Convex Relaxations and Integrality Gaps

Eden Chlamtac[1] and Madhur Tulsiani[2]

[1] Tel Aviv University, Tel Aviv, Israel. chlamtac@post.tau.ac.il
[2] Institute for Advanced Study, Princeton, NJ. madhurt@math.ias.edu

Summary. We discuss the effectiveness of linear and semidefinite relaxations in approximating the optimum for combinatorial optimization problems. Various hierarchies of these relaxations, such as the ones defined by Lovász and Schrijver [47], Sherali and Adams [55] and Lasserre [42] generate increasingly strong linear and semidefinite programming relaxations starting from a basic one. We survey some positive applications of these hierarchies, where their use yields improved approximation algorithms. We also discuss known lower bounds on the integrality gaps of relaxations arising from these hierarchies, demonstrating limits on the applicability of such hierarchies for certain optimization problems.

1 Introduction

Convex relaxations are one of the most powerful techniques for designing polynomial time approximation algorithms for NP-hard optimization problems such as Chromatic Number, MAX-CUT, Minimum Vertex Cover etc. Approximation algorithms for these problems are developed by formulating the problem at hand as an integer program. One then relaxes the integer program to a convex program which can be solved in polynomial time, such as a linear program (LP) or semidefinite program (SDP). A solution to the combinatorial problem is then obtained by designing a (possibly randomized) polynomial-time algorithm to convert the solution of such a convex relaxation, to an integer solution for the combinatorial problem, often referred to as "rounding".

If we are dealing with (say) a maximization problem for which the true combinatorial optimum is OPT, then the convex relaxation will achieve a value FRAC which is at least as large as OPT (as the integer solution is also a feasible solution to the convex program). The rounding algorithm then uses the solution of the convex relaxation with objective value FRAC to produce an integer solution with (possibly suboptimal) value ROUND. The analysis of the algorithm then boils down to a comparison of these three quantities which satisfy $\mathsf{ROUND} \leq \mathsf{OPT} \leq \mathsf{FRAC}$. The inequalities are reversed for a minimization problem.

If one just thinks of the combinatorial problem as a question of finding the (say) maximum *value* of the objective (e.g. the *size* of the maximum cut in a graph),

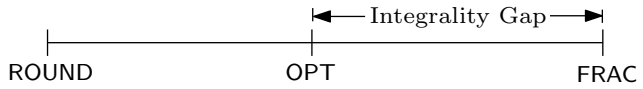

Fig. 1. The integrality gap for a maximization problem

then the rounding algorithm is not needed and the quality of approximation is measured by the ratio FRAC/OPT. If instead, the question is to *search* for an optimum integer solution (e.g. a cut of maximum value), then one is interested in the ratio OPT/ROUND. However, the analyses of most approximation algorithms actually require proving an upper bound on FRAC/ROUND (which in turn bounds OPT/ROUND), simply because OPT is not known for an arbitrary instance of the problem! Thus a lower bound on the ratio FRAC/OPT not only gives a lower bound on the quality of approximation for the objective value, but also on the performance of rounding algorithms (if the algorithms are analyzed by comparing FRAC and ROUND). This ratio is called the *integrality gap* of the program. Figure 1 shows the relationship between these quantities.

For a (maximization) problem, the integrality gap of a program is defined as the supremum of the ratio FRAC/OPT over all instances of the problem. For a minimization, we take it to be the supremum of the inverse ratio. Note that the integrality gap is always at least 1 and a large gap indicates a poor approximation ratio. Conversly, the analysis of an approximation algorithm bounding FRAC/ROUND in turn gives an *upper bound* on the integrality gap. In cases when the integrality gap is infinite, we express it as a function of the size of the instance, in which case it is defined as the maximum of the relevant ratio over all instances of the given size.

2 Hierarchies of Convex Relaxations

Convex relaxations for various combinatorial problems can be strengthened by including additional constraints which are satisfied by an integer solution. This process of generating stronger relaxations by adding larger (but still, local) constraints is captured by various hierarchies of convex relaxations such as the ones defined by Lovász and Schrijver [47], Sherali and Adams [55] and Lasserre [42]. Starting from a basic relaxation, these hierarchies define various *levels* of convex relaxations for a problem, with the relaxations at a higher level being more powerful than the relaxations at lower levels.

These hierarchies are known to capture the convex relaxations used in the best available algorithms for many problems, such as the SDP relaxation for Sparsest Cut by Arora, Rao and Vazirani [7] and the ϑ-function of Lovász for Maximum Independent Set [46], within a constant number of levels. It is also known that for an integer program with n variables taking values in $\{0, 1\}$, the convex program obtained by n levels of any of the above hierarchies has integrality gap 1, i.e., it gives the exact solution. However, solving the program obtained by t levels of these hierarchies takes time $n^{O(t)}$ which is exponential in n for $t = \Omega(n)$.

The interesting question is then to characterize the problems for which the t^{th} level of these hierarchies yields a better approximation, for a small t. On the other hand, a lower bound showing that the integrality gap of the program obtained after many (say even $\Omega(n)$) levels of a hierarchy remains large, is a strong lower bound against a class of algorithms capturing most known ones.

We describe below each of these hierarchies [3] . We shall use the example of Maximum Independent Set throughout this chapter to illustrate the differences in the programs obtained by the various hierarchies. An excellent comparison of all the three hierarchies mentioned above is also available in [43]. The basic LP and SDP relaxations for Maximum Independent Set (also known as Maximum Stable Set) are given in Figure 2. We follow the convention used in the study of approximation algorithms of writing the SDP directly in terms of inner products of vectors in the Cholesky decomposition[4] of the PSD matrix of variables.

LP relaxation		SDP relaxation	
maximize	$\sum_{i \in V} x_i$	maximize	$\sum_{i \in V} \|\mathbf{u}_i\|^2$
subject to	$x_i + x_j \leq 1 \quad \forall (i,j) \in E$	subject to	$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \quad \forall (i,j) \in E$
	$x_i \in [0,1]$		$\langle \mathbf{u}_i, \mathbf{u}_0 \rangle = \|\mathbf{u}_i\|^2 \quad \forall i \in V$
			$\|\mathbf{u}_0\| = 1$

Fig. 2. LP and SDP relaxations for Maximum Independent Set

2.1 The Lovász- Schrijver Hierarchies

Lovász and Schrijver [47] describe two versions of a "lift-and-project" method. This can be thought of as an operator which when applied to a convex programming relaxation $\mathcal{P}$ of a 0/1 integer linear program, produces a tighter relaxation. A weaker version of the method, denoted LS, adds auxiliary variables and linear inequalities, and the projection of the new relaxation on the original variables is denoted by $N(\mathcal{P})$; a stronger version, denoted LS+, adds semidefinite programming constraints as well, and the projection on the original variables is denoted by $N_+(\mathcal{P})$.

Starting from a basic relaxation and iteratively applying the operator N (N_+) one gets higher and higher levels (which are called *rounds* for the Lovász-Schrijver hierarchies due to their iterative nature) of the LS (LS+) hierarchy. Thus, the relaxation obtained by r rounds of the hierarchy is given by $N(\cdots N(\mathcal{P}) \cdots)$ where the operator is applied t times. We denote it as $N^t(\mathcal{P})$.

[3] Some of these hierarchies can be defined in a more general context. However, we shall limit our discussion to relaxations of 0/1 (or $-1/1$) integer programs.

[4] The Cholesky decomposition of an $n \times n$ PSD matrix X is a collection of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ satisfying for all i, j, $X_{ij} = \langle \mathbf{u}_i, \mathbf{u}_j \rangle$.

Lovász and Schrijver also prove that if we start from a linear programming relaxation of a 0/1 integer program with n variables, then n applications of the LS procedures are sufficient to obtain a tight relaxation where the only feasible solutions are convex combinations of integral solutions. If one starts from a linear program with poly(n) inequalities, then it is possible to optimize over the set of solutions defined by t rounds of LS or LS+ in $n^{O(t)}$ time.

To describe these hierarchies it will be more convenient to work with *convex cones* rather than arbitrary convex subsets of $[0,1]^n$. Specifically, if we are interested in a convex set $\mathcal{P} \subseteq [0,1]^n$ (which might be the feasible region of our starting convex relaxation), we first convert it into the cone $\text{cone}(\mathcal{P}) \subseteq \mathbb{R}^{n+1}$ defined as the set of all vectors $(\lambda, \lambda y_1, \ldots, \lambda y_N)$ such that $\lambda \geq 0$ and $(y_1, \ldots, y_n) \in \mathcal{P}$. For example, in the "cone" linear programming relaxation of the Maximum Independent Set problem $(y_0, y_1, \ldots, y_n)$ is in the feasible region (denoted by $\text{cone}(IS(G))$) if and only if

$$\begin{aligned} y_i + y_i &\leq y_0 \quad \forall (i,j) \in E \\ 0 \leq y_i &\leq y_0 \quad \forall i \in V \\ y_0 &\geq 0 \qquad\qquad\qquad\qquad (\text{cone}(IS(G))) \end{aligned}$$

We would now like to "tighten" the relaxation by adding inequalities (on the solution obtained after scaling to get $y_0 = 1$) that are valid for 0/1 solutions but that are violated by other solutions. Ideally, we would like to say that a solution $(1, y_1, \ldots, y_n)$ must satisfy the conditions $y_i^2 = y_i$, because such a condition is satisfied only by 0/1 solutions. Equivalently, we could introduce n^2 new variables $Y_{i,j}$ and add the conditions (i) $Y_{i,j} = y_i \cdot y_j$ and (ii) $Y_{i,i} = y_i$. Unfortunately, condition (i) is neither linear nor convex, and so we will instead "approximate" condition (i) by enforcing a set of linear conditions that are implied by (but not equivalent to) (i). This is formalized in the definition below.

Definition 1. *For a cone $K \subseteq \mathbb{R}^d$ we define the set $N(K)$ (also a cone in $\mathbb{R}^d$) as follows: a vector $\mathbf{y} = (y_0, \ldots, y_{d-1}) \in \mathbb{R}^d$ is in $N(K)$ if and only if there is a matrix $Y \in \mathbb{R}^{d \times d}$ such that*

1. *Y is symmetric*
2. *For every $i \in \{0, 1, \ldots, d-1\}$, $Y_{0,i} = Y_{i,i} = y_i$*
3. *Each row Y_i is an element of K*
4. *Each vector $Y_0 - Y_i$ is an element of K*

In such a case, Y is called the protection matrix *of $\mathbf{y}$. If, in addition, Y is* positive semidefinite*, then $\mathbf{y} \in N_+(K)$. We define $N^0(K)$ and $N_+^0(K)$ as K, and $N^t(K)$ (respectively, $N_+^t(K)$) as $N(N^{t-1}(K))$ (respectively, $N_+(N_+^{t-1}(K))$). When $K = cone(\mathcal{P})$ for $\mathcal{P} \subseteq \mathbb{R}^{d-1}$, we denote as $N^t(\mathcal{P})$ the set $\left\{\mathbf{y} \in \mathbb{R}^{d-1} \mid (1, \mathbf{y}) \in N^t(cone(\mathcal{P}))\right\}$, and similarly for $N_+^t(\mathcal{P})$.*

Let us see that these operators are in fact relaxations for condition (i) above. Indeed, if $\mathbf{y} = (1, y_1, \ldots, y_{d-1}) \in \{0,1\}^d$, then we can set $Y_{i,j} = y_i \cdot y_j$. Such a matrix Y is clearly positive semidefinite, and it satisfies $Y_{i,i} = y_i^2 = y_i$ if the y_i are in $\{0,1\}$.

