_{n1} = \frac{k}{n}$ and $U_{ij} = 0$ for all $j > 1$.

For a subset I of index pairs, which index into components of U, and $\mathbf{b} \in \{0,1\}^{|I|}$, define $\text{round}(U, I, \mathbf{b})$ analogously to $\text{round}(\mathbf{x}, I, \mathbf{b})$ to be the result of rounding the values indexed by I using the values in $\mathbf{b}$. For any row $\mathbf{a}$ of A, since the components of $\mathbf{a}$ are all in $\{-1, 1\}$, rounding any entry in U can change any component of $AU\mathbf{1}$ by at most 1. Thus, if $\mathbf{x} = U\mathbf{1}$, we have

$$\begin{aligned}\mathbf{Pr}_A(\exists I,\ |I| = r, \mathbf{b} \in \{0,1\}^r, \neg A\ \text{round}(U, I, \mathbf{b})\ \mathbf{1} \ge \mathbf{1}) &\le \mathbf{Pr}_A(\neg AU\mathbf{1} \ge (r+1)\mathbf{1}) \\ &= \mathbf{Pr}_A(\neg A\mathbf{x} \ge (r+1)\mathbf{1}).\end{aligned}$$

For any row $\mathbf{a}$ of A, $\mathbf{E}_{\mathbf{a}}(\mathbf{a} \cdot \mathbf{x}) = 2\gamma k = 2(r+1)$. Applying Lemma 1 and a union bound,

$$\mathbf{Pr}(\neg A\mathbf{x} \ge (r+1)\mathbf{1}) \le m \exp\left(\frac{-2(r+1)^2}{n(2k/n)^2}\right) = m \exp\left(-2 \ln m\right) \le 1/3$$

for large enough k, proving that $\mathbf{Pr}(\text{opt}_{LS,r,\boldsymbol{\beta}}(A) \le k) \ge 2/3$. ∎

5.5 Putting it together

Proof (of Theorem 5): Combining Lemmas 6 and 9, a random A satisfies the requirements of Theorem 5 with probability at least 1/3. ∎

6 Gaps for Sherali-Adams relaxations

In this section, we establish integrality gaps for Sherali-Adams (SA) relaxations of an integer programming formulation of the minimum majority problem.

6.1 Definition of SA relaxations and statement of gap theorem

The description is based on [13], which also gives useful intuition.

Let $P = \{\mathbf{x} : A\mathbf{x} \ge b\}$. Assume that the constraints that define P include $0 \le x_i \le 1$ for all variables i.

The level-r SA relaxation is obtained through the following steps: For each constraint $\mathbf{a} \cdot \mathbf{x} \ge b$ from the original problem, and each pair I and J of disjoint subsets of $[n]$ such that $|I| + |J| = r$, add the constraint

$$(\mathbf{a} \cdot \mathbf{x} - b)\left(\prod_{i \in I} x_i\right)\left(\prod_{j \in J}(1 - x_j)\right) \ge 0.$$

Expand each LHS to express it as a sum of monomials. For each term, replace any x_j^k for $k > 1$ with x_j (the *degree reduction step*). Define a variable Y_S for each set of at most $r+1$ variables. For each LHS, replace each term (which is now a product of a set S of at most $r+1$ variables) with Y_S. (The result is a set of linear inequalities over $\{Y_S : S \subseteq [n], |S| \leq r+1\}$, which defines a polytope Q in $\sum_{j=0}^{r+1} \binom{n}{j}$ dimensions.) Finally, project the result onto $\{Y_S : S \subseteq [n], |S| = 1\}$.

If we denote $\mathrm{SA}^r(P)$ be the level-r SA relaxation of P, then $\mathbf{x} \in \mathrm{SA}^r(P)$ if and only if there is an assignment $\mathbf{y}$ to $\{Y_S : S \subseteq [n], |S| \leq r+1\}$ such that $\mathbf{y} \in Q$ and $x_i = y_{\{i\}}$ for all $i \in [n]$.

Let $N = \sum_{k=0}^{r+1} \binom{n}{k}$ be the dimension of this higher dimensional space.

Definition 10 *Let* $\mathrm{opt}_{SA,r,\boldsymbol{\beta}}(A)$ *be the value of the optimal solution to minimizing*

$$\sum_{j=1}^{n} \sum_{\ell=1}^{q} \beta_\ell U_{j\ell}$$

subject to membership in $\mathrm{SA}^r(P)$.

