

The Lasserre hierarchy in Approximation algorithms

Lecture Notes for the MAPSP 2013 Tutorial

Preliminary version*

Thomas Rothvoß

June 24, 2013

Abstract

The *Lasserre hierarchy* is a systematic procedure to strengthen a relaxation for an optimization problem by adding additional variables and SDP constraints. In the last years this hierarchy moved into the focus of researchers in approximation algorithms as the obtained relaxations have provably nice properties. In particular on the t -th level, the relaxation can be solved in time $n^{O(t)}$ and every constraint that one could derive from looking just at t variables is automatically satisfied. Additionally, it provides a vector embedding of events so that probabilities are expressible as inner products.

The goal of these lecture notes is to give short but rigorous proofs of all key properties of the Lasserre hierarchy. In the second part we will demonstrate how the Lasserre SDP can be applied to (mostly **NP**-hard) optimization problems such as **KNAPSACK**, **MATCHING**, **MAXCUT** (in general and in dense graphs), **3-COLORING** and **SETCOVER**.

Contents

1	Introduction	2
2	Properties of the Lasserre hierarchy	3
2.1	Definition of Lasserre	3
2.2	Some basic properties	4
2.3	Writing y as a convex combination	6
2.4	Locally consistent probability distributions	10

*I'm still fixing mistakes and extending this document. If you spot any non-trivial flaw, I would be glad for any feedback. My email address is `rothvoss@math.mit.edu`.

2.5	The vector representation	11
2.6	The Decomposition Theorem	11
3	Applications	14
3.1	Scheduling on 2 machines with precedence constraints	14
3.2	Set Cover	17
3.3	Matching	18
3.4	MaxCut	19
3.5	Global Correlation Rounding	21
3.5.1	Application to MaxCut in dense graphs	23

1 Introduction

A large fraction of results in approximation algorithms is obtained by solving a *linear program*, say $\max\{c^T x \mid Ax \geq b\}$ and then rounding the obtained fractional solution to an integral one. For some problems, this already provides an approximate answer whose quality matches known hardness results. For others however, the *integrality gap* may be larger than the suspected hardness. In this case, one has the option to inspect the integrality gap construction and try to derive additional inequalities which are violated, hoping that this decreases the overall gap. Another option is to think about the use of *semidefinite programs*, see the example of MAXCUT [GW95]. It may even be the case that the natural semidefinite program is not enough and even additional constraints for the SDP are needed in order to make progress on existing approximation algorithms (SPARSE CUT is an example [ARV04]).

Instead of following the heuristic approach of finding inequalities that may be helpful for an LP or SDP, there is a more systematic (and potentially more powerful) approach lying in the use of *LP or SDP hierarchies*. In particular there are procedures by Balas, Ceria, Cornuéjols [BCC93]; Lovász, Schrijver [LS91] (with LP-strengthening LS and an SDP-strengthening LS_+); Sherali, Adams [SA90] or Lasserre [Las01a, Las01b]. On the t -th level, they all use $n^{O(t)}$ additional variables to strengthen an initial relaxation $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ (thus the term *Lift-and-Project Methods*) and they all can be solved in time $n^{O(t)}$. Moreover, for $t = n$ they define the integral hull K_I and for any set of $|S| \leq t$ variables, a solution x can be written as convex combinations of vectors from K that are integral on S . Despite these similarities, the Lasserre SDP relaxation is strictly stronger than all the others. Thus for the sake of a simpler exposition, we will only consider the Lasserre hierarchy in this lecture notes. We refer to the survey of Laurent [Lau03] for a detailed comparison with other hierarchies.

Up to now, there have been few (positive) result on the use of hierarchies in approximation algorithms. One successful application of Chlamtáč [Chl07] uses the 3rd level of the Lasserre relaxation to find $O(n^{0.2072})$ -colorings for 3-colorable graphs. It lies in the range of possibilities that $O(\log n)$ levels of Lasserre might be enough to obtain a coloring with $O(\log n)$ colors in 3-colorable graphs [ACC06]. In fact, for special graph classes,

there has been recent progress by Arora and Ge [AG11]. Chlamtáč and Singh [CS08] showed that $O(1/\gamma^2)$ rounds of a mixed hierarchy can be used to obtain an independent set of size $n^{\Omega(1/\gamma^2)}$ in a 3-uniform hypergraph, whenever it has an independent set of size γn . The Sherali-Adams hierarchy was used in [BCG09] to find degree lower-bounded arborescences.

Guruswami and Sinop provide approximation algorithms for quadratic integer programming problems whose performance guarantees depend on the eigenvalues of the graph Laplacian [GS11]. Also the Lasserre-based approach of [BRS11] for UNIQUE GAMES depends on the eigenvalues of the underlying graph adjacency matrix. Though the $O(\sqrt{\log n})$ -apx of Arora, Rao and Vazirani [ARV04] for SPARSEST CUT does not explicitly use hierarchies, their triangle inequality is implied by $O(1)$ rounds of Lasserre. For a more detailed overview on the use of hierarchies in approximation algorithms, see the recent survey of Chlamtáč and Tulsiani [CT11].

The goal of this lecture notes is two-fold.

- *Part 1:* We give a complete introduction of all basic properties of the Lasserre hierarchy. Here our focus lies on short and simple proofs.
- *Part 2:* We provide several short, but interesting applications in approximation algorithms.

2 Properties of the Lasserre hierarchy

In this section, we provide a definition of the Lasserre hierarchy and all properties that are necessary for our purpose. In our notation, we follow to some extent the survey of Laurent [Lau03]. We want to briefly remind the reader that a matrix $M \in \mathbb{R}^{n \times n}$ is called *positive semidefinite* if $\mathbf{x}^T M \mathbf{x} \geq 0$ for all vectors $\mathbf{x} \in \mathbb{R}^n$. Moreover, if M is *symmetric*, then the following conditions are equivalent:

- a) M is positive semidefinite
- b) Any principal submatrix U of M has $\det(U) \geq 0$.
- c) For all $i \in [n]$, there is a vector \mathbf{v}_i such that $M_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$.