Consider now a row Y_i of Y, that is, the vector $\mathbf{r}$ such that $r_j := Y_{i,j} = y_i \cdot y_j$. Then, either $y_i = 0$, in which case $\mathbf{r} = (0, \ldots, 0)$ is in every cone, or $y_i = 1$, and $\mathbf{r} = \mathbf{y}$. Similarly, if we consider $r_j := Y_{0,j} - Y_{i,j} = (1 - y_i) \cdot y_j$ we find that $\mathbf{r}$ either equals the all-zero vector or it equals $\mathbf{y}$. This shows that if $\mathbf{y} = (1, y_1, \ldots, y_{d-1}) \in \{0,1\}^d$ and $\mathbf{y} \in K$, then also $\mathbf{y} \in N_+^t(K)$ for every t. Hence, if $K \cap \{y_0 = 1\}$ defines a relaxation of the integral problem, so does $N_+^t(K) \cap \{y_0 = 1\}$, and hence also $N^t(K) \cap \{y_0 = 1\}$.

For a graph G, the relaxation of the Maximum Independent Set problem resulting from t rounds of LS+ is the result of

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} y_i \\ \text{subject to} \quad & (y_0, \ldots, y_n) \in N_+^t(\text{cone}(IS(G))) \\ & y_0 = 1 \end{aligned}$$

2.2 The Sherali-Adams Hierarchy

The Sherali-Adams hierarchy [55] defines a hierarchy of linear programs which give increasingly tighter relaxations. To see the intuition behind the hierarchy, we may view it as a strengthening of the LS procedure. Recall that the solution to a 0/1 integer program can be specified by a vector $\mathbf{y} \in \{0,1\}^n$. In the Lovász-Schrijver hierarchy we defined auxiliary variables Y_{ij} and wanted to express the constraint that $Y_{ij} = y_i \cdot y_j$. We then expressed it by some implied linear conditions on the variables Y_{ij}.

Consider a solution $(1, y_1, \ldots, y_n)$ which is feasible at the second level of the LS hierarchy. Then the row Y_i of the protection matrix must also define a feasible solution to the "cone" version of the relaxation, say cone$(IS(G))$. Now, the solution $\mathbf{y}' = Y_i = (Y_{i0}, Y_{i1}, \ldots, Y_{in})$ must be feasible for the first level, and so there exists a protection matrix Y' for it. Now, we would also like to think of Y'_{jk} as a relaxation for $y_i y_j y_k$. However, notice that the choice of protection matrix Y' was dependent on the fact that we first chose the row Y_i. In particular, if we looked at the protection matrix Y'' for the solution $\mathbf{y}'' = Y_j = (Y_{j0}, Y_{j1}, \ldots, Y_{jn})$, it need not be true that $Y'_{jk} = Y''_{ik}$.

The Sherali-Adams hierarchy solves this problem by introducing all the auxiliary variables at once instead of by an inductive process. In particular, we define a variable Y_S for each $S \subseteq [n]$ with $|S| \leq t + 1$. The intuition again is that we want to impose $Y_S = \prod_{i \in S} y_i$. However, we instead impose some linear conditions implied by this. For every constraint $a^\mathsf{T}\mathbf{y} - b \leq 0$ of the starting LP relaxation, we consider sets S, T such that $|S| + |T| \leq t$ and impose a linear implication of $(a^\mathsf{T}\mathbf{y} - b) \cdot \prod_{i \in S} y_i \cdot \prod_{j \in T}(1 - y_j) \leq 0$, by requiring that

$$\sum_{T' \subseteq T} (-1)^{|T'|} \cdot \left(\sum_{i=1}^{n} a_i \cdot Y_{S \cup T' \cup \{i\}} - b \cdot Y_{S \cup T'} \right) \leq 0$$

Note again that the number of variables and constraints in the LP at level t is $n^{O(t)}$ and hence it can be solved in time $n^{O(t)}$. Also, each such program is a relaxation,

since for any $\mathbf{y} \in \{0,1\}^n$ satisfying the initial constraints, $Y_S = \prod_{i \in S} y_i$ defines a valid level-t solution. The program below gives the relaxation of Maximum Independent Set obtained at the t^{th} level of the Sherali-Adams hierarchy.

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{i=1}^{n} Y_{\{i\}} & \\
\text{subject to} & \displaystyle\sum_{T' \subseteq T} (-1)^{|T'|} \cdot \left[Y_{S \cup T' \cup \{j\}} + Y_{S \cup T' \cup \{i\}} - Y_{S \cup T'} \right] \leq 0 & |S| + |T| \leq t,\ (i,j) \in E \\
& 0 \leq \displaystyle\sum_{T' \subseteq T} (-1)^{|T'|} \cdot Y_{S \cup T' \cup \{i\}} \leq \sum_{T' \subseteq T} (-1)^{|T'|} \cdot Y_{S \cup T'} & |S| + |T| \leq t,\ i \in V \\
& Y_{\emptyset} = 1 &
\end{array}
$$

Fig. 3. Sherali-Adams relaxation for Maximum Independent Set

Since the above program is a convex relaxation, any convex combination of 0/1 solutions is also a solution to the program. It is convenient to think of the convex combination as defining a *distribution* over 0/1 solutions. With this interpretation, we can think of Y_S as the *probability* that all variables in set S are equal to 1. It is easy to show that the feasible sets for Sherali-Adams relaxations are characterized by solutions which "locally" (for every small subset of variables) look like valid distributions.

Lemma 1. *Consider a family of distributions $\{\mathcal{D}(S)\}_{S \subseteq [n]:|S| \leq t+2}$, where each $\mathcal{D}(S)$ is defined over $\{0,1\}^S$. If the distributions satisfy*

1. *For all $(i,j) \in E$ and $S \supseteq \{i,j\}$, $\mathbb{P}_{\mathcal{D}(S)}\left[(y_i = 1) \wedge (y_j = 1)\right] = 0$, and*
2. *For all $S' \subseteq S \subseteq [n]$ with $|S| \leq t+1$, the distributions $\mathcal{D}(S'), \mathcal{D}(S)$ agree on S'.*

Then $Y_S = \mathbb{P}_{\mathcal{D}(S)}\left[\bigwedge_{i \in S}(y_i = 1)\right]$ is a feasible solution for the above level-t Sherali-Adams relaxation. Conversely, for any feasible solution $\{Y'_S\}$ for the level-$(t+1)$ Sherali-Adams relaxation, there exists a family of distributions satisfying the above properties, as well as $\mathbb{P}_{\mathcal{D}(S)}\left[\bigwedge_{i \in S'}(y_i = 1)\right] = Y'_{S'}$ for all $S' \subseteq S \subset [n]$ s.t. $|S| \leq t+1$.

Extending further this intuition of the variables Y_S as probabilities, we can also define variables for arbitrary events over a set S of size at most t. A basic event is given by a *partial assignment* $\alpha \in \{0,1\}^S$ which assigns value 0 to some variables in S (which we denote by $\alpha^{-1}(0)$) and 1 to the others (denoted $\alpha^{-1}(1)$). We can define variables $X_{(S,\alpha)}$ when $|S| \leq t$ and $\alpha \in \{0,1\}^S$ as

$$X_{(S,\alpha)} := \sum_{T \subseteq \alpha^{-1}(0)} (-1)^{|T|} Y_{\alpha^{-1}(1) \cup T}.$$

Note that the previous constraints imply that $X_{(S,\alpha)} \geq 0$ for all (S,α). For a 0/1 solution, the intended values are $X_{(S,\alpha)} = \prod_{i \in \alpha^{-1}(1)} y_i \prod_{i \in \alpha_{-1}(0)} (1 - y_i)$. The previous program can easily be re-written in terms of the variables $X_{(S,\alpha)}$. This formulation in

terms of variables for partial assignments also extends to problems where the variables in the integer program take values not in $\{0, 1\}$ but in a larger finite domain $[q]$.

2.3 The Lasserre Hierarchy

The Lasserre hierarchy gives a sequence of increasingly tight semidefinite programming relaxations for a quadratic 0/1 program. As in the case of the Sherali-Adams hierarchy, the semidefinite program after t rounds of the Lasserre hierarchy also introduces a new (vector valued) variable for the product of every t variables in the original program.

For concreteness, we consider the program for Maximum Independent Set. The same procedure can be used to derive the level-t SDP for any problem formulated as a quadratic integer program, with variables taking values in $\{0, 1\}$. Given a graph $G = (V, E)$, the integer program would have a variable X_i for each $i \in V$ with $y_i = 1$ if i is in the independent set and $y_i = 0$ otherwise. To ensure that the solution is an independent set, we would enforce that $y_i \cdot y_j = 0$ for all $(i, j) \in E$.

To obtain the Lasserre relaxation, we first think of an integer program which has a variable Y_S for each $S \subseteq V, |S| \leq t$ where the intended solution, as before, is $Y_S = 1$ iff all vertices in S are in the independent set. We can then add the constraint that the product $Y_{S_1} \cdot Y_{S_2}$ must only depend on $S_1 \cup S_2$. For homogenization, we introduce an extra variable $Y_\emptyset$ which is always supposed to be 1. Replacing the integer variables Y_S by vectors $\mathbf{U}_S$ gives the semidefinite relaxation as below, where all sets S_i are assumed to be of cardinality at most $t + 1$.

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{i \in V} \left\|\mathbf{U}_{\{i\}}\right\|^2 & \\
\text{subject to} & \left\langle \mathbf{U}_{\{i\}}, \mathbf{U}_{\{j\}} \right\rangle = 0 & \forall\ (i, j) \in E \\
& \left\langle \mathbf{U}_{S_1}, \mathbf{U}_{S_2} \right\rangle = \left\langle \mathbf{U}_{S_3}, \mathbf{U}_{S_4} \right\rangle & \forall\ S_1 \cup S_2 = S_3 \cup S_4 \\
& \|\mathbf{U}_\emptyset\|^2 = 1 &
\end{array}
$$

Fig. 4. Lasserre SDP for Maximum Independent Set

Note that the program for level t only has vectors for sets of size at most $t + 1$. It can be shown that for any set S with $|S| \leq t$, the vectors $\mathbf{U}_{S'}, S' \subseteq S$ induce a probability distribution over valid independent sets of the subgraph induced by S. However, unlike the Sherali-Adams hierarchy, the existence of such distributions is not a sufficient condition for the existence of a feasible solution for the semidefinite program.

As in the case of the Sherali-Adams hierarchy, one can also write the above program in terms of vectors $\mathbf{V}_{(S,\alpha)}$ for partial assignments, which can be derived from the vectors $\mathbf{U}_S$ (for 0/1 integer programs) as

$$\mathbf{V}_{(S,\alpha)} \quad := \quad \sum_{T \subseteq \alpha^{-1}(0)} (-1)^{|T|} \mathbf{U}_{\alpha^{-1}(1) \cup T}.$$

2.4 A comparison

Let $SA^{(t)}(\mathcal{P})$ denote the feasible set of the linear program obtained by starting from a basic linear relaxation $\mathcal{P}$ (for some 0/1 program) and augmenting variables for t levels of the Sherali-Adams hierarchy. Similarly, let $LS^{(t)}(\mathcal{P})$, $LS_+^{(t)}(\mathcal{P})$, $Las^{(t)}(\mathcal{P})$ represent feasible sets corresponding respectively to t levels of the LS, LS+ and Lasserre hierarchies. We summarize in the facts below, a comparison of these relaxations. The reader is referred to the excellent survey by Laurent [43] for a more detailed comparison.