Theorem 11 *For all* $r \in N$, *there are constants* $k_0, c > 0$ *and a polynomial* p *such that, for all* $k \in \mathbb{Z}^+$ *such that* $k > \max\{k_0, 5(r+1)\}$ *and* $m \geq k^3/(r+1)^2$, *there is an* $n \in N$ *with* $n \leq p(k, \ln m)$ *and there is an* $A \in \{-1,1\}^{m \times n}$, *such that, for any* $q \geq 1$, *for any* $\boldsymbol{\beta} \in \mathbb{N}^q$, $\mathrm{opt}_{SA,r,\boldsymbol{\beta}}(A) \leq k$, *but* $\mathrm{opt}(A) \geq c \left(\frac{k}{r+1}\right)^2 \ln m$.

6.2 A protection lemma

A protection lemma analogous to Lemma 8 also holds for SA relaxations.

Lemma 12 *If* $\mathbf{x} \in P$ *and, for any* $I \in \{1, ..., n\}$ *with* $|I| \leq r$ *and any* $\mathbf{b} \in \{0,1\}^{|I|}$, $\mathrm{round}(\mathbf{x}, I, \mathbf{b}) \in P$, *then* $\mathbf{x} \in \mathrm{SA}^r(P)$.

Proof: Let Q be the polytope in $\mathbb{R}^N$ whose projection yields $\mathrm{SA}^r(P)$. Let ϕ be the mapping from $\mathbb{R}^n$ to $\mathbb{R}^N$ defined by $\phi(\mathbf{x}) = \mathbf{y}$, where $\mathbf{y}$ is indexed by subsets of at most $r+1$ elements of $[n]$, and $y_S = \prod_{i \in S} x_i$. Note that, if $\phi(\mathbf{x}) \in Q$, we have $\mathbf{x} \in \mathrm{SA}^r(P)$.

Now, choose $\mathbf{x}$ satisfying the hypotheses of the lemma. To prove that $\mathbf{x} \in \mathrm{SA}^r(P)$, it suffices to prove that $\phi(\mathbf{x}) \in Q$. Choose a constraint from among those defining Q: suppose that it was derived from the original constraint $\mathbf{ax} \geq b$ together with index sets I and J. The constraint in $\mathbb{R}^N$ arising from the Sherali-Adams process is $\sum_{J' \subseteq J} (-1)^{|J'|} \left(\sum_{\ell=1}^{n} a_i Y_{S \cup J' \cup \{\ell\}} - b Y_{S \cup J'} \right) \geq 0$ (see [13]). Thus, it suffices to prove

$$\sum_{J' \subseteq J} (-1)^{|J'|} \left(\sum_{i=1}^{n} a_i \prod_{\ell \in I \cup J' \cup \{i\}} x_\ell - b \prod_{\ell \in I \cup J'} x_\ell \right) \geq 0. \tag{4}$$

Toward this end, consider $\mathbf{x}'$ obtained by rounding $\mathbf{x}$ as follows. Note that the LHS of (4) is linear in each x_ℓ, if the other components of $\mathbf{x}$ are fixed. Thus, for each $\ell \in I \cup J$, the LHS of (4) is either non-increasing or non-decreasing in x_ℓ as it varies between 0 and 1. Thus, there is a choice of how to round x_ℓ to either 0 or 1 that does not change the sign of the LHS from non-negative to negative, or from negative to non-negative. Let $\mathbf{x}'$ be constructed from $\mathbf{x}$ by rounding elements of $I \cup J$ one at a time in this manner.

We consider two cases. First suppose that a member of I was rounded to 0 or a member of J was rounded to 1. Then $(\mathbf{a}\cdot\mathbf{x}'-b)(\prod_{i\in I} x'_i)(\prod_{j\in J}(1-x'_j)) = 0$, and, since all the variables affected by the degree reduction step are in $I \cup J$, and this degree reduction has no effect for elements of $\{0,1\}$, this implies

$$\sum_{J'\subseteq J}(-1)^{|J'|}\left(\sum_{i=1}^{n} a_i \prod_{\ell\in I\cup J'\cup\{i\}} x'_\ell - b\prod_{\ell\in I\cup J'} x'_\ell\right) = 0.$$

Since the transformation from $\mathbf{x}$ to $\mathbf{x}'$ did not change the sign from negative to non-negative, this implies that (4) holds.