For any vector $\mathbf{y} \in \mathbb{R}^n$, recall that the *outer product* $\mathbf{y}\mathbf{y}^T \in \mathbb{R}^{n \times n}$ is a rank-1 matrix that is positive semidefinite as $\mathbf{x}^T (\mathbf{y}\mathbf{y}^T) \mathbf{x} = \langle \mathbf{x}, \mathbf{y} \rangle^2 \geq 0$.

2.1 Definition of Lasserre

Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ some linear relaxation of a binary optimization problem and our goal is to add additional variables and constraints in order to make the relaxation stronger. Our final (and unreachable) goal is to have $x \in \text{conv}(K \cap \{0, 1\}^n)$, in which case we could consider x as a probability distribution over 0/1 vectors such that if we draw a vector $X \in K \cap \{0, 1\}^n$ from that distribution, then $\Pr[X_i = 1] = x_i$. Observe that we can

interpret the fractional values x_i as marginal probabilities, but they do not say much about the probability of *joint* events. For example from those marginals we just know that we should have $\Pr[X_i = X_j = 1] \in [\max\{x_i + x_j - 1, 0\}, \min\{x_i, x_j\}]$ for $i \neq j$. So our idea is to add additional variables that also give a value on joint probabilities.

Let $y \in \mathbb{R}^{2^m}$ be a 2^m dimensional vector, indexed by subsets of variables. What we ideally want is that $y_I = \Pr[\bigwedge_{i \in I} (X_i = 1)]$ (and in particular $y_{\{i\}} = x_i$ and $y_\emptyset = 1$), and in fact, it is possible to explicitly write down constraints that would imply this properties. However, as y is 2^m -dimensional and we aim at polynomial time algorithms, we will only introduce constraints that involve entries y_I for $|I| \leq O(t)$ and t is the *number of rounds* or the *level* of our solution. This way we can still solve the system in time $n^{O(t)}$.

Let $\mathcal{P}_t(\{n\}) := \{I \subseteq [n] \mid |I| \leq t\}$ be the set of all index sets of cardinality at most t .

Definition 1. Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$. We define the t -th level of the Lasserre hierarchy $\text{LAS}_t(K)$ as the set of vectors $y \in \mathbb{R}^{2^m}$ that satisfy

$$M_t(y) := (y_{I \cup J})_{|I|, |J| \leq t} \geq 0; \quad M_t^\ell(y) := \left(\sum_{i=1}^n A_{\ell i} y_{I \cup J \cup \{i\}} - b_\ell y_{I \cup J} \right)_{|I|, |J| \leq t} \geq 0 \quad \forall \ell \in [m]; \quad y_\emptyset = 1.$$

The matrix $M_t(y)$ is called the *moment matrix* of y and the second type $M_t^\ell(y)$ is called *moment matrix of slacks*. Furthermore, let $\text{LAS}_t^{\text{proj}}(K) := \{(y_{\{1\}}, \dots, y_{\{m\}}) \mid y \in \text{LAS}_t(K)\}$ be the projection on the original variables.

We want to emphasize that the t -th level of Lasserre cares only for entries y_I with $|I| \leq 2t + 1$ – all other entries are ignored (if the reader prefers a proper definition, he might impose $y_I = 0$ for $|I| > 2t + 1$). The set of positive semidefinite matrices forms a non-polyhedral convex cone, thus also $\text{LAS}_t(K)$ is a convex set. The separation problem for $\text{LAS}_t(K)$ boils down to computing a negative Eigenvector of the involved matrices, which can be done in polynomial time. As a consequence, if m denotes the number of linear constraints, one optimize over $\text{LAS}_t(K)$ in time $n^{O(t)} \cdot m^{O(1)}$ up to numerical errors that can be neglected as we are talking about approximation algorithms anyway.

Intuitively speaking, the constraint $M_t(y) \geq 0$ takes care that the variables are consistent (e.g. it guarantees that $y_{\{1,2\}} \in [y_{\{1\}} + y_{\{2\}} - 1, \min\{y_{\{1\}}, y_{\{2\}}\}]$). Moreover, the psdness of the ℓ th slack matrices $M_t^\ell(y)$ guarantees that y satisfies the ℓ -th linear constraint. More generally, the Lasserre hierarchy can even be applied to non-convex semi-algebraic sets – but for the sake of a simple presentation we stick to polytopes, again the interested reader is referred to [Lau03]¹.

2.2 Some basic properties

Let us begin with discussing some of the basic properties, especially that $\text{LAS}_t(K)$ is indeed a relaxation of $K \cap \mathbb{Z}^n$. Moreover we have $0 \leq y_I \leq 1$ regardless to whether or not

¹in fact, we do deviate in terms of the notation from [Lau03]. In particular, we consider y as a 2^m dimensional vector, while Laurent considers it as $n^{O(t)}$ dimensional. Moreover, Laurent uses $M_{t+1}(y)$ as moment matrix on the t -th level.

these constraints are included in $Ax \geq b$. For a square matrix M , we abbreviate the determinant by $|M| = \det(M)$.

Lemma 1. *Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ and $y \in \text{LAS}_t(K)$ for $t \geq 0$. Then the following holds:*

- a) $K \cap \{0, 1\}^n \subseteq \text{LAS}_t^{\text{proj}}(K)$.
- b) $0 \leq y_I \leq 1$ for all $|I| \leq t$
- c) One has $0 \leq y_J \leq y_I \leq 1$ for all $I \subseteq J$ with $0 \leq |I| \leq |J| \leq t$.
- d) $|y_{I \cup J}| \leq \sqrt{y_I \cdot y_J}$ for $|I|, |J| \leq t$.
- e) $\text{LAS}_t^{\text{proj}}(K) \subseteq K$
- f) $\text{LAS}_0(K) \supseteq \text{LAS}_1(K) \supseteq \dots \supseteq \text{LAS}_n(K)$

Proof. The properties can be proven as follows:

- a) Let $x \in K \cap \{0, 1\}^n$ be a feasible integral solution. Then we can define $y_I := \prod_{i \in I} x_i$. We claim that indeed $y \in \text{LAS}_t(K)$ for any $t \geq 0$. First of all, the moment matrix $M_t(y)$ is actually a submatrix of the positive semidefinite rank-1 matrix yy^T since for any entry (I, J) , we have

$$(M_t(y))_{I,J} = y_{I \cup J} = \prod_{i \in I \cup J} x_i = \left(\prod_{i \in I} x_i \right) \cdot \left(\prod_{i \in J} x_i \right) = y_I \cdot y_J = (yy^T)_{I,J}$$

using that $x_i^2 = x_i$ for $x_i \in \{0, 1\}$. Now consider one of the constraints in $Ax \geq b$ and say it is of the form $ax \geq \beta$. We claim that the corresponding slack moment matrix is a submatrix of $(ax - \beta) \cdot yy^T$ which is again a positive semidefinite rank-1 matrix. Here we use crucially that the scalar $ax - \beta$ is non-negative because x is a feasible solution to K . To verify this we inspect the (I, J) -entry which is of the form

$$\sum_{i=1}^n a_i y_{I \cup J \cup i} - \beta y_{I \cup J} = (ax - \beta) \cdot y_I \cdot y_J$$

where we again use that $x_i^2 = x_i$.

- b) Let $|I| \leq t$. The determinant of the principal submatrix of $M_t(y)$ indexed by $\{\emptyset, I\}$ is

$$\begin{array}{c} \emptyset \quad I \\ \emptyset \quad \left| \begin{pmatrix} y_\emptyset & y_I \\ y_I & y_I \end{pmatrix} \right| \\ I \quad \left(\begin{array}{c} y_I \\ y_I \end{array} \right) \end{array} = y_I(1 - y_I) \geq 0$$

hence $0 \leq y_I \leq 1$.

- c) Similarly to *b*), we consider the determinant of the principal submatrix of $M_t(y)$ indexed by $\{I, J\}$ and obtain

$$\begin{matrix} I & J \\ \left(\begin{array}{cc} y_I & y_J \\ y_J & y_J \end{array} \right) \end{matrix} = y_J(y_I - y_J) \geq 0$$

using that $I \cup J = J$. We know from *b*), that $y_J, y_I \geq 0$, hence $y_I \geq y_J$.

- d) This case is actually useful because $y_{I \cup J}$ does not need to appear on the diagonal. But still we can consider the determinant of the principal submatrix of $M_t(y)$ indexed by $\{I, J\}$ and obtain

$$\begin{matrix} I & J \\ \left(\begin{array}{cc} y_I & y_{I \cup J} \\ y_{I \cup J} & y_J \end{array} \right) \end{matrix} = y_I y_J - y_{I \cup J}^2 \geq 0$$

which shows that $|y_{I \cup J}| \leq \sqrt{y_I y_J}$.

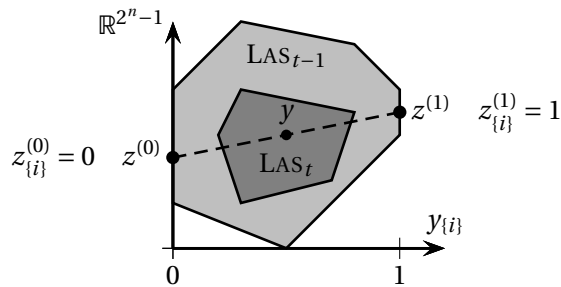
- e) If we consider the slack moment matrix of a constraint $ax \geq \beta$ then the diagonal entry indexed with (\emptyset, \emptyset) is of the form $\sum_{i=1}^n a_i y_{\emptyset \cup \{i\}} - \beta y_\emptyset = a^T(y_1, \dots, y_n) - \beta \geq 0$, thus $(y_1, \dots, y_n) \in K$.
- f) This is clear as the $M_t(y)$ is a submatrix of $M_{t+1}(y)$ and in particular all principal submatrices of a PSD matrix are again PSD. The same holds true for the moment slack matrices.

□

2.3 Writing y as a convex combination

If we have a vector $x \in \text{conv}(K \cap \{0, 1\}^n)$, then it can be written as a convex combination of integral vectors in K . This is not true anymore for any $x \in K$. But for a round- t Lasserre solution y a weaker claim is true: for any subset S of t variables, y can be written as a convex combination of fractional solutions that are integral at least on those variables in S . It actually suffices to show this claim for one variable – the rest follows by induction.

More concrete it means that for a solution $y \in \text{LAS}_t(K)$ we can pick any variable i where y_i is not yet integral. Then it can be written as convex combination of vectors $z^{(0)}$ and $z^{(1)}$ such that the i th variable in $z^{(0)}$ is 0 and in $z^{(1)}$ the entry is 1. A schematic (but non-sense) visualization can be found below.



In particular we can give a precise formula for $z^{(0)}$ and $z^{(1)}$.

Lemma 2. For $t \geq 1$, let $y \in \text{LAS}_t(K)$ and $i \in [n]$ be a variable with $0 < y_i < 1$. If we define

$$z_I^{(1)} := \frac{y_{I \cup \{i\}}}{y_i} \quad \text{and} \quad z_I^{(0)} := \frac{y_I - y_{I \cup \{i\}}}{1 - y_i}$$

then we have $y = y_i \cdot z^{(1)} + (1 - y_i) \cdot z^{(0)}$ with $z^{(0)}, z^{(1)} \in \text{LAS}_{t-1}(K)$ and $z_i^{(0)} = 0, z_i^{(1)} = 1$.

As this is a crucial statement, we want to give two proofs.

Proof 1. Clearly $z_i^{(1)} = \frac{y_i}{y_i} = 1, z_i^{(0)} = \frac{y_i - y_i}{1 - y_i} = 0$ and

$$y_i \cdot z_I^{(1)} + (1 - y_i) \cdot z_I^{(0)} = y_{I \cup \{i\}} + (y_I - y_{I \cup \{i\}}) = y_I$$

thus it only remains to argue that $z^{(0)}, z^{(1)} \in \text{LAS}_{t-1}(K)$.