1. $LS^{(n)}(\mathcal{P}) = LS_+^{(n)}(\mathcal{P}) = SA^{(n)}(\mathcal{P}) = Las^{(n)}(\mathcal{P}) = \mathcal{I}$, where $\mathcal{I}$ denotes the convex hull of the 0/1 solutions to the starting integer program with n variables.
2. For all $t \leq n$, $LS^{(t)}(\mathcal{P}) \subseteq LS_+^{(t)}(\mathcal{P}) \subseteq Las^{(t)}(\mathcal{P})$, and also $LS^{(t)}(\mathcal{P}) \subseteq SA^{(t)}(\mathcal{P}) \subseteq Las^{(t)}(\mathcal{P})$. Hence, the relaxations provided by the Lasserre hierarchy at each level are the strongest (most constrained) among the relaxations at the corresponding level of all the hierarchies discussed above.
3. If the starting relaxation P has $n^{O(1)}$ constraints, then one can optimize over the sets $LS^{(t)}(\mathcal{P}), LS_+^{(t)}(\mathcal{P})$, $SA^{(t)}(\mathcal{P})$ and $Las^{(t)}(\mathcal{P})$ in time $n^{O(t)}$. This is known to be true for $LS^{(t)}(\mathcal{P}), LS_+^{(t)}(\mathcal{P})$ and $SA^{(t)}(\mathcal{P})$ even if we only assume that $\mathcal{P}$ has a weak separation oracle running in time $n^{O(1)}$. It is not known if one can optimize efficiently over $Las^{(t)}(\mathcal{P})$ using an efficient separation oracle for $\mathcal{P}$.

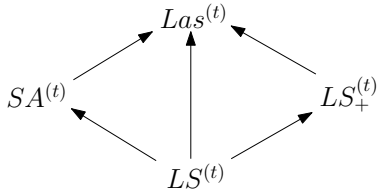

Fig. 5. A comparison (the direction of the arrows denotes tighter relaxations)

3 Approximation Algorithms

The use of LP relaxations in approximation algorithms, as described in Section 1, is a well-established approach which has spawned much work and a large variety of techniques. On the other hand, the use of SDPs in approximation algorithms is a more recent development, starting with the seminal work of Goemans and Williamson on MAX-CUT [28]. This algorithm gave a better approximation than is achievable by known LP approaches (see Section 4), and placed MAX-CUT in what is now a large

body of NP-hard problems for which a straightforward SDP-based algorithm gives an approximation which is conjectured to give the optimal approximation guarantee (see Section 5).

The main idea in the work of Goemans and Williamson [28] is to use the vectors arising from the Cholesky decomposition (defined in Section 2) of the PSD matrix of variables, and then to apply a randomized rounding technique now known as *hyperplane rounding*, which we will describe shortly. This approach (along with some refinements) quickly gave rise to improved approximation algorithms for a host of problems over the next decade.

A new approach, which included hyperplane rounding, but also introduced a subtler investigation of the geometry of finite metric spaces arising from feasible solutions to certain SDPs, was given in the celebrated work of Arora, Rao and Vazirani [7] on Sparsest Cut. Among other things, this work showed that tightening SDPs by adding certain valid constraints (in this case, the triangle inequality for squared-distances) could yield improvements by significantly reducing the integrality gap of certain relaxations.

More generally, for any given NP-hard optimization problem, one may hope to gain improvements in the approximation guarantee by systematically strengthening an LP or SDP relaxation with additional constraints, as long as the strengthened relaxation can be solved in polynomial time. This sort of systematic strengthening is precisely what is offered by the various LP and SDP hierarchies described in Section 2, all of which produce polynomial time solvable convex relaxations in the first $O(1)$ levels (in retrospect, such relaxations subsume certain important SDP relaxations, such as the Sparsest Cut relaxation in Section 3.2). Unfortunately, most results along these lines have been negative, showing that even relaxations at superconstant (and sometimes even linear) levels of certain hierarchies hardly yield any reduction in the integrality gap. These are discussed in Section 4.

In contrast, relatively few results, where improved approximation guarantees arise from the first $O(1)$ levels of a hierarchy, have recently begun to emerge. Such positive results have been obtained, for example, for MAX-CUT in dense graphs [21], for Minimum Vertex Cover in planar graphs [48], and for MaxMin Allocation [8]. Later in this section, we will examine three such results. The first two, for Chromatic Number [18] and Hypergraph Independent Set [19] make explicit use of the interpretation of feasible solutions to relaxations arising from the Sherali-Adams (and by extension, Lasserre) hierarchy as families of distributions over local 0/1-assignments. Finally, we will consider the work of Karlin, Nguyen and Mathieu on Knapsack [34], which takes quite a different approach.

3.1 Max Cut and hyperplane rounding

The approximation algorithm of Goemans and Williamson [28] for MAX-CUT is perhaps the best known example of a simple SDP-based approximation algorithm which gives an approximation guarantee for the problem it approximates which is unmatched by any other method. Their algorithm gives an approximation ratio of

1.138..., whereas until recently, all other known methods (including LP-based approaches – see Section 4) gave a 2-approximation in the worst case[5]. They propose the following SDP relaxation for the corresponding -1/1 program:

$$\begin{aligned} \text{maximize} \quad & \tfrac{1}{4} \sum_{(i,j)\in E} \|v_i - v_j\|^2 \\ \text{subject to} \quad & \|v_i\|^2 = 1 \qquad \forall\, i \in V \end{aligned}$$

The rounding algorithm and analysis are equally simple and straightforward. Sample a vector $z \in \mathbb{R}^n$ uniformly at random from the unit sphere $\{u \in \mathbb{R}^n \mid \|u\| = 1\}$, and output cut $(S, \overline{S})$, where $S = \{i \in V \mid \langle z, v_i \rangle > 0\}$. That is, separate the vectors on either side of the hyperplane orthogonal to z, and take the corresponding cut in G.

It is not hard to see that the probability that any two vectors are separated is proportional to the angle between them (where for the maximal angle, π, the vectors are separated with probability 1). For any edge $(i, j) \in E$, let $\theta_{ij} = \frac{1}{4}\|v_i - v_j\|^2$ be its contribution to the objective function. Then the probability that this edge will be cut is at least

$$\frac{\arccos\langle v_i, v_j\rangle}{\pi} = \frac{\arccos(1 - 2\theta_{ij})}{\pi} \geq \theta_{ij} \cdot \min_{0<\theta\leq 1} \frac{\arccos(1-2\theta)}{\theta} = C_{\mathrm{GW}} \cdot \theta_{ij},$$

where $C_{\mathrm{GW}} = 0.878\ldots$ is the Goemans-Williamson constant. It follows immediately, by linearity of expectation, that the expected size of the cut is at least a factor C_{GW} times the value of the objective function of the SDP, thus giving a $1/C_{\mathrm{GW}} \approx 1.138$ approximation.

Surprisingly, there is now some evidence that the Goemans-Williamson constant C_{GW} is not simply an artifact of the above analysis, but in fact the best possible. That is, assuming a conjecture about the hardness of a certain problem known as Unique Games, achieving an approximation of $1/C_{\mathrm{GW}} - \varepsilon$ is computationally intractable for any $\varepsilon > 0$. The Unique Games problem has similar consequences for many optimization problems, discussed further in Section 5.

3.2 Sparsest Cut and metric embeddings

A common feature of the MAX-CUT algorithm above, and other SDP-based algorithms that followed, is the use of "local" SDP relaxations. In a local relaxation, every SDP constraint involves a single combinatorial item – a vertex, an edge, or a clause (in a Constraint Satisfaction Problem). In turn, the analysis of the rounding algorithm involves a local examination of the (expected) contribution of each item to the rounded solution, either taking into account the local constraints for that item, or comparing its contribution in the rounded solution to its contribution in the objective function of the SDP (as in the MAX-CUT algorithm). The bound on the approximation ratio then follows directly from linearity of expectation.

[5] More recently, 1.6281...-approximations were obtained using spectral techniques [58, 56] and combinatorially using random walks [32].

We find a departure from this approach in the work of Arora, Rao and Vazirani [7] on Sparsest Cut. This problem is defined as follows: Given a graph $G = (V, E)$, find a cut $(S, \overline{S})$ that minimizes the ratio

$$\frac{\left|E(S,\overline{S})\right|}{|S|\left|\overline{S}\right|}.$$

Their algorithm gives a $\sqrt{\log n}$-approximation, which relies on the following SDP relaxation:

minimize	$\sum_{\{i,j\}\in E} \|v_i - v_j\|^2$		(1)
subject to	$\|v_i - v_j\|^2 + \|v_j - v_k\|^2 \geq \|v_i - v_k\|^2$	$\forall i, j, k$	(2)
	$\sum_{i<j} \|v_i - v_j\|^2 = 1$		(3)

Fig. 6. Sparsest Cut SDP relaxation

This relaxation is derived from the integer programming formulation as follows: in a -1/1 solution, the objective function (1) and the left hand side of (3) would represent $4\left|E(S,\overline{S})\right|$ and $4|S|\left|\overline{S}\right|$, respectively. We then scale the solution so that (1) represents the ratio of the two expressions above. Note the use of the triangle inequality for squared-distances (2), which is already a slightly non-local constraint, as it involves more than simply the variables associated with a single vertex or edge. In fact, the triangle inequality is necessary, since otherwise the integrality gap may be as large as $\Omega(n)$. To see this, consider the case of the undirected cycle C_n with vertices $\{0, \ldots, n-1\}$ and edges $\{(i, j) \mid j = i + 1 \mod n\}$. It is easy to see that the sparsest cut has sparsity $8/(n-2)^2$. On the other hand, without triangle inequality, the above relaxation has the following 2-dimensional solution: $v_j = \frac{1}{n}(\cos(2j\pi/n), \sin(2j\pi/n))$, for which the objective function has value $(4 - o(1))\pi^2/n^3$.

The triangle-inequality constraint (2) means that the vectors $\{v_i\}$ form a finite metric space, known as L_2^2, or *negative type* metric. A crucial component of the algorithm and analysis is the following structure theorem for certain L_2^2 metrics (slightly simplified here):

Theorem 1. *Let $\{v_i \mid i \in [n]\}$ be a set of vectors in $\mathbb{R}^n$ of constant length (i.e. which all belong to some annulus $\{v \in \mathbb{R}^n \mid 0 < c_1 < \|v\| < c_2\}$) which form an L_2^2 metric, and which satisfy*

$$\sum_{i<j} \|v_i - v_j\|^2 \geq n^2/3. \tag{4}$$

Then there is a polynomial-time randomized algorithm which outputs two disjoint subsets $S, T \subset [n]$ both of size $\Omega(n)$, such that for every $i \in S$ and $j \in T$ we have

$$\|v_i - v_j\|^2 \geq 1/\sqrt{\log n}. \tag{5}$$

Consider the following algorithm, which takes as a parameter $\delta > 0$, the intended L_2^2 distance between S and T:

1. Let $z = (z_1, \ldots z_n)$ where the coordinates are independent Gaussians $z_i \sim \mathcal{N}(0, 1)$.
2. For some constant $c > 0$, let $S' = \{i \mid \langle z, v_i \rangle \geq c/2\}$ and $T' = \left\{j \middle| \langle z, v_j \rangle \leq -c/2\right\}$.
3. While there is any pair $(i, j) \in S' \times T'$ such that $\|v_i - v_j\|^2 < c^2\delta$, remove i and j from S', T'.
4. Output the remaining sets $S = S'$ and $T = T'$.