Now, suppose that all members of I were rounded to 1, and all members of J were rounded to 0. Then $\sum_{J'\subseteq J}(-1)^{|J'|}\left(\sum_{i=1}^{n} a_i \prod_{\ell\in I\cup J'\cup\{i\}} x'_\ell - b\prod_{\ell\in I\cup J'} x'_\ell\right) = \mathbf{a}\cdot\mathbf{x}'-b \geq 0$, since $\mathbf{x}'$ was obtained from $\mathbf{x}$ by rounding at most r components. This completes the proof. ∎

6.3 Putting it together

Proof (of Theorem 11): The proof is exactly the same as the proof of Theorem 5, except replacing Lemma 8 with Lemma 12. ∎

7 A simple and direct proof of an upper bound

The algorithm solves the more general problem: for $A \in \{-1,1\}^{m\times n}$ and $\mathbf{b} \in \mathbb{Z}^m$, minimize $\sum_{i=1}^n x_i$ subject to $A\mathbf{x} \geq \mathbf{b}$, and $\mathbf{x} \in (\mathbb{Z}^+)^n$. We have a problem for each $(A,\mathbf{b})$ pair; let $\mathrm{opt}(A,\mathbf{b})$ be the value of its optimal solution. (Let us use the shorthand $\mathrm{opt}(A)$ for $\mathrm{opt}(A,(1,1,...,1)^T)$. Then, if, for each variable i, we denote the ith column of A by A_i, we have

$$\mathrm{opt}(A,\mathbf{b}) = 1 + \min_i \mathrm{opt}(A, \mathbf{b} - A_i). \tag{5}$$

To see this, consider commiting to make $x_i \geq 1$. This gives rise to a subproblem of a similar form with the constraints updated as indicated. Obviously, if no component of $\mathbf{b}$ is positive, then $\mathrm{opt}(A,\mathbf{b}) = 0$.

The algorithm exploits this recursive structure using a parameter $\eta \in (0,1/2)$ that will be set using the analysis:

- if no component of $\mathbf{b}$ is positive, return $(0,0,...,0)$ and halt,
- otherwise,
  - choose i to minimize $\sum_{t=1}^m \exp(\eta(b_t - A_{ti}))$,
  - recurse to solve the problem with $\mathbf{b}$ replaced by $\mathbf{b} - A_i$, getting $\mathbf{x}$,
  - return the solution obtained from $\mathbf{x}$ by adding 1 to x_i.

Let us call $\sum_{t=1}^m \exp(\eta b_t)$ the *potential.*

Our analysis of this algorithm uses the following, which is essentially the discriminator lemma [24].

Lemma 13 *For any $\mathbf{r} \in [0,\infty)^m$, there is an i such that $\mathbf{r}^T A_i \geq \frac{\mathbf{r}^T\mathbf{b}}{\mathrm{opt}(A,\mathbf{b})}$.*

Proof: First, since all components of $\mathbf{r}$ are non-negative, $\mathbf{r}^T A\mathbf{x} \geq \mathbf{r}^T\mathbf{b}$ for all feasible $\mathbf{x}$. If $\mathbf{x}^*$ is an optimum, if we sample i from $\mathbf{x}^*/||\mathbf{x}^*||_1 = \mathbf{x}^*/\text{opt}(A, \mathbf{b})$, we have $\mathbf{E}(\mathbf{r}^T A_i) = \frac{\mathbf{r}^T A\mathbf{x}^*}{||\mathbf{x}^*||_1} \geq \frac{\mathbf{r}^T b}{||\mathbf{x}^*||_1}$, completing the proof. ∎

Lemma 13 implies that,

$$\forall \mathbf{r} \in (\mathbb{R}^+)^m,\ \exists i,\ \frac{\mathbf{r} \cdot A_i}{||\mathbf{r}||_1} \geq \frac{1}{\text{opt}(A)}. \tag{6}$$

Note that this is a property of A, and that A does not change as the algorithm progresses, so that (6) always remains true throughout the recursion.

Now, we want to bound the reduction in the potential prior to the recursive call. Let $P_{new} = \min_{i\in[n]} \sum_{t=1}^m \exp(\eta(b_t - A_{ti}))$ be the new potential, and $P_{old} = \sum_{t=1}^m \exp(\eta b_t)$ be the old potential. We have

$$\begin{aligned}
&P_{new}/P_{old} \\
&= \frac{1}{P_{old}} \min_{i\in[n]} \sum_{t=1}^{m} \exp(\eta(b_t - A_{ti})) \\
&= \frac{1}{\sum_{t=1}^m \exp(\eta b_t)} \min_{i\in[n]} \sum_{t=1}^{m} \left(\frac{e^{-\eta} + e^{\eta}}{2} + \frac{e^{-\eta} - e^{\eta}}{2} A_{ti}\right) \exp(\eta b_t) \\
&\leq \frac{e^{-\eta} + e^{\eta}}{2} + \frac{e^{-\eta} - e^{\eta}}{2\text{opt}},
\end{aligned}$$