For this sake, define index sets Let $\mathcal{I}_{-i} := \{I \mid |I| < t, i \notin I\}$ and $\mathcal{I}_{+i} := \{I \cup \{i\} \mid |I| < t, i \notin I\}$. Let M_{-i} and M_{+i} be the principal submatrices of $M_t(y)$ indexed by \mathcal{I}_{-i} and \mathcal{I}_{+i} , resp. In other words, M_{-i}, M_{+i} are square matrices of the same size and in particular $M_{-i}, M_{+i} \geq 0$. We already noted that y is expressed as convex combination – but the same is also true for the moment matrix. In fact, we can write:

$$M_t(y) = \begin{array}{c} \mathcal{I}_{-i} \\ \mathcal{I}_{+i} \end{array} \begin{array}{c} \mathcal{I}_{-i} \quad \mathcal{I}_{+i} \\ \begin{bmatrix} M_{-i} & M_{+i} & * \\ M_{+i} & M_{+i} & * \\ * & * & * \end{bmatrix} \end{array} = y_i \cdot \begin{array}{c} M_{t-1}(z^{(1)}) \\ \cap \\ \begin{bmatrix} \frac{1}{y_i} M_{+i} & \frac{1}{y_i} M_{+i} & * \\ \frac{1}{y_i} M_{+i} & \frac{1}{y_i} M_{+i} & * \\ * & * & * \end{bmatrix} \end{array} + (1 - y_i) \cdot \begin{array}{c} M_{t-1}(z^{(0)}) \\ \cap \\ \begin{bmatrix} \frac{1}{1-y_i} (M_{-i} - M_{+i}) & \mathbf{0} & * \\ \mathbf{0} & \mathbf{0} & * \\ * & * & * \end{bmatrix} \end{array}$$

In other words, the new moment matrices $M_{t-1}(z^{(1)})$ and $M_{t-1}(z^{(0)})$ are submatrices of the dashed matrices on the right hand side. As $M_{t-1}(z^{(1)})$ is a submatrix of a scaled and cloned copy of M_{+i} , we know that $M_{t-1}(z^{(1)}) \geq 0$. After scaling and removing zeros, $M_{t-1}(z^{(0)})$ becomes $M_{-i} - M_{+i}$. While it is clear, that both M_{-i} and M_{+i} are positive semidefinite, this is less obvious for their difference. However, for each vector w we have

$$w^T (M_{-i} - M_{+i}) w = (w, -w)^T \begin{pmatrix} M_{-i} & M_{+i} \\ M_{+i} & M_{+i} \end{pmatrix} (w, -w) \geq 0$$

and hence $M_{-i} - M_{+i} \geq 0$.

Now consider the ℓ th constraint for K and abbreviate it with $ax \geq \beta$. Let M_{-i}^ℓ and M_{+i}^ℓ be the principal submatrices of $M_t^\ell(y)$ that are indexed by \mathcal{I}_{-i} and \mathcal{I}_{+i} , resp. Then we get a picture that is identical to the one before, namely we can consider $M_t^\ell(y)$ as convex combination

Lemma 4. Let $y \in \text{LAS}_t(K)$ and $S \subseteq [n]$. For $J_0, J_1 \subseteq [n]$ abbreviate $y_I^{J_0, J_1} = \sum_{H \subseteq J_0} (-1)^{|H|} y_{I \cup J_1 \cup H}$. Then we can express y as convex combination

$$y = \sum_{J_0 \cup J_1 = S: y_{\emptyset}^{J_0, J_1} > 0} y_{\emptyset}^{J_0, J_1} \cdot \underbrace{\left(\frac{y_{\emptyset}^{J_0, J_1}}{y_{\emptyset}^{J_0, J_1}} \right)}_{=: z^{J_0, J_1}} \quad \text{with} \quad z_i^{J_0, J_1} = \begin{cases} 1 & i \in J_1 \\ 0 & i \in J_0 \end{cases} \quad (1)$$

Moreover $z^{J_0, J_1} \in \text{LAS}_{t-|S|}(K)$.

Proof. Suppose by induction the claim (1) is shown for S and we want to show it for $S \cup \{i\}$. By definition $y_{I \cup \{i\}}^{J_0, J_1} = \sum_{H \subseteq J_0} (-1)^{|H|} y_{I \cup \{i\} \cup J_1 \cup H} = y_I^{J_0, J_1 \cup \{i\}}$. Furthermore

$$y_I^{J_0, J_1} - y_{I \cup \{i\}}^{J_0, J_1} = \sum_{H \subseteq J_0} (-1)^{|H|} y_{I \cup J_1 \cup H} - \sum_{H \subseteq J_0} (-1)^{|H|} y_{I \cup J_1 \cup H \cup \{i\}} = \sum_{H \subseteq J_0 \cup \{i\}} (-1)^{|H|} y_{I \cup J_1 \cup H} = y_I^{J_0 \cup \{i\}, J_1}.$$

By the properties of Lemma 2 the claim follows. \square

One crucial observation is that for the convex combination in Lemma 4, the order in which we condition on variables is irrelevant – we always end up with the same formula.