Fig. 7. Algorithm ARV-Round(δ)

Note that the distribution of vector z is rotationally invariant. Thus for any fixed unit vector $v \in \mathbb{R}^n$, we have $\langle z, v \rangle \sim \mathcal{N}(0, 1)$. Using this fact, the constant density of the standard normal distribution near 0, assumption (4), and Markov's inequality, an easy argument shows that for an appropriate constant $c > 0$, Step 2 yields sets S' and T' of size $\Omega(n)$ with constant probability. Thus, it suffices to show that in Step 3 only $o(n)$ pairs are removed in expectation for $\delta = \Theta(1/\sqrt{\log n})$.

Let us first see that a "local" analysis, based on linearity of expectation, works for $\delta = O(1/\log n)$. By the properties of Gaussian distribution, we have the following tail bound on the projection of any fixed vector $v \in \mathbb{R}^n$, for $C \geq \|v\|$:

$$\mathbb{P}\left[\langle z, v \rangle \geq C\right] \leq e^{-\frac{1}{2}C^2/\|v\|^2} \tag{6}$$

Note that every pair $(i, j) \in S' \times T'$ satisfies $\langle z, v_i - v_j \rangle \geq c$. By (6) the probability that this occurs for any pair that is sufficiently close to be eliminated in Step 3 is at most $e^{-1/(2\delta)}$, which is at most $1/n^3$ for some $\delta = O(1/\log n)$. Thus, by linearity of expectation, the expected number of pairs eliminated in this case is $o(1)$.

To show that the algorithm works even for $\delta = \Theta(1/\sqrt{\log n})$ requires a much more subtle argument. We give an overview of the simplified proof in [45], with some technical details omitted. We will need the following notation. For $i \in [n]$ let us denote $\Gamma(i) = \left\{j \middle| \|v_j - v_i\|^2 < c^2\delta\right\}$, for a set $I \subseteq n$ denote $\Gamma(I) = \bigcup_{i \in I} \Gamma(i)$, and let $\Gamma^t(i)$ be $\Gamma(\ldots(\Gamma(i))\ldots)$, applied t times. Finally, denote by $R^t(i)$ the event "$\exists j \in \Gamma^t(i)$ s.t. $\langle z, v_j - v_i \rangle \geq ct$". Note that i is eliminated in Step 3 only when the event $R^1(i)$ occurs. At a high level, the analysis relies on the following idea introduced in [7], known as a *chaining argument*: By way of contradiction, assume that every $i \in [n]$ is eliminated at step 3 with probability $\Omega(1)$ (this condition can be shown to be essentially equivalent to algorithm ARV-Round failing). That is, for every $i \in [n]$, event $R^1(i)$ occurs with constant probability. Suppose we have already shown that $R^{t-1}(j)$ occurs with constant probability for all j. Then for any $i \in [n]$, with constant probability there is some $j \in \Gamma(i)$ because of which $R^1(i)$ occurs, and with constant probability $R^{t-1}(j)$ occurs *for this particular* j. Together, for some $k \in \Gamma^{t-1}(j) \subseteq \Gamma^t(i)$ the two events together imply

$$\langle z, v_k - v_i \rangle = \langle z, v_k - v_j \rangle + \langle z, v_j - v_i \rangle \geq c(t-1) + c = ct,$$

or in other words, the event $R^t(i)$.

To summarize, the chaining argument, applied for t steps, says that if algorithm ARV-Round fails, then for every vertex $i \in [n]$, event $R^t(i)$ occurs with constant probability. This has the following immediate implication:

Lemma 2. *If the chaining argument works after t steps, then ARV-Round works for* $\delta = \Omega(t/\log n)$.

Proof. By way of contradiction, assume that ARV-Round fails for this δ. Then applying the chaining argument for t steps, we get, for all $i \in [n]$,

$$\mathbb{P}\left[\exists j \in \Gamma^t(i) : \langle z, v_j - v_i\rangle \geq ct\right] = \Omega(1). \tag{7}$$

We now make crucial use of the L_2^2 property (2): it implies that for all $j \in \Gamma^t$ we have $\|v_j - v_i\|^2 < c^2\delta t$. Thus, bound (6) and a union bound imply

$$\mathbb{P}\left[\exists j \in \Gamma^t(i) : \langle z, v_j - v_i\rangle \geq ct\right] \leq ne^{-t/(2\delta)},$$

which contradicts (7).

For how many steps can we continue the above chaining argument? The main obstacle to making the argument rigorous is that, even though the events $R^t(j)$ for the various $j \in \Gamma(i)$ have constant probability, we cannot chain them to the event $R^1(i)$ since there is no guarantee that any of them will intersect the event $R^1(i)$. Instead, we have to settle for slightly weaker events $\tilde{R}^t(j)$ which have probability close to 1. Let $\tilde{R}^t(j)$ be the event "$\exists k \in \Gamma^t(j)$ s.t. $\langle z, v_k - v_j\rangle \geq c(t+1)/2$". By considerations of measure concentration, it can be shown that if $\tilde{R}^t(j)$ occurs with constant probability, then

$$\mathbb{P}\left[\exists k \in \Gamma^t(j) : \langle z, v_k - v_j\rangle \geq ct/2\right] \geq 1 - e^{-1/(9t\delta)},$$

as long as $t\delta \leq c'$ for some constant $c' > 0$. If these events have probability close to 1, we can then chain them with the event $R^1(i)$ to imply the event $R^{t+1}(i)$. For the above probability to be close to 1, it suffices to require that $t\delta \leq c''$ for some constant $c'' > 0$. Thus the chaining argument works for $t = \Omega(1/\delta)$ steps, which by Lemma 2 means that ARV-Round works for $\delta = \Omega(1/\sqrt{\log n})$.

The Sparsest Cut problem has a generalization known as the *general demand* version of Sparsest Cut. In this variant, every pair of nodes $i, j \in V$ has some edge weight $w_{ij} \geq 0$ ($w_{ij} = 0$ for non-edges) and some demand $d_{ij} \geq 0$, and the objective is to find a cut $(S, \overline{S})$ minimizing the ratio $\sum_{i\in S, j\in\overline{S}} w_{ij} / \sum_{i\in S, j\in\overline{S}} d_{ij}$. For any graph G, the worst case integrality gap of the standard LP relaxation for this problem on G (over all choices of edge weights and demands) is known to be equivalent to the L_1-embeddability of that graph (the least distortion with which a shortest-path metric on G can be embedded into L_1, maximized over all possible edge lengths). One might also ask whether the SDP relaxation also has connections to embeddings into L_1. Indeed, in later work, Arora, Lee and Naor [3] combined Theorem 1 with a careful accounting scheme, to show that any n-point L_2^2 metric embeds into L_2 (and hence into L_1) with distortion $O(\sqrt{\log n}\log\log n)$.

3.3 Chromatic Number: coloring 3-colorable graphs

The 3-Coloring problem is a classic NP-complete problem. Its optimization variant, Chromatic Number in NP-hard to approximate within an $n^{1-\varepsilon}$-factor for any constant $\varepsilon > 0$ [23, 60]. Therefore, there has been much focus on approximation algorithms when the chromatic number is small. Specifically, we are interested in the following question: given a 3-colorable graph (where the coloring itself is not given), what is the least number of colors with which we can efficiently legally color the graph?

All algorithms for this problem (or subroutines thereof) involve some assumption on either the minimum or maximum degree in the graph. For simplicity of presentation, let us only consider d-regular graphs here. After a series of purely combinatorial algorithms, Karger, Motwani and Sudan [33] gave an $\tilde{O}\left(d^{1/3}\right)$-coloring for such graphs ($\tilde{O}(\cdot)$ hides polylogarithmic factors) in one of the early SDP-based algorithms following [28]. Subsequently and Blum and Karger [11] combined this with an earlier combinatorial $\tilde{O}\left((n/d)^{3/5}\right)$-coloring algorithm of Blum [12] to give a $\tilde{O}(n^{3/14})$-coloring (or roughly $n^{0.2143}$).

This was the state of the art until a series of two papers, by Arora, Charikar and Chlamtac [6], and by Chlamtac [18] improved the above guarantee to roughly $n^{0.2111}$ and $n^{0.2072}$, respectively, by carefully characterizing the tight case of the analysis of the SDP rounding in [33] and showing it cannot occur (at least near the Blum-Karger threshold of $d = n^{9/14}$). Whereas Arora et al. [6] achieve this by giving a chaining and measure-concentration based argument reminiscent of [7], Chlamtac [18] makes explicit use of a level-2 Lasserre relaxation and its implicit interpretation as a family of local-distributions (as discussed earlier). Let us first examine the various SDP relaxations used in these works, before giving a brief high-level description of the intuition behind the Lasserre hierarchy-based improvement in [18].

Karger, Motwani and Sudan [33] proposed the following SDP relaxation for Chromatic Number of a graph $G = (V, E)$, which they called the *vector chromatic number*:

$$\begin{aligned} \text{minimize} \quad & \kappa && && (8)\\ \text{subject to} \quad & \|v_i\|^2 = 1 && \forall i \in V && (9)\\ & \langle v_i, v_j \rangle \leq -\tfrac{1}{\kappa - 1} && \forall (i,j) \in E && (10) \end{aligned}$$

This relaxation is based on the following observation: For any $k \in \mathbb{N}$, there is a set of k unit-vectors whose pairwise inner-products are all $-1/(k-1)$, and moreover this is the minimum value for which such vectors exist. The vector chromatic number is equivalent to $\vartheta'(\overline{G})$, where $\overline{G}$ is the complement graph, and ϑ' is the variant of the Lovász ϑ-function introduced by Schrijver [54][6].

Let us focus on relaxations for 3-Coloring. Since the above relaxation is not based on a 0/1 program, we require some manipulation in order to strengthen it using an SDP hierarchy based on 0/1 programs, such as the Lasserre hierarchy. Such

[6] The Lovász ϑ function itself is actually equivalent to a variant of the above relaxation, called the *strict vector chromatic number*, in which we have equality in (10).

an approach has been suggested by Gvozdenović and Laurent [29], based on the following observation of Chvátal [20]: Let X be a set of three colors $X = \{R, B, Y\}$. For a graph $G = (V, E)$, define a new graph $G_X = (V_X, E_X)$, where $V_X = V \times X$, and $E_X = \{((i, C), (j, C)) \mid (i, j) \in E \text{ and } C \in X\} \cup \{((i, C_1), (i, C_2)) \mid i \in V \text{ and } C_1 \neq C_2 \in X\}$. Then G is 3-colorable iff G_X contains an independent set of size n (note that it can never contain a larger independent set). Moreover, every independent set I corresponds to a unique 3-coloring $f_I : V \to X$ in the natural way: $f_I(i) = C$ for all $(i, C) \in I$. We can now apply any of the SDP hierarchies described earlier for Maximum Independent Set to the graph G_X.