by (6). Applying Taylor series, we have

$$P_{new}/P_{old} \leq 1 - \frac{\eta}{\text{opt}} + (1/2 + o(1))\eta^2.$$

Setting $\eta = \frac{1}{\text{opt}}$ (and we may assume w.l.o.g. that the algorithm "knows" opt, since it can guess progressively larger values), we get

$$P_{new}/P_{old} \leq 1 - (1/2 - o(1))\frac{1}{\text{opt}^2}.$$

If $\mathbf{b} = (1, 1, \ldots 1)$, the potential starts at $m \exp(\eta)$. After s recursive calls, it is at most

$$\left(1 - (1/2 - o(1))\frac{1}{\text{opt}^2}\right)^s m \exp(\eta).$$

When all components of $\mathbf{b}$ are at most 0, we may stop. If we have not stopped, the potential is at least e^η. Thus, on input $(A, (1, 1, \ldots, 1)^T)$, the number of recursive calls before stopping is at most $(2 + o(1))\text{opt}(A)^2 \ln m$. This bounds the total number of times that the solution is incremented when returning from the nested recursion, completing the proof.

Acknowledgments

I would like to thank Dave Helmbold and Kunal Talwar for valuable conversations. I am also grateful to anonymous reviewers for their careful reading of the paper, and their helpful comments and suggestions.

References

[1] Michael Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong nonapproximability results in the Lovász-Schrijver hierarchy. In *STOC*, pages 294–303, 2005.

[2] Mikhail Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong nonapproximability results in the Lovász-Schrijver hierarchy. *Computational Complexity*, 20(4):615–648, 2011.

[3] Edoardo Amaldi and Viggo Kann. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, 209(1-2):237–260, 1998.

[4] Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis. Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(2):19–51, 2006.

[5] Abbas Bazzi, Samuel Fiorini, Sebastian Pokutta, and Ola Svensson. No small linear program approximates vertex cover within a factor 2 - ε. *Math. Oper. Res.*, 44(1):147–172, 2019.

[6] Siavosh Benabbas and Avner Magen. Extending SDP integrality gaps to Sherali-Adams with applications to quadratic programming and maxcutgain. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 299–312. Springer, 2010.

[7] Aditya Bhaskara, Moses Charikar, Aravindan Vijayaraghavan, Venkatesan Guruswami, and Yuan Zhou. Polynomial integrality gaps for strong sdp relaxations of densest k-subgraph. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 388–405. Society for Industrial and Applied Mathematics, 2012.

[8] Cristian Bucilu, Rich Caruana, and Alexandru Niculescu-Mizil. Model compression. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 535–541. ACM, 2006.

[9] Joshua Buresh-Oppenheim, Nicola Galesi, Shlomo Hoory, Avner Magen, and Toniann Pitassi. Rank bounds and integrality gaps for cutting planes procedures. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 318–327. IEEE, 2003.

[10] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 283–292. ACM, 2009.

[11] Jeff Cheeger, Bruce Kleiner, and Assaf Naor. A $(\log n)^{\Omega(1)}$ integrality gap for the sparsest cut SDP. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 555–564. IEEE, 2009.

[12] Eden Chlamtác and Pasin Manurangsi. Sherali-Adams integrality gaps matching the log-density threshold. In *APPROX*, pages 10:1–10:19, 2018.

[13] Eden Chlamtac and Madhur Tulsiani. Convex relaxations and integrality gaps. In *Handbook on semidefinite, conic and polynomial optimization*, pages 139–169. Springer, 2012.

[14] Ali Civril. A note on the hardness of sparse approximation. *Information Processing Letters*, 113(14):543–545, 2013.

[15] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–61. Society for Industrial and Applied Mathematics, 2007.

[16] N. Duffy and D. Helmbold. Potential boosters? In *Proceedings of NIPS*, pages 258–264, 1999.

[17] Noah Fleming, Pravesh Kothari, Toniann Pitassi, et al. Semialgebraic proofs and efficient algorithm design. *Foundations and Trends® in Theoretical Computer Science*, 14(1-2):1–221, 2019.

[18] Dean Foster, Howard Karloff, and Justin Thaler. Variable selection is hard. In *Conference on Learning Theory*, pages 696–709, 2015.