We wonder, where the formula in Lemma 4 is coming from. Suppose that $y \in \text{LAS}_n(K)$, thus there is indeed a consistent random variable $X \in K \cap \{0, 1\}^n$ with $\Pr[\bigwedge_{i \in I} (X_i = 1)] = y_I$. Now consider indices $J_0, J_1 \subseteq [n]$, then the *inclusion-exclusion formula* says that

$$\Pr \left[\bigvee_{i \in J_0} (X_i = 1) \right] = \sum_{\emptyset \subset H \subseteq J_0} (-1)^{|H|+1} \Pr \left[\bigwedge_{i \in H} (X_i = 1) \right].$$

Negating this event yields

$$\Pr \left[\bigwedge_{i \in J_0} (X_i = 0) \right] = 1 - \Pr \left[\bigvee_{i \in J_0} (X_i = 1) \right] = \sum_{H \subseteq J_0} (-1)^{|H|} \Pr \left[\bigwedge_{i \in H} (X_i = 1) \right]. \quad (2)$$

Observe that Equation (2) remains valid if all events are intersected with the same event $\bigwedge_{i \in J_1} (X_i = 1)$. In other words we arrive at the *generalized inclusion exclusion formula* (sometimes called *Möbius inversion*)

$$\Pr \left[\bigwedge_{i \in J_1} (X_i = 1) \wedge \bigwedge_{i \in J_0} (X_i = 0) \right] = \sum_{H \subseteq J_0} (-1)^{|H|} \Pr \left[\bigwedge_{i \in J_1 \cup H} (X_i = 1) \right] = \sum_{H \subseteq J_0} (-1)^{|H|} y_{J_1 \cup H} = y_{\emptyset}^{J_0, J_1} \quad (3)$$

We can use this to rewrite

$$\begin{aligned} z_I^{J_0, J_1} &= \frac{y_I^{J_0, J_1}}{y_{\emptyset}^{J_0, J_1}} = \frac{\Pr[\bigwedge_{i \in J_1 \cup I} (X_i = 1) \wedge \bigwedge_{i \in J_0} (X_i = 0)]}{\Pr[\bigwedge_{i \in J_1} (X_i = 1) \wedge \bigwedge_{i \in J_0} (X_i = 0)]} \\ &= \Pr \left[\bigwedge_{i \in I} (X_i = 1) \mid \bigwedge_{i \in J_1} (X_i = 1) \wedge \bigwedge_{i \in J_0} (X_i = 0) \right] \end{aligned}$$

If the reader prefers to see some applications right now, we have all tools together that we need for Sections 3.1 and 3.2.

2.4 Locally consistent probability distributions

If we have a solution $x \in \text{conv}(K \cap \{0, 1\}^n)$ then there is an obvious way to consider x as a probability distribution over integral solutions in K . This interpretation carries over nicely to t -round Lasserre solution, just that we cannot expect a *globally* feasible probability distribution, but at least one that is *locally* consistent when considering subsets of at most t variables.

Lemma 5. *Let $y \in \text{LAS}_t(K)$. Then for any subset $S \subseteq [n]$ of size $|S| \leq t$ there is a distribution $\mathcal{D}(S)$ over $\{0, 1\}^S$ such that*

$$\Pr_{z \sim \mathcal{D}(S)} \left[\bigwedge_{i \in I} (z_i = 1) \right] = y_I \quad \forall I \subseteq S$$

Proof. We write y as convex combination of vectors $y^{(z)}$ that are integral on S following Cor. 3. In fact, let $y = \sum_{z \in \{0, 1\}^S} \lambda_z y^{(z)}$ such that $y_i^{(z)} = z_i \in \{0, 1\}$ for $i \in S$ and $y^{(z)} \in \text{LAS}_{t-|S|}(K)$. \square

This property make more sense, when talking about problems that are defined using local constraints. Thus we want to make this concrete using the GRAPH 3-COLORING problem. Let $G = (V, E)$ be an undirected graph and suppose it is 3-colorable, i.e. the nodes can be colored with **red**, **green** and **blue** such that adjacent node get assigned different colors. Then the following relaxation looks for a such an assignment

$$\{x \in [0, 1]^n \mid x_{i,c} + x_{j,c} \leq 1 \quad \forall c \in \{R, G, B\} \quad \forall (i, j) \in E\}$$

and the decision variables $x_{i,c}$ tells whether node i gets assigned color $c \in \{R, G, B\}$.

Lemma 6. *Let $y \in \text{LAS}_{3t}(K)$ be any solution for 3-coloring then there is family of distributions $\{\mathcal{D}(S)\}_{S \subseteq V: |S| \leq t}$ such that for any $S \subseteq V$ with $|S| \leq t$ one has*

- a) *each event $\chi \sim \mathcal{D}(S)$ is a valid 3-coloring $\chi : S \rightarrow \{R, G, B\}$ for the induced subgraph $G[S]$*
- b) $\Pr_{\chi \sim \mathcal{D}(S)} [\chi(i_1) = c_1, \dots, \chi(i_k) = c_k] = y_{\{(i_1, c_1), \dots, (i_k, c_k)\}} \quad \forall i_1, \dots, i_k \in S \quad \forall c_1, \dots, c_k \in \{R, G, B\}$

Proof. Fix a set S and write y as a convex combination of $\text{LAS}_0(K)$ solutions that are integral on variables $x_{i,c}$ for $(i, c) \in S \times \{R, G, B\}$. Let $z \in \text{LAS}_0(K)$ be one of those vectors. The vector z induces 3-coloring $\chi : S \rightarrow \{R, G, B\}$ with $\chi(i) = c \Leftrightarrow z_{i,c} = 1$. Now consider an event defined by $i_1, \dots, i_k \in S$ and $c_1, \dots, c_k \in \{R, G, B\}$. \square

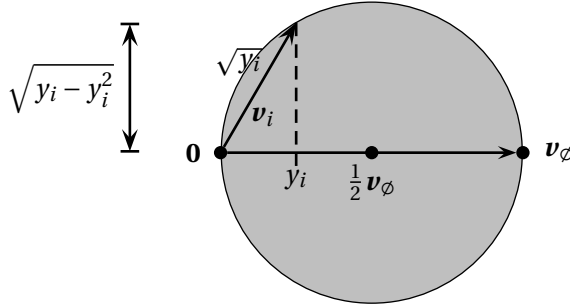
In particular, (b) implies that different distributions $\mathcal{D}(S)$ and $\mathcal{D}(S')$ are consistent with each other, i.e. events that are defined for their intersection $S \cap S'$ have the same probability in $\mathcal{D}(S)$ and $\mathcal{D}(S')$.