How does such a hierarchy give improvements over the algorithm and analysis in [33, 11]? It suffices to consider the tight (or nearly tight) case of the SDP rounding in [33], and show that when the analysis is tight, there is an alternative rounding which performs vastly better than all currently known algorithms. To understand the tight case of the analysis, it helps to examine the interpretation of the SDP solution as a family of distributions on local colorings (as arise in Sherali-Adams and Lasserre relaxations, e.g. as in Lemma 1). In particular, let us fix a single vertex i, and some coloring, say (i, R), and consider the random color assignments (over $\{B, Y\}$) to its various neighbors $N(i)$. A simple probabilistic argument says that it cannot be the case that most pairs of vertices $j, k \in N(i)$ are assigned the same color with probability much less than $\frac{1}{2}$ (since the average correlation cannot be very negative). On the other hand, it turns out that in the tight case, these probabilities cannot be much larger on average. Indeed, if many such pairs are assigned the same color with significantly higher probability, then the corresponding vectors u_j, u_k will have a larger inner product. At a very high level, this facilitates a better SDP rounding, since when the SDP solution is more clustered, it becomes easier to separate vertices into fewer color classes. To summarize, if the SDP analysis of [33] is tight, then in the interpretation of the SDP solution locally as a distribution on colorings, most vertices at distance two will be assigned the same color with probability $\approx \frac{1}{2}$.

When can a distribution on 3-colorings with this property exist? Let us fix two vertices j, k at distance 2 from each other which are assigned the same color with probability $\frac{1}{2}$, which also have many common neighbors $i \in N(j) \cap N(k)$. We know that when j and k are assigned different colors, say B and Y, then all vertices in $N(j) \cap N(k)$ will have *the same* color as each other (here R), while if j and k are assigned the same color, then as before, most pairs of vertices $i, i' \in N(j) \cap N(k)$ will have the same color with probability $\approx \frac{1}{2}$. Thus, most pairs of common neighbors i, i' will receive the same color with probability $> \frac{1}{2}$, contradicting our earlier assertion for vertices at distance 2. The only resolution of this contradiction is for all common neighborhoods $N(j) \cap N(k)$ to all be quite small. A simple counting argument then gives a lower bound on the size of any 2-neighborhood $N(N(j))$.

The precise formalization of the above argument gives a bound of $|N(N(j))| \geq d^{3/2}$. This is already a contradiction when $d > n^{2/3}$ (thus ruling out any tight case of the previous analysis for this range of parameters). However, when the degree d is near the Blum-Karger threshold of $n^{9/14}$, this is not quite enough. In this case we make use of the fact that each vertex in $N(N(j))$ has the same color as j with probability $\frac{1}{2}$, which implies an independent set of size $\frac{1}{2}|N(N(j))|$ in the 2-neighborhood.

Standard Minimum Vertex Cover approximations then allow us to extract a large independent set (a color class in our coloring), here of size $\tilde{\Omega}(d^{3/2}) = \tilde{\Omega}(n^{27/28})$, which makes progress towards an $\tilde{O}(n^{1/28})$ coloring. To summarize, this argument shows that either the analysis of the SDP rounding in [33] is not tight (and then the algorithm performs better), or a different algorithm yields a far better coloring than the current approach. Formalizing the argument involves a careful analysis of the SDP rounding in [33], and is beyond the scope of this chapter.

3.4 Mixed hierarchies and Hypergraph Independent Set

In all SDP-based approximation algorithms, such as the MAX-CUT and Sparsest Cut algorithms we've seen, the rounding algorithm and analysis rely on the geometry of SDP solutions. More recently, algorithms such as the coloring algorithm of [18] also use the interpretation of solutions to Lasserre hierarchies as families of distributions. However, this interpretation also arises in LP hierarchies such as Sherali-Adams. We might ask whether weaker SDP hierarchies than Lasserre, which combine an LP characterized by local distributions with a simple SDP could also yield improved approximations. Such mixed hierarchies arise naturally in approximation algorithms. In fact, under certain complexity theoretic assumptions, they already give optimal approximations for k-CSPs at level k (see Section 5).

One example where such a hierarchy gives an infinite sequence of improved approximation guarantees is described in the work of Chlamtac and Singh [19] on Hypergraph Independent Set. Let us focus on the 3-uniform variant of this problem: We are given a hypergraph $H = (V, E)$, where the hyperedges in E are all subsets $e \subset V$ of cardinality $|e| = 3$. Find the maximum independent set in H, where an independent set is any subset $S \subseteq V$ that does not contain any hyperedge as a subset.

Let us now define a mixed hierarchy of relaxations for this problem. First, for any integer $t \geq 1$, define

$$I_t = \mathrm{conv}\left(\left\{ (\textstyle\prod_{i \in S} x_i)_{S \subseteq [t]} \middle| \, x \in \{0,1\}^t \right\}\right).$$

Note that there is a one-to-one correspondence between vectors in I_t and distributions over assignments $f : [t] \to \{0, 1\}$. Now, define $M_1(Y)$ for any vector (Y_S) to be the $(n+1)\times(n+1)$ moment matrix $(Y_{S \cup T})_{|S|,|T| \leq 1}$. For $t \geq 3$, Figure 8 (below) gives a level-t relaxation for Hypergraph Independent Set. Note that the constraint $M_1(Y) \succeq 0$ simply means that there exists a set of $n + 1$ vectors $v_\emptyset, v_{\{1\}}, \ldots v_{\{n\}}$ whose pairwise inner-products are consistent with the LP values, i.e. $\langle v_S, v_T \rangle = Y_{S \cup T}$.

Like Chromatic Number, the problems Maximum Independent Set and Hypergraph Independent Set are notoriously hard to approximate. Therefore, for polynomial-size relaxations, we cannot expect integrality gaps smaller than $n^{1-o(1)}$ (unless P = NP). Instead, we focus on integrality gaps parametrized by the optimum value of the relaxation. As before, denote by FRAC and OPT the optimum of the LP or SDP relaxation, and the 0/1 optimum, respectively. Then for $\gamma \in [0, 1]$ we can define the approximation curve of the relaxation to be

$$\mathsf{MINOPT}_n(\gamma) = \min\{\mathsf{OPT} \mid \mathsf{FRAC} \geq \gamma n\},$$

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{i=1}^{n} Y_{\{i\}} & \\
\text{subject to} & Y_{\{i,j,k\}} = 0 & (i,j,k) \in E \\
& (Y_S)_{S \subseteq T} \in I_t & |T| = t \\
& M_1(Y) \succeq 0 &
\end{array}
$$

Fig. 8. Mixed level-t relaxation for Hypergraph Independent Set

where the minimization is over all problem instances on n variables. Thus, the approximation curve can be meaningful even when the integrality gap (in this case, $\max_\gamma(\gamma n/\mathsf{MINOPT}(\gamma))$ is $\Omega(n)$.

For example, a series of papers [2, 30] shows that for Maximum Independent Set, the SDP relaxation given by the Lovász ϑ function has an approximation curve of $\mathsf{MINOPT}_n(\gamma) = \tilde{\Omega}(n^{f(\gamma)})$, where f satisfies $f(\gamma) \geq 3\gamma/(1+\gamma)$ for $1/\gamma \in \mathbb{N} \setminus \{1\}$, and $f(\gamma) \geq \min\{2\gamma, 1\}$ in general. Attempts to extend this approach to Hypergraph Independent Set by Krivelevich, Nathaniel and Sudakov [41] yielded only partial results. In particular, for all $\gamma \leq \frac{1}{2}$ the approximation curve given by their SDP relaxation for this problem is quite poor: $\mathsf{MINOPT}_n(\gamma) = 2$. This shortcoming is in fact inherent. The SDP relaxation arising from level-3 of the mixed hierarchy in Figure 8, which is at least as strong as the SDP used by [41], always has optimum value at least $n/2$.

While no fixed level of the mixed hierarchy gives non-trivial guarantees for all $\gamma > 0$, the range of γ for which the approximation curve is non-trivial grows as we use higher levels, and tends to $[0, 1]$ in the limit. Let us denote by $\mathsf{MINOPT}_{n,t}$ the value of MINOPT_n as defined above with respect to the level-t mixed hierarchy in Figure 8. Then the results of [19] for this hierarchy can be summarized as follows:

$$A_{n,t}(\gamma) \leq \mathsf{MINOPT}_{n,t}(\gamma) \leq B_{n,t}(\gamma),$$

where

$$A_{n,t}(\gamma) = \begin{cases} 2, & \gamma < \frac{2}{\sqrt{t}} \\ n^{\gamma^2/32}, & \gamma \geq \frac{2}{\sqrt{t}} \end{cases} \qquad B_{n,t}(\gamma) = \begin{cases} 2, \gamma \leq \frac{1}{t-1} \\ n, \gamma > \frac{1}{t-1}. \end{cases}$$

Note that the second case in bound $A_{n,t}(\gamma)$ together with the first case in bound $B_{n,t}(\gamma)$ shows that indeed, there is an infinite sequence of strict improvements in the approximation guarantee as t increases (for instance, $\mathsf{MINOPT}_{n,9t^2}(\gamma)$ is much larger than $\mathsf{MINOPT}_{n,t+1}(\gamma)$ in the range $2/(3t) \leq \gamma \leq 1/t$). While slightly better parameters for $A_{n,t}(\gamma)$ are known for the Lasserre hierarchy, there are currently no bounds such as $B_{n,t}(\gamma)$ for this hierarchy (in particular, no bounds which would preclude the possibility that even the level-2 Lasserre relaxation gives a non-trivial guarantee for every $\gamma > 0$).

3.5 Sherali-Adams and Lasserre relaxations for Knapsack

The Knapsack problem is defined as follows: Given a set of n items with costs $c_i \geq 0$ and rewards $r_i \geq 0$, and some capacity $C > 0$, find a subset $S \subseteq [n]$ of cost $\sum_{i \in S} c_i$ at most C which maximizes the reward $\sum_{i \in S} r_i$. This is a well-understood classic NP-complete problem which is easy to approximate – it admits an FPTAS [31, 44]. While from the perspective of approximation algorithms, there is nothing to be gained by applying convex optimization techniques to this problem, it is a useful tool for gaining a better understanding of the strengths and properties of various hierarchies of relaxations. Let us review the results of Karlin, Mathieu and Nguyen [34], who recently investigated this topic.

Consider the natural LP relaxation for Knapsack:

$$\text{maximize} \quad \sum_i r_i y_i \tag{11}$$

$$\text{subject to} \quad \sum_i c_i y_i \leq C \tag{12}$$

$$0 \leq y_i \leq 1 \qquad \forall i \in [n] \tag{13}$$

Fig. 9. Standard LP relaxation for Knapsack

This LP relaxation has an integrality gap of 2. Due to the existence of simple combinatorial $(1+\varepsilon)$-approximations for Knapsack, one would expect that strengthening this relaxation would quickly reduce the integrality gap. Nevertheless, the authors of [34] show that this is not the case for Sherali-Adams:

Theorem 2. *For any $0 < \delta < \frac{1}{3}$, the integrality gap of the level-δn Sherali-Adams relaxation for Knapsack is at least $2/(1+\delta)$.*

This follows by considering a simple instance of Knapsack: let all the costs and rewards be 1, and the capacity be $C = 2 - \varepsilon$ for some sufficiently small $\varepsilon > 0$. The optimum in this instance has reward 1, while the LP admits a solution of value $(2-\varepsilon)/(1+(1-\varepsilon)\delta)$ (the lemma follows by letting ε tend to 0). To see this, consider the Sherali-Adams solution $Y_\emptyset = 1$, $Y_{\{i\}} = p$ for all $i \in [n]$, where $p = (2-\varepsilon)/((1+(1-\varepsilon)\delta)n)$, and $Y_I = 0$ for all $|I| > 1$. This clearly gives valid distributions over assignments to sets of size $1/p > \delta n$. It can be checked that for this solution, the only relevant constraints that need to be verified are certain lifts of constraint (12). Specifically, under the interpretation of Sherali-Adams as a relaxation for a global distribution over knapsack solutions $K \subseteq [n]$, for all $J \subseteq [n]$ of cardinality $|J| \leq \delta n$, we need to verify the constraint corresponding to $\mathbb{E}[|K| \mid J \cap K = \emptyset] \leq C$. But for the above LP solution, the expression corresponding to this expectation is simply $(n - |J|)p/(1 - |J|p) \leq 2 - \varepsilon$ (assuming $|J| \leq \delta n$).