[19] Y. Freund and Robert E. Schapire:. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997. Special Issue for EuroCOLT '95.

[20] Naveen Garg, Vijay V. Vazirani, and Mihalis Yannakakis. Primal-dual approximation algorithms for integral flow and multicut in trees. *Algorithmica*, 18(1):3–20, 1997.

[21] Konstantinos Georgiou, Avner Magen, Toniann Pitassi, and Iannis Tourlakis. Integrality gaps of 2-o(1) for vertex cover SDPs in the Lovász–Schrijver hierarchy. *SIAM Journal on Computing*, 39(8):3553–3570, 2010.

[22] Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. Optimal Sherali-Adams gaps from pairwise independence. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 125–139. Springer, 2009.

[23] Michel X Goemans and Levent Tunçel. When does the positive semidefiniteness constraint help in lifting procedures? *Mathematics of Operations Research*, 26(4):796–815, 2001.

[24] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J. Comp. Sys. Sci.*, 46:129–154, 1993.

[25] J. Jackson and M. Craven. Learning sparse perceptrons. In *NIPS 8*, pages 654–660, 1996.

[26] Anna R Karlin, Claire Mathieu, and C Thach Nguyen. Integrality gaps of linear and semidefinite programming relaxations for knapsack. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 301–314. Springer, 2011.

[27] Steven Kelk and Georgios Stamoulis. Integrality gaps for colorful matchings. *Discrete Optimization*, 32:73–92, 2019.

[28] Subhash Khot and Rishi Saket. SDP integrality gaps with local ℓ_1-embeddability. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 565–574. IEEE, 2009.

[29] Subhash A Khot and Nisheeth K Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into ℓ_1. *Journal of the ACM (JACM)*, 62(1):8, 2015.

[30] M. Krishnapur. Anti-concentration inequalities, 2016.

[31] Adam Kurpisz, Monaldo Mastrolilli, Claire Mathieu, Tobias Mömke, Victor Verdugo, and Andreas Wiese. Semidefinite and linear programming integrality gaps for scheduling identical machines. *Math. Program.*, 172(1-2):231–248, 2018.

[32] Monique Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0–1 programming. *Mathematics of Operations Research*, 28(3):470–496, 2003.

[33] P. M. Long. Using the pseudo-dimension to analyze approximation algorithms for integer programming. *Proceedings of the Seventh International Workshop on Algorithms and Data Structures*, 2001.

[34] P. M. Long. Minimum majority classification and boosting. *Proceedings of the The Eighteenth National Conference on Artificial Intelligence*, 2002.

[35] L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13:383–390, 1975.

[36] László Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0–1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

[37] L. Mason, J. Baxter, P. L. Bartlett, and M. Frean. Boosting algorithms as gradient descent. In *Advances in Neural Information Processing Systems 12*, pages 512–518. MIT Press, 2000.

[38] Claire Mathieu and Alistair Sinclair. Sherali-Adams relaxations of the matching polytope. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 293–302. ACM, 2009.

[39] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[40] V. V. Petrov. *Limit theorems of probability theory*. Oxford Science Publications, Oxford, England, 1995.

[41] D. Pollard. *Convergence of Stochastic Processes*. Springer Verlag, Berlin, 1984.

[42] J. R. Quinlan. Some elements of machine learning. *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 523–524, 1999.

[43] Prasad Raghavendra and David Steurer. Integrality gaps for strong SDP relaxations of unique games. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 575–585. IEEE, 2009.

[44] R. E. Schapire, Y. Freund, P. L. Bartlett, and W. S. Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics*, 26(5):1651–1686, 1998.

[45] R.E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, 1999.

[46] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 593–602. IEEE, 2008.

[47] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovász-Schrijver LP relaxations of vertex cover and max cut. In *STOC*, pages 302–310. ACM, 2007.

[48] R. A. Servedio. Every linear threshold function has a low-weight approximator. *Computational Complexity*, 16(2):180–209, 2007.

[49] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3(3):411–430, 1990.

[50] E. Slud. Distribution inequalities for the binomial law. *Annals of Probability*, 5:404–412, 1977.

[51] Christino Tamon and Jie Xiang. On the boosting pruning problem. In *European Conference on Machine Learning*, pages 404–412. Springer, 2000.

[52] Iannis Tourlakis. Towards optimal integrality gaps for hypergraph vertex cover in the Lovász-Schrijver hierarchy. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 233–244. Springer, 2005.

[53] Yuchen Zhang, Martin J Wainwright, and Michael I Jordan. Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In *Conference on Learning Theory*, pages 921–948, 2014.