2.5 The vector representation

After discussing the local consistence of Lasserre solutions we want to discuss a very powerful global property which is the *vector representation*. In particular, for each event $\bigwedge_{i \in I} (x_i = 1)$ with $|I| \leq t$ there is a vector \mathbf{v}_I representing it in a consistent way. More formally, the positive semidefiniteness of $M_t(y)$ implies

Lemma 7 (Vector Representation Lemma). *Let $y \in \text{LAS}_t(K)$. Then there is a family of vectors $(\mathbf{v}_I)_{|I| \leq t}$ such that $\langle \mathbf{v}_I, \mathbf{v}_J \rangle = y_{I \cup J}$ for all $|I|, |J| \leq t$. In particular $\|\mathbf{v}_I\|_2^2 = y_I$ and $\|\mathbf{v}_\emptyset\|_2^2 = 1$.*

It helps to have a geometric picture in mind. Each vector \mathbf{v}_I lies on the sphere with radius $\frac{1}{2}$ and center $\frac{1}{2}\mathbf{v}_\emptyset$. This is easy to check as $\|\mathbf{v}_I - \frac{1}{2}\mathbf{v}_\emptyset\|_2^2 = \|\mathbf{v}_I\|_2^2 - 2 \cdot \frac{1}{2}\mathbf{v}_I \mathbf{v}_\emptyset + \frac{1}{4}\|\mathbf{v}_\emptyset\|_2^2 = \frac{1}{4}$.



More generally for $I \supseteq J$, \mathbf{v}_I lies on the sphere of radius $\frac{1}{2}\sqrt{y_J}$ and center $\frac{1}{2}\mathbf{v}_J$. Moreover, all angles between vectors are between 0° and 90° . For example if we have two events I, J with $y_I + y_J = 1$ that are disjoint (i.e. $y_{I \cup J} = 0$), then \mathbf{v}_I and \mathbf{v}_J will be antipodal w.r.t. center $\frac{1}{2}\mathbf{v}_{\text{emptyset}}$. This vector representation is crucial in many rounding algorithm as we will see in Section 3.4.

2.6 The Decomposition Theorem

In this chapter we want to discuss a very strong property of the Lasserre hierarchy that is actually implied by the SDP constraints and which somewhat distinguishes Lasserre from hierarchies that are solely based on linear programming.

We want to demonstrate the issue using a concrete instance of the KNAPSACK problem. Let's say we have n items of unit weight and unit profit and a knapsack of capacity 1.9 as well. Obviously the best KNAPSACK solution has $OPT = 1$, but let us investigate how the Lasserre hierarchy deals with the relaxation

$$K = \left\{ x \in \mathbb{R}_{\geq 0}^n \mid \sum_{i=1}^n x_i \leq 1.9 \right\},$$

²Again, because $\|\mathbf{v}_I - \frac{1}{2}\mathbf{v}_J\|_2^2 = \|\mathbf{v}_I\|_2^2 - \mathbf{v}_I \mathbf{v}_J + \frac{1}{4}\|\mathbf{v}_J\|_2^2 = \frac{1}{4}y_J$.

Let us consider a 2-round Lasserre solution $y \in \text{LAS}_2(K)$. Choose a variable y with $0 < y_i < 1$. We can consider the convex combination $y = y_i \cdot y^{(1)} + (1 - y_i)y^{(0)}$ with $y^{(1)}, y^{(0)} \in \text{LAS}_1(K)$ and $y_i^{(1)} = 1, y_i^{(0)} = 0$. Since this is a convex combination, one of both solutions $y^{(1)}$ or $y^{(0)}$ has to be at least as good as y in terms of the objective function. If $y^{(1)}$ is the better solution, we are done as $y^{(1)} \in \{0, 1\}^n$ is already completely integral (if not then there is a $j \neq i$ with $y_j^{(1)} > 0$; then we can induce on j and obtain a solution $y^{(1,1)} \in \text{LAS}_0(K)$ fully containing 2 items which is a contradiction). On the other hand, if $y^{(0)}$ is the better solution then we have a problem since we have already used one of our 2 precious rounds to make just one out of n variables integral. In fact, this is precisely what happens for the Sherali-Adams hierarchy which does have an integrality gap of $1.9 - O(\frac{1}{n})$ for this case [KMN11].

Fortunately it turns out, we can use a smarter argument for Lasserre. We already saw in the last section that there are vectors \mathbf{v}_I such that $\mathbf{v}_I \cdot \mathbf{v}_J = y_{I \cup J}$. According to our arguments above, $y_{\{i,j\}} = 0$ for all $i \neq j$. In particular, this means that $\mathbf{v}_i \cdot \mathbf{v}_j = y_{\{i,j\}} = 0$ or in other words: the vectors representing different items are orthogonal. Now we can get

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \|\mathbf{v}_i\|_2^2 \stackrel{\mathbf{v}_i \cdot \mathbf{v}_i = \mathbf{v}_i \cdot \mathbf{v}_\emptyset}{=} \sum_{i=1}^n \langle \mathbf{v}_\emptyset, \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|_2} \rangle^2 \leq \|\mathbf{v}_\emptyset\|_2^2 = 1$$

where we use *Pythagoras Theorem* in the inequality. This shows that we have no integrality gap left in this case, i.e. $\text{LAS}_2^{\text{proj}}(K) = \text{conv}(K \cap \{0, 1\}^n)$.

The next question would be, what happens if we have a knapsack polytope $K = \{x \mid \sum_{i=1}^n x_i \leq k + 0.9\}$ for some $k \in \mathbb{N}$. Will the gap of a round $t = k + 1$ Lasserre lifting be 0? In fact, exactly this is the case. The crucial observation is that $y_I = 0$ for all $k < |I| \leq t$. For $x \in \mathbb{R}^n$, let $\text{ones}(x) := \{i \mid x_i = 1\}$ be the number of ones.