Why is the integrality gap so large after so many rounds? The main reason is that Sherali-Adams is very inconsistent with respect to the value of the objective function. Specifically, in the above example, the objective function takes value $\sum_i Y_{\{i\}}$

which is nearly 2. However, if we lift this expression by one round by "conditioning on $x_j = 1$" for a fixed $j \in [n]$, we get $(Y_{\{j\}} + \sum_{i\in[n],i\neq j} Y_{\{i,j\}})/Y_{\{j\}} = 1$. We can circumvent this problem by rephrasing the initial LP as a feasibility LP, with the objective function as an added constraint, as in Figure 10. Note that R in constraint (14) is an external parameter, and not an LP variable. We can now take this relaxation and apply Sherali-Adams to it, and find the maximum R that for which the new Sherali-Adams relaxation is feasible (say, by binary search).

$$\begin{aligned}
&\text{find} && y_1, \ldots, y_n && \\
&\text{which satisfy} && \sum_i r_i y_i \geq R && (14)\\
& && \sum_i c_i y_i \leq C && (15)\\
& && 0 \leq y_i \leq 1 \qquad \forall i \in [n] && (16)
\end{aligned}$$

Fig. 10. Parametrized feasibility LP relaxation for Knapsack

Applying Sherali-Adams to the feasibility LP above significantly reduces the integrality gap:

Theorem 3. *The integrality gap of level-t of the Sherali-Adams hiersarchy applied to the feasibility* LP *in Figure 10 is at most* $1 + 1/(t-2)$.

To see why lifting the objective function helps, we first need the following well-known lemma (here, Greedy is the value of the combinatorial greedy algorithm for Knapsack, which does not use any convex relaxations of the problem):

Lemma 3. *The standard* LP *relaxation satisfies* $\mathsf{FRAC} \leq \mathsf{Greedy} + \max_i r_i (\leq 2\mathsf{OPT})$.

The bound in Theorem 3 now follows from a simple rounding algorithm: Let $S_{t-1} = \{i \mid r_i > \mathsf{OPT}/(t-1)\}$ (we can guess this set in n trials by sorting). At most $(t-2)$ items in S_{t-1} can fit in the knapsack (otherwise OPT would be higher). So, as long as there exists an item $i \in S_{t-1}$ with non-zero LP value, condition the LP solution on picking this item. After at most $(t-1)$ steps, all the remaining items in S_{t-1} have LP value 0. Let K_0 be the set of items picked so far, with reward $R_0 = r(K_0)$, and consider the current LP solution $\{Y'\}$ restricted to $[n] \setminus S_{t-1}$. Since we lifted the objective function, the value of this LP is now $\sum_{i\notin S_{t-1}} Y'_i \geq R - R_0$. Now apply the greedy algorithm to the remaining items, giving some additional reward R_g. By the above bound on the LP value, and Lemma 3, we have

$$R - R_0 \leq \sum_{i\notin S_{t-1}} Y'_i \leq R_g + \max_{i\notin S_{t-1}} r_i < R_g + \mathsf{OPT}/(t-1) \leq R_g + R/(t-1).$$

Therefore, the value of the rounded solution is $R_0 + R_g \geq R - R/(t-1) = R(1-1/(t-1))$.

Surprisingly, the Lasserre hierarchy does not require such manipulations. Applying it directly to the standard (maximization) LP immediately yields essentially the same guarantee as in Theorem 3:

Theorem 4. *Level-t of the Lasserre hierarchy applied to the standard relaxation for* Knapsack *has integrality gap* $\leq 1 + \frac{1}{t-1}$.

This follows from a similar rounding algorithm as above. We need to adapt the rounding of the fractional solution for the high-reward items S_t to the current setting, after which we can apply the greedy algorithm to the remaining items, just as before. Note that for the same analysis to go through, the rounding for S_t should be lossless, in the sense that it does not decrease the overall objective value.

The existence of a lossless rounding for S_t follows immediately from the following key observation: the solution to the above SDP restricted to the set S_t is already in the *integral hull*. To see this, consider any set S for which the SDP solution satisfies $\|\mathbf{U}_T\| = 0$ for all subsets $T \subseteq S$ of cardinality $|T| = t$. Then we can add zero vectors $\mathbf{U}_T = 0$ for *all* subsets $T \subseteq S$ of cardinality $|T| > t$, without violating the essential consistency constraints

$$\forall T_1, T_2, T_3, T_4 \subseteq S \text{ s.t. } T_1 \cup T_2 = T_3 \cup T_4 : \langle \mathbf{U}_{T_1}, \mathbf{U}_{T_2} \rangle = \langle \mathbf{U}_{T_3}, \mathbf{U}_{T_4} \rangle .$$

This would extend the current solution to a valid level-$|S|$ Lasserre solution. Such a solution must be in the integral hull (as would be any level-$|S|$ relaxation restricted to $|S|$ vertices for any of the hierarchies we consider here).

In our case, this defines a distribution over assignments $f : S_t \to \{0, 1\}$ and feasible LP solutions $\{y_i^f\}_i$ which are integral on S_t, such that for all $i \in [n]$ we have $\mathbb{E}_f \left[y_i^f \right] = \|\mathbf{U}_{\{i\}}\|^2$. Thus, by linearity of expectation, at least one such LP solution has objective value at least $\sum_i r_i \|\mathbf{U}_{\{i\}}\|^2$. Moreover, since there are at most $|S_t|^{t-1}$ assignments in the support of this distribution (each assignment assigns 1 to at most $t - 1$ items), we can enumerate over all such assignments (and correpsonding LP solutions) in polynomial time.

4 Lower bounds on Integrality Gaps

From the perspective of complexity theory, one can view the various hierarchies of programs as restricted models of computation, with the number of applications of these operators as a resource. This also corresponds naturally to time, as optimizing over the level-t relaxations takes time $n^{O(t)}$.

Showing that the integrality gap for a problem remains large after many levels of the hierarchy then corresponds to a strong lower bound, which *unconditionally* (not even assuming P $\neq$ NP) rules out a large and general class of algorithms. Also, for some problems where NP-hardness results are not known, such lower bounds give some evidence of hardness for the problem.

4.1 Integrality gaps for Vertex Cover

Minimum Vertex Cover, the problem of finding the smallest subset of vertices in a graph such that every edge is incident to some vertex in the subset, is perhaps the most studied problem with regard to integrality gaps. While a simple LP relaxation for the problem gives a factor 2 approximation, it can also be shown that the integrality gap after many levels of the different LP hierarchies remains at least $2 - \varepsilon$. However, among the SDP hierarchies, lower bounds close to factor 2 are known only in the LS+ hierarchy.

Viewing solutions as local distributions

The technique most useful for proving lower bounds on LP integrality gaps has been the view of LP solutions as "local distributions" as was stated for the Sherali-Adams hierarchy in Lemma 1 (the lemma was stated for Maximum Independent Set, but an identical claim also holds for Minimum Vertex Cover). The LS hierarchy of linear programs does not have such a direct characterization in terms of local distributions, but intuition of probability distributions can still be applied when reasoning about it.

To describe this view for the LS hierarchy, we re-interpret what its conditions mean for a point $\mathbf{x} \in \mathcal{P}$ which is indeed a convex combination of integer solutions. Such a point $\mathbf{x}$ is expressible as $\mathbf{x} = \sum_i \lambda_i \mathbf{z}^{(i)}$ where $\sum_i \lambda_i = 1$ and $\forall i.\ \ \mathbf{z}^{(i)} \in \mathcal{P} \cap \{0,1\}^n,\ \ \lambda_i \geq 0$. Then, we can consider a random variable $\mathbf{z}$ which takes value $\mathbf{z}^{(i)}$ with probability λ_i. For $j \in \{1, \ldots, n\}$, the numbers x_j are then equal to $\mathbb{P}\left[z_j = 1\right]$ i.e. the marginals of this distribution.

To prove that $\mathbf{x} \in N(\mathcal{P})$, we then require a matrix $Y \in \mathbb{R}^{n+1}$ which satisfies the conditions stated in Definition 1. For each $\mathbf{z}^{(i)}$ such a matrix $Y^{(i)}$ can be given as $Y^{(i)} = (1, \mathbf{z}^{(i)})(1, \mathbf{z}^{(i)})^T$ where $(1, \mathbf{z}^{(i)}) \in \{0,1\}^{n+1} \in \mathbb{R}^{n+1}$. The matrix Y for $\mathbf{x}$ can then be exhibited as $Y = \sum_i \lambda_i Y^{(i)}$, where each entry $Y_{ij} = \mathbb{P}\left[(z_i = 1) \wedge (z_j = 1)\right]$. Arguing that the vector $Y_i \in \text{cone}(\mathcal{P})$ is then equivalent to arguing that the vector $\mathbf{x}^{(i,1)} \in \mathbb{R}^n$ defined as

$$x_j^{(i,1)} \ = \ Y_{ij}/x_i \ = \ \mathbb{P}\left[z_j = 1 \mid z_i = 1\right]$$

is in $\mathcal{P}$. Similarly, $Y_0 - Y_i \in \text{cone}(\mathcal{P})$ is equivalent to proving that the vector $\mathbf{x}^{(i,0)}$ with coordinates $x_j^{(i,0)} = (Y_{0j} - Y_{ij})/(1 - x_i) \ = \ \mathbb{P}\left[z_j = 1 | z_i = 0\right]$, is in $\mathcal{P}$.

Thus, to prove that a vector $\mathbf{x}$ of marginal probabilities is in $N(\mathcal{P})$, we need to provide a vector of *conditional* probabilities, where the conditioning is on being an arbitrary variable chosen by an *adversary* being 0 or 1. For proving $\mathbf{x} \in N^t(\mathcal{P})$, we can think of t-step game, where at each step we are required to provide the conditional probabilities and the adversary can further condition on one more variable.

Integrality gaps in the LS hierarchy

The study of integrality gaps in this model was initiated by the works of Arora, Bollobás, Lovász and Tourlakis [4, 5]. They showed that the integrality gap remains at least $2 - \varepsilon$ even after $\Omega(\log n)$ levels of the hierarchy. Since the integrality gap can be easily shown to be *at most* 2 even for the starting linear relaxation, this showed that even using time $n^{O(\log n)}$ in the computational model of the LS hierarchy yields

no significant improvement. The results were later improved to $3/2$ for $\Omega(\log^2 n)$ levels by Tourlakis [57] and to an optimal lower bound of $2-\varepsilon$ for $\Omega(n)$ levels by Schoenebeck, Trevisan and Tulsiani [53]. Note that the last result even rules out exponential time algorithms (as $\Omega(n/\log n)$ levels would correspond to $2^{\Omega(n)}$ time) in this model.

To produce instances with a large integrality gap, one considers sparse random graphs which have no cycles of size less than $\Omega(\log n)$ so that any subgraph with $O(\log n)$ vertices is a tree. One then starts with a solution which has fractional value $1/2+\varepsilon$ on every vertex. The move of the adversary then corresponds to selecting a vertex where the solution has a fractional value, and fixing it to 1 or 0 i.e. conditioning it to be in or out of the vertex cover. As long as the adversary conditions on $O(\log n)$ vertices, the set of conditioned vertices form a tree, restricted to which there is an *actual distribution of vertex covers* with marginal values $1/2+\varepsilon$. Hence, one can use this to produce the required conditional distribution. We remark that this is just an intuition for the proof in [5]. The actual proof proceeds by looking at the duals of the linear programs obtained by the LS hierarchy and involves significantly more work.

The result in [53] for $\Omega(n)$ levels uses an explicit version of the above intuitive argument together with a more careful use of the sparsity of these graphs. Their techniques also give an integrality gap of $2-\varepsilon$ for relaxation of MAX-CUT obtained by $\Omega(n)$ levels of the LS hierarchy. The results for MAX-CUT also exhibit a separation between linear and semidefinite programs. As shown in Section 3.1, even a basic semidefinite program at the first level of the SDP hierarchy can be shown to have integrality gap at most $1/0.878 \approx 1.139$; while even the linear programs obtained by $\Omega(n)$ levels have integrality gap close to 2.

Integrality gaps in the Sherali-Adams hierarchy

Charikar, Makarychev and Makarychev [16] proved that for any $\varepsilon > 0$, there is a δ such that the integrality gap of the LP relaxation for Minimum Vertex Cover obtained by n^δ levels of the Sherali-Adams hierarchy is at least $2-\varepsilon$. They used an intuition similar to that in [53], where one defines a process to sample vertex covers on trees by including the root with probability $1/2+\varepsilon$; and including a child with probability 1 if the parent is excluded and ε' otherwise. The value of ε' is chosen so that the marginal probability for each vertex is $1/2+\varepsilon$. They also defined an extension of this process on graphs which are not trees, to produce local distributions for subsets of size n^δ. The number of levels is less than in [53] as the conditions imposed on local distributions by the Sherali-Adams hierarchy are stronger than those imposed by the LS hierarchy.

Using similar techniques, Charikar, Makarychev and Makarychev also showed a gap of $2-\varepsilon$ for MAX-CUT after n^δ levels. They also extended it show Sherali-Adams integrality gaps for Unique Games and many other problems to which Unique Games can be reduced (see Section 5 for more on Unique Games).

Integrality gaps for semidefinite programs

Integrality gaps for the semidefinite hierarchies have been somewhat harder to prove. It was shown by Goemans and Kleinberg [39] that the integrality gap of an SDP relaxation for Minimum Vertex Cover based on the ϑ-function of Lovász (which is weaker than the relaxation obtained by one application of N_+) is at least $2 - \varepsilon$. The result was strengthened by Charikar [14], who showed that the same gap holds even when the relaxation is augmented with a subset of the "triangle inequalities" discussed in Section 3.2 . . The gap was extended to $\Omega(\sqrt{\log n / \log \log n})$ levels of the LS_+ hierarchy by Georgiou et al. [26]. Interestingly, all the these results for LS+ were proven for the same family of graphs, inspired by a paper of Frankl and Rödl [25]. It is an interesting problem to construct an alternate family which is also an integrality gap instance, or to extend the above results to even $\Omega(\log n)$ levels.

Somewhat incomparable to the above results, lower bounds of factor $7/6$ for $\Omega(n)$ levels of the LS+ hierarchy were obtained by Schoenebeck et al. [52]. These were later strengthened to $7/6$ for $\Omega(n)$ levels of the Lasserre hierarchy by Schoenebeck [51] and 1.36 for $n^{\Omega(1)}$ levels of Lasserre by Tulsiani [59]. These results also differ from the ones discussed above in that they do not directly exhibit a family of integrality gap instances for vertex cover. Instead, they start with an integrality gap instance for a constraint satisfaction problem, and proceed by using a reduction from the constraint satisfaction problem to vertex cover.

4.2 Results for Constraint Satisfaction Problems

For constraint satisfaction problems (CSPs) with 3 or more variables in each constraint, very strong lower bounds have been shown even for the Lovász-Schrijver and Lasserre semidefinite (and hence also the linear) hierarchies. For these problems one studies how well convex relaxations approximate the maximum number of constraints that can be satisfied by any assignment to the variables. Instances exhibiting a large integrality gap for CSPs are also useful as they can often be transformed to lower bounds for other problems using reductions (see Section 4.3).

Proofs of integrality gaps for CSPs were significantly influenced by arguments in proof complexity and crucially used an *expansion* property of the problem instances. In proof complexity, expansion arguments were used for proving exponential lower bounds on the size of proofs in the *resolution* proof system to show that a certain SAT formula was unsatisfiable [7]. Specifically, they showed that φ is an unsatisfiable formula in n variables in conjunctive normal form and 3 variables in each clause, with each set of s clauses (for say $s \leq n/1000$) involving at least (say) $3s/2$ variables; then any proof of the unsatisfiability of φ in the resolution proof system must have exponential size (see [10], [9]).

For proving large integrality gaps, one is often interested in showing that such an unsatisfiable SAT formula (or instance of some other CSP) "seems highly satisfiable" to a convex relaxation. In the context of the hierarchies, expansion guarantees

[7] Resolution is the proof system where one uses two clauses of the form $(\psi_1 \vee x)$ and $(\psi_2 \vee \neg x)$ to derive $(\psi_1 \vee \psi_2)$. Unsatisfiability is proved by deriving the empty clause.

that any small set of variables is scattered across various clauses and hence the formula restricted to these variables looks highly satisfiable. This intuition is formalized differently for different hierarchies leading to the corresponding bounds.

Integrality gaps for the Lovász-Schrijver hierarchies

For the LS+ hierarchy, optimal lower bounds of factor $2^k/(2^k - 1)$ for MAX k-SAT with $k \geq 5$ and $\Omega(n)$ levels were shown by Buresh-Oppenheim et al. [13]. They were also the first to use the expansion arguments in the context of Lovász-Schrijver hierarchy. Their results were later extended to the important remaining case of MAX 3-SAT by Alekhnovich et al. [1], who also proved strong lower bounds for approximating Minimum Vertex Cover in hypergraphs.

The arguments for the above results start with a vector with a fractional value for each variable and prove that the vector is in $N_+^t(\mathcal{P})$ for $t = \Omega(n)$. As before, we think of an adversary fixing one of the fractional variables to 1 or 0 (i.e. true or false) at each of t steps, and one is required to provide a fractional assignment consistent with the fixing which is still in the polytope of feasible solutions. However, at each step, instead of proving that the solutions they provide are in the polytope, they express it as a convex combination of a set O of fractional solutions, and prove that all the solutions in O are in the polytope.

The set O is obtained by fixing *additional* variables at each step to maintain the invariant that if one considers the formula restricted only to the variables which have not been fixed to 0 or 1, then the formula is still expanding (in the sense that a set of s clauses will contain at least $3s/2$ unfixed variables). Expansion essentially means that even when $O(n)$ variables are fixed, most clauses still have a large number of unfixed variables, whose value can be modified suitably to satisfy the constraints of the convex program.

Integrality gaps in the Lasserre hierarchy

Optimal $\Omega(n)$ level lower bounds for other constraint satisfaction problems were also proved for the relaxations in the Lasserre hierarchy by Schoenebeck [51] and Tulsiani [59]. Schoenebeck proved the first integrality gaps in the Lasserre hierarchy for the MAX k-XOR problem where each constraint is a linear equation in $\mathbb{F}_2$ involving k variables. He showed an optimal integrality gap of $2 - \varepsilon$ for $\Omega(n)$ levels of the hierarchy. His techniques were extended in [59] to a large family of CSPs, also showing that for the general problem MAX k-CSP with arbitrary constraints, the integrality gap is at least $2^k/2k$ after $\Omega(n)$ levels. The latter result matches, up to a constant factor, the result of Charikar, Makarychev and Makarychev [15], who gave an SDP based approximation algorithm for MAX k-CSP achieving an approximation ratio of $O(2^k/k)$.

Schoenebeck's result was based on a significant extension of a technique for creating SDP solutions for MAX k-XOR, previously used by Feige and Ofek [24] and Schoenebeck et al. [52]. He showed that for a random instance of MAX k-XOR, the SDP relaxation has value 1 (the objective is the maximum fraction of constraints that can be satisfied) while the integer optimum can be easily shown to be at most $1/2 +$

ε. He created the SDP solutions for relaxation obtained by t levels of the Lasserre hierarchy, by creating vectors with one coordinate for each linear form in at most $2t$ variables over $\mathbb{F}_2$. The vector for a partial assignment to the variables then takes value 1 or -1 depending on the parity of the corresponding linear form according to the partial assignment e.g. the vector corresponding to the partial assignment ($x_1 = 1, x_2 = 1$) takes value $(-1)^{x_1} = -1$ in the coordinate for x_1 and $(-1)^{x_1+x_2} = 1$ in the coordinate for $x_1 + x_2$. In addition, the linear equations in the constraints (say $x_1 + x_2 + x_3 = 1$) imply certain additional constraints on the parities ($(-1)^{x_1+x_2} = -(-1)^{x_3}$ for all assignments). These were imposed by grouping linear forms into equivalence classes if they were related by an equation, and having a single coordinate for each class instead. Expansion was used to prove that it was indeed possible to partition the linear forms consistently.

The above technique was extended in [59] to handle more general constraints, but it still required them to be expressible in some way as linear equations. On the other hand, optimal Sherali-Adams integrality gaps were proved by Georgiou et al. [27] for a much more general class of constraints (which do not have such a linear structure). It remains an interesting open problem to prove the corresponding gaps in the Lasserre hierarchy.

4.3 Results for other problems

A very useful technique in proving integrality gaps for other problems has been the use of *reductions*. This was first used by Khot and Vishnoi [38] in converting integrality gaps for Unique Games to those for Sparsest Cut. For the purposes of this discussion, we may think of Unique Games as simply a CSP with each constraint being a linear equation in two variables modulo a large prime p. It has been shown to be a very convenient starting point for many reductions. Khot and Vishnoi exhibited an integrality gap for Unique Games, and using a reduction to Sparsest Cut[8] , proved that the integrality gap is at least $\Omega((\log\log n)^{1/6})$ for the SDP in figure 3.2. The bound was later improved to $\Omega(\log\log n)$ by Krauthgamer and Rabani [40].

Their results were extended by Raghavendra and Steurer [50], who showed that the integrality gap remains at least $\Omega((\log\log n)^{\delta})$ even for the programs obtained by $\Omega((\log\log n)^{\gamma})$ levels of the mixed hierarchy, for some absolute constants $\delta, \gamma > 0$. A similar result was independently obtained by Khot and Saket [37]. The lower bound for the SDP in figure 3.2 has been recently improved to $\Omega((\log n)^{\delta})$ for some small positive constant δ, by Cheeger, Kleiner and Naor [17]. Note that this still remains far from the best known upper bound of $O(\sqrt{\log n})$.

Reductions were also used by Tulsiani [59] to convert integrality gaps for CSPs in the Lasserre hierarchy, to those for Maximum Independent Set, Minimum Vertex Cover and coloring problems. The arguments there involve considering the reduc-

[8] The results in [38] were actually for a generalized version of the Sparsest Cut problem, where the denominator is not the total number of pairs $|S||\overline{S}|$ with one vertex in S, but rather each pair has a different cost associated with it. This is known as the *non-uniform* version of the problem. The result for the uniform version was proven later by Devanur et al. [22].

tions used for proving the hardness of approximating theses problems, and generalizing the proofs of correctness of the reductions to work with vector solutions, instead of integer solutions.