Lemma 8. *Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ be any relaxation and suppose that $\max\{|\text{ones}(x)| \mid x \in K\} \leq t$, then $\text{LAS}_{t+1}^{\text{proj}}(K) = \text{conv}(K \cap \{0, 1\}^n)$.*

Proof. Let $y \in \text{LAS}_{t+1}(K)$. First of all, we claim that for all $|I| = t + 1$, we have $y_I = 0$. Suppose for the sake of contradiction that $y_I > 0$. Then we can induce on all variables in I and obtain $\tilde{y} \in \text{LAS}_0(K)$ with $\tilde{y}_i = 1$ for all $i \in I$, which contradicts the assumption that no feasible fractional solution may contain more than t ones. Recall that we actually defined y as a $2^{\lfloor n \rfloor}$ dimensional vector. We define $y_I = 0$ for all $|I| > t$. Note that this is consistent with the off diagonal matrix entries in $M_{t+1}(y)$ as $|y_{I \cup J}| \leq \sqrt{y_I \cdot y_J} = 0$ if $t < |I \cup J| \leq 2t$.

What happens if instead of $M_{t+1}(y)$, we consider the bigger matrix $M_n(y)$?

$$M_n(y) = \begin{pmatrix} \boxed{M_{t+1}(y)} & \boxed{\mathbf{0}} \\ \boxed{\mathbf{0}} & \boxed{\mathbf{0}} \end{pmatrix} \begin{matrix} |I| \leq t+1 & |I| > t+1 \\ |I| \leq t+1 & |I| > t+1 \end{matrix}$$

Obviously extending the matrix by zeros does not violate the positive semidefiniteness. The only potential problem is that there might be entries (I, J) and (I', J') with $I \cup J =$

$I' \cup J'$ such that (I, J) is in $M_t(y)$, but (I', J') is not. But then $|I \cup J| = |I' \cup J'| > t$, thus $y_{I \cup J} = y_{I' \cup J'} = 0$ and the matrix is consistent. The same arguments also holds for $M_{t+1}^\ell(y)$ ³ \square

Finally, we want to give the most general form of the above arguments which is the *Decomposition Theorem* of Karlin, Mathieu and Nguyen [KMN11].

Theorem 9 (Decomposition Theorem). *Let $0 \leq k \leq t$, $y \in \text{LAS}_t(K)$, $S \subseteq [n]$ so that $k \geq |\text{ones}(x) \cap S|$ for all $x \in K$. Then*

$$y \in \text{conv}\{z \mid z \in \text{LAS}_{t-k}(K); z_{\{i\}} \in \{0, 1\} \forall i \in S\}$$

Proof. First, let us make a thought experiment and let us again use the formula from Lemma 4. In other words, we again define $y_I^{J_0, J_1} = \sum_{H \subseteq J_0} (-1)^{|H|} y_{I \cup J_1 \cup H}$. Then the equation

$$y = \sum_{J_0 \dot{\cup} J_1 = S} y^{J_0, J_1}$$

still holds true, no matter what values the y_I 's have. Moreover, if $y_\emptyset^{J_0, J_1} \neq 0$, then $z_i^{J_0, J_1} = 1$ if $i \in J_1$ and $z_i^{J_0, J_1} = 0$ otherwise. What is not clear apriori whether $M_{t-k}(y^{J_0, J_1}) \geq 0$ and $M_{t-k}^\ell(y^{J_0, J_1}) \geq 0$ (in particular if yes, this implies that the used coefficient $y_\emptyset^{J_0, J_1}$ are non-negative).

So far, we were not much concerned about entries that did not appear in $M_t(y)$. For now let us define them to be $y_I := 0$ for $|I| > 2t$. We know that $y_I = 0$ for $|I \cap S| = k + 1$. But by monotonicity (Lemma 1. 1) we still have $y_I = 0$ for all $|I \cap S| \geq k + 1$ with $|I| \leq t$. We should be a bit careful that there are also entries $y_{I \cup J}$ in $M_t(y)$ with $|(I \cup J) \cap S| \geq k + 1$ but $|I \cup J| > t$. However, Lemma 1 provides that $|y_{I \cup J}| \leq \sqrt{y_I \cdot y_J} = 0$ (if we split $I \cup J$ so that $|I \cap S| \geq k + 1$). So we will be consistent if we define $y_I := 0$ for *all* $I \subseteq [n]$ with $|I \cap S| \geq k + 1$.

We know that $M_t(y) \geq 0$, thus there are vectors \mathbf{v}_I such that $\langle \mathbf{v}_I, \mathbf{v}_J \rangle = y_{I \cup J}$ for $|I|, |J| \leq t$. In fact, by choosing $\mathbf{v}_I := \mathbf{0}$ for $|I \cap S| > k$, we know that $\langle \mathbf{v}_I, \mathbf{v}_J \rangle = y_{I \cup J}$ is true for all $I, J \subseteq [n]$ with $|J \setminus S|, |I \setminus S| \leq t - k$.

Now fix $J_0 \dot{\cup} J_1 = S$ and we want to show that $M_{t-k}(y^{J_0, J_1}) \geq 0$. We define $\mathbf{u}_I := \sum_{H \subseteq J_0} (-1)^{|H|} \mathbf{v}_{I \cup J_1 \cup H}$. Then

$$\begin{aligned} \langle \mathbf{u}_I, \mathbf{u}_J \rangle &= \left\langle \sum_{H \subseteq J_0} (-1)^{|H|} \mathbf{v}_{I \cup J_1 \cup H}, \sum_{H' \subseteq J_0} (-1)^{|H'|} \mathbf{v}_{J \cup J_1 \cup H'} \right\rangle \\ &\stackrel{(*)}{=} \sum_{G \subseteq J_0} (-1)^{|G|} y_{I \cup J_1 \cup G} \sum_{H \subseteq G} \underbrace{\sum_{A \subseteq H} (-1)^{|A|}}_{=0 \text{ if } H \neq \emptyset, 1 \text{ ow.}} = \sum_{G \subseteq J_0} (-1)^{|G|} y_{I \cup J \cup J_1 \cup G} = y_{I \cup J}^{J_0, J_1} \end{aligned}$$

In step (*), we multiply out the term and use that $\langle \mathbf{v}_{I \cup J_1 \cup H}, \mathbf{v}_{J \cup J_1 \cup H'} \rangle = y_{I \cup J_1 \cup H \cup H'}$ as $|(I \cup J_1 \cup H) \setminus S| \leq |I| \leq t - k$ and $|(J \cup J_1 \cup H') \setminus S| \leq |J| \leq t - k$. Moreover, we reindexed $G = H \cup H'$ and $A = H \cap H'$ and use that $(-1)^{|H|+|H'|} = (-1)^{|G|} \cdot (-1)^{|A|}$. It follows that $M_{t-k}(y^{J_0, J_1}) \geq 0$.