5 Integrality Gaps and Hardness of Approximation

For constraint satisfaction problems, a very elegant connection between the integrality gaps and the NP-hardness of approximating these problems was exhibited by Raghavendra [49]. He considered a basic semidefinite relaxation for any constraint satisfaction problem, and showed that assuming a conjecture about the hardness of approximating the Unique Games problem, it is NP-hard to achieve a better approximation ratio than the basic SDP.

Recall that Unique Games is a constraint satisfaction problem with each constraint being a linear equation in two variables, modulo a large prime p. It was conjectured by Khot [35] that the for all ε, there is a p such that it is NP-hard to distinguish instances of Unique Games (with equations modulo p) in which $1-\varepsilon$ fraction of the constraints are satisfiable from those in which at most ε fraction are satisfiable. This conjecture, known as the Unique Games Conjecture, has been used as an assumption in a large number of complexity results.

The semidefinite relaxation considered by Raghavendra is best stated in terms of the mixed hierarchy defined in Section 3.4. For an instance of MAX k-CSP, where each constraint is over k variables, we consider the program given by the k^{th} level of the mixed hierarchy. To describe the relaxation when the variables in the CSP are boolean, we introduce real variables Y_S for all subsets of variables with $|S| \leq k$.

Let m be the number of constraints. If S_i is the set of variables involved in the i^{th} constraint, then one can find a multilinear polynomial in the variables $\{x_j\}_{j\in S_i}$ which is 1 when the values of the variables satisfy the constraint and 0 otherwise. Let the polynomial be $\sum_{T_i\subseteq S_i} c_{i,T_i}\cdot\left(\prod_{j\in T_i} x_j\right)$. We then let the term in the objective function corresponding to the i^{th} constraint be $\sum_{T_i\subseteq S_i} c_{i,T_i} Y_{T_i}$. For example, in the case of an inequality constraint between x_1 and x_2 as in MAX-CUT, the polynomial is $1-x_1-x_2+2x_1x_2$ and the term in the objective function is $Y_\emptyset - Y_{\{1\}} - Y_{\{2\}} + 2Y_{\{1,2\}}$ (where $Y_\emptyset = 1$).

$$
\begin{array}{lll}
\text{maximize} & \displaystyle\sum_{i=1}^{m}\sum_{T_i\subseteq S_i} c_{i,T_i}\cdot Y_{T_i} & \\
\text{subject to} & (Y_S)_{S\subseteq T}\in I_k & |T| = k \\
& M_1(Y)\succeq 0 &
\end{array}
$$

Fig. 11. Mixed level-k relaxation for a k-CSP

Note that the condition $M_1(Y) \succeq 0$ is equivalent to the existence of vectors $\mathbf{u}_0,\ldots,\mathbf{u}_n$ such that $Y = \|\mathbf{u}_0\|^2 = 1$ and $\langle \mathbf{u}_0, \mathbf{u}_i\rangle = \langle \mathbf{u}_i, \mathbf{u}_i\rangle = Y_{\{i\}}$, $\left\langle \mathbf{u}_i, \mathbf{u}_j\right\rangle = Y_{\{i,j\}}$

for $1 \leq i, j \leq n$. The above program can also expressed in terms of the variables $X_{(S,\alpha)}$ for partial assignments as described in Section 2.2. The formulation in terms of partial assignments can then be generalized for CSPs with non-boolean variables while the above program is specialized to the case of boolean variables.

It is easy to check that for MAX-CUT, the above relaxation is equivalent to the one discussed in Section 3.1. For MAX-CUT, it was shown by Khot et al. [36] that it is NP-hard to achieve an approximation better than $1/C_{GW} - \varepsilon$ for any $\varepsilon > 0$ (assuming the Unique Games Conjecture). This was significantly generalized by Raghavendra who showed that given an instance with integrality gap α for the above SDP for a given type of constraints, one can convert an algorithm achieving an approximation $\alpha - \varepsilon$ for the corresponding CSP, to an algorithm for the Unique Games problem. Assuming the Unique Games Conjecture, one then gets that it is NP-hard to achieve an approximation ratio better than the integrality gap of the above program. This points to a very interesting connection between the the the power of convex relaxations and those of general polynomial time algorithms.

References

1. Michael Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong nonapproximability results in the Lovasz-Schrijver hierarchy. In *Proceedings of the 37th ACM Symposium on Theory of Computing*, pages 294–303, 2005.
2. Noga Alon and Nabil Kahale. Approximating the independence number via the theta-function. *Math. Program.*, 80:253–264, 1998.
3. S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21(1):1–21, 2008.
4. Sanjeev Arora, Béla Bollobás, and László Lovász. Proving integrality gaps without knowing the linear program. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pages 313–322, 2002.
5. 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.
6. Sanjeev Arora, Moses Charikar, and Eden Chlamtac. New approximation guarantee for chromatic number. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 215–224, New York, NY, USA, 2006. ACM.
7. Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows and a $\sqrt{\log n}$-approximation to sparsest cut. In *Proceedings of the 36th ACM Symposium on Theory of Computing*, 2004.
8. MohammadHossein Bateni, Moses Charikar, and Venkatesan Guruswami. Maxmin allocation via degree lower-bounded arborescences. In *41st annual ACM symposium on Theory of computing*, pages 543–552, New York, NY, USA, 2009. ACM.
9. Eli Ben-Sasson. *Expansion in Proof Complexity*. PhD thesis, Hebrew University, 2001.
10. Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow: Resolution made simple. *Journal of the ACM*, 48(2), 2001.
11. A. Blum and D. Karger. An $\tilde{O}(n^{3/14})$ coloring algorithm for 3-colorable graphs. *Information Processing Letters*, 61(1):49–53, 1997.
12. Avrim Blum. New approximation algorithms for graph coloring. *J. ACM*, 41(3):470–516, 1994.

13. Josh Buresh-Oppenheim, Nicola Galesi, Shlomo Hoory, Avner Magen, and Toniann Pitassi. Rank bounds and integrality gaps for cutting planes procedures. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pages 318–327, 2003.
14. Moses Charikar. On semidefinite programming relaxations for graph coloring and vertex cover. In *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms*, pages 616–620, 2002.
15. Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. In *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms*, pages 62–68, 2007.
16. Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for sherali-adams relaxations. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009.
17. Jeff Cheeger, Bruce Kleiner, and Assaf Naor. A (log n)$^{\mathrm{omega(1)}}$ integrality gap for the sparsest cut sdp. In *FOCS*, pages 555–564, 2009.
18. Eden Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *FOCS*, pages 691–701, 2007.
19. Eden Chlamtac and Gyanit Singh. Improved approximation guarantees through higher levels of sdp hierarchies. In *APPROX-RANDOM*, pages 49–62, 2008.
20. V. Chvtal. Edmonds polytopes and a hierarchy of combinatorial problems. *Discrete Mathematics*, 4(4):305 – 337, 1973.
21. Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *SODA*, pages 53–61, 2007.
22. Nikhil R. Devanur, Subhash Khot, Rishi Saket, and Nisheeth K. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *STOC*, pages 537–546, 2006.
23. Uriel Feige and Joe Kilian. Zero knowledge and the chromatic number. *Journal of Computer and System Sciences*, 57(2):187–199, 1998.
24. Uriel Feige and Eran Ofek. Random 3CNF formulas elude the Lovász theta function. Manuscript, 2006.
25. Peter Frankl and Vojtech Rodl. Forbidden intersections. *Transactions of the American Mathematical Society*, 300(1):259–286, 1987.
26. Konstantinos Georgiou, Avner Magen, Toniann Pitassi, and Iannis Tourlakis. Integrality gaps of 2 - o(1) for vertex cover SDPs in the Lovász-Schrijver hierarchy. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 702–712, 2007.
27. Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. Optimal sherali-adams gaps from pairwise independence. In *APPROX-RANDOM*, 2009.
28. M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995. Preliminary version in *Proc. of STOC'94*.
29. Nebojša Gvozdenović and Monique Laurent. The operator ψ for the chromatic number of a graph. *SIAM J. on Optimization*, 19(2):572–591, 2008.
30. Eran Halperin, Ram Nathaniel, and Uri Zwick. Coloring k-colorable graphs using relatively small palettes. *J. Algorithms*, 45(1):72–90, 2002.
31. O. H. Ibarra and C. E. Kim. Fast approximation for the knapsack and sum of subset problems. *Journal of the ACM*, 22:463–468, 1975.
32. Satyen Kale and C. Seshadhri. Combinatorial approximation algorithms for maxcut using random walks. *CoRR*, abs/1008.3938, 2010.

33. D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. In *Proceedings of the 35th IEEE Symposium on Foundations of Computer Science*, pages 2–13, 1994.
34. Anna R. Karlin, Claire Mathieu, and C. Thach Nguyen. Integrality gaps of linear and semi-definite programming relaxations for knapsack. *CoRR*, abs/1007.1283, 2010.
35. Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 767–775, 2002.
36. Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 146–154, 2004.
37. Subhash Khot and Rishi Saket. SDP integrality gaps with local ℓ_1-embeddability. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, 2009.
38. Subhash Khot and Nisheeth Vishnoi. The unique games conjecture, integrality gap for cut problems and the embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, pages 53–63, 2005.
39. Jon M. Kleinberg and Michel X. Goemans. The Lovász Theta function and a semidefinite programming relaxation of vertex cover. *SIAM Journal on Discrete Mathematics*, 11:196–204, 1998.
40. Robert Krauthgamer and Yuval Rabani. Improved lower bounds for embeddings intol$_1$\$. *SIAM J. Comput.*, 38(6):2487–2498, 2009.
41. Michael Krivelevich, Ram Nathaniel, and Benny Sudakov. Approximating coloring and maximum independent sets in 3-uniform hypergraphs. In *SODA*, pages 327–328, 2001.
42. Jean B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Proceedings of the 8th Conference on Integer Programming and Combinatorial Optimization*, pages 293–303, London, UK, 2001. Springer-Verlag.
43. Monique Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, 28(3):470–496, 2003.
44. Eugene L. Lawler. Fast approximation algorithms for knapsack problems. *Math. Oper. Res.*, 4(4):339–356, 1979.
45. James R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *SODA*, pages 92–101, 2005.
46. L. Lovasz. On the shannon capacity of a graph. *Information Theory, IEEE Transactions on*, 25(1):1–7, Jan 1979.
47. L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. on Optimization*, 1(12):166–190, 1991.
48. A. Magen and M. Moharrami. Robust algorithms for maximum independent set on minor-free graphs based on the Sherali-Adams hierarchy. In *12th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, pages 258–271, 2009.
49. Prasad Raghavendra. Optimal algorithms and inapproximability results for every csp? In *STOC*, pages 245–254, 2008.
50. Prasad Raghavendra and David Steurer. Integrality gaps for strong SDP relaxations of unique games. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, 2009.
51. Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-csps. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, 2008.
52. Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. A linear round lower bound for Lovász-Schrijver SDP relaxations of vertex cover. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*, 2007.

53. Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of vertex cover and max cut. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, 2007.
54. Alexander Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Transactions on Information Theory*, 25(4):425–429, 1979.
55. Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990.
56. José Soto. Improved analysis of a max cut algorithm based on spectral partitioning. *CoRR*, abs/0910.0504, 2009.
57. Iannis Tourlakis. New lower bounds for vertex cover in the Lovasz-Schrijver hierarchy. In *Proceedings of the 21st IEEE Conference on Computational Complexity*, 2006.
58. Luca Trevisan. Max cut and the smallest eigenvalue. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 263–272, New York, NY, USA, 2009. ACM.
59. Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009.
60. David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(1):103–128, 2007.