³Let us abbreviate $s_I := \sum_{\ell=1}^n A_{\ell i} y_{I \cup \{i\}} - y_I b_\ell$. Then for $|I| = t + 1$, we have $s_I = 0$. Again also off-diagonal entries will be zero as $|s_{I \cup J}| \leq \sqrt{s_I \cdot s_J}$ for $|I|, |J| \leq t + 1$.

Next, we want to show that $M_{t-k}^\ell(y^{J_0, J_1}) \geq 0$. We abbreviate the ℓ th constraint by $ax \geq \beta$. Again we know that $M_t^\ell(y) \geq 0$, thus there are vectors \mathbf{v}_I with $\langle \mathbf{v}_I, \mathbf{v}_J \rangle = \sum_{i=1}^n a_i y_{I \cup J \cup \{i\}} - \beta y_{I \cup J}$. We define $\mathbf{u}_I := \sum_{H \in J_0} (-1)^{|H|} \mathbf{v}_{I \cup J_1 \cup H}$ and calculate that

$$\begin{aligned} \langle \mathbf{u}_I, \mathbf{u}_J \rangle &= \left\langle \sum_{H \in J_0} (-1)^{|H|} \mathbf{v}_{I \cup J_1 \cup H}, \sum_{H' \in J_0} (-1)^{|H'|} \mathbf{v}_{I \cup J_1 \cup H'} \right\rangle \\ &= \sum_{G \in J_0} (-1)^{|G|} \left(\sum_{i=1}^n a_i y_{I \cup J \cup G \cup \{i\}} - \beta y_{I \cup J \cup G} \right) = \sum_{i=1}^n a_i y_{I \cup J \cup \{i\}}^{J_0, J_1} - \beta y_{I \cup J}^{J_0, J_1} \end{aligned}$$

and the claim is proven. \square

3 Applications

3.1 Scheduling on 2 machines with precedence constraints

As these are lecture notes for a scheduling workshop, we feel morally obliged to start with a scheduling application. We consider a problem, which in the standard scheduling notation is denoted by $P2 \mid \text{prec}, p_j = 1 \mid C_{\max}$. In words: we are given a set of n jobs J of *unit processing time* with *precedence constraints* on the jobs as well as $m = 2$ identical machines. We will write $i < j$ if job i has to finish before job j can be started. The goal is to schedule the jobs on the machines without preemption or migration so that the makespan (i.e. the time that the last job is finished) is minimized.

For a general number of machines m and running times, Graham's classical *list scheduling* provides a $2 - \frac{1}{m}$ approximation, while the problem is **NP**-hard for general m even with identical running times $p_j = 1$. For the case of two machines that we consider here, it is known that a polynomial algorithm based on matching techniques finds the optimum solution [JG72]. However, here we want to demonstrate a neat application of the Lasserre hierarchy that is due to Svensson [Sve11].

So we consider T time slots of unit length and want to decide whether a makespan of T is possible. We can restrict the jobs to start at integral time units only. It seems natural to consider a straightforward time-indexed formulation with variables x_{jt} saying whether we want to process job $j \in J$ in the interval $[t-1, t]$ for $t \in \{1, \dots, T\}$. Let $K(T)$ be the set of solutions to the following LP:

$$\begin{aligned} \sum_{t=1}^T x_{jt} &= 1 & \forall j \in J \\ \sum_{j \in J} x_{jt} &\leq 2 & \forall t \in [T] \\ \sum_{t' \leq t} x_{it'} &\geq \sum_{t' \leq t+1} x_{jt'} & \forall i < j \quad \forall t \in [T] \\ x_{jt} &\geq 0 & \forall j \in J \quad \forall t \in [T] \end{aligned} \tag{4}$$

First, let us verify that this LP itself does not yet give a correct answer. In fact, the simple instance depicted in Figure 1 already gives a gap of $4/3$.

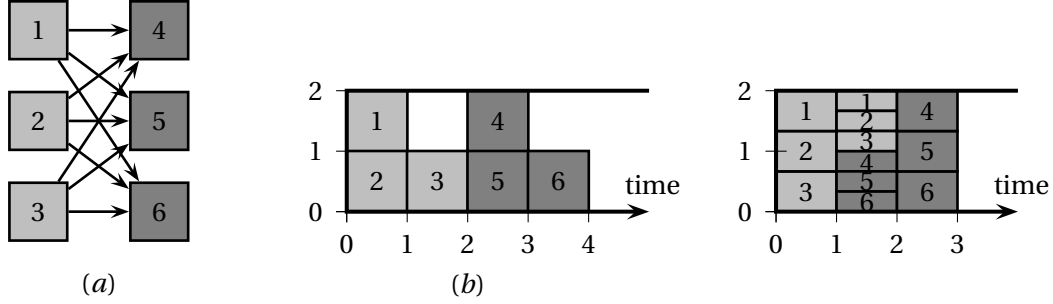


Figure 1: Instance for $P2 \mid \text{prec}, p_j = 1 \mid C_{\max}$ with integrality gap in (a). (b) shows optimum schedule with makespan 4. (c) shows a feasible fractional solution for $T = 3$ where each of the jobs $j \in \{1, 2, 3\}$ has $x_{j1} = \frac{2}{3}$ and $x_{j2} = \frac{1}{3}$. This solution is feasible as such j is scheduled in $[0, 1]$ to an extent of $\frac{2}{3}$ and a dependent job $j' \in \{4, 5, 6\}$ is scheduled in $[0, 2]$ to at most $\frac{1}{3}$.

We claim that in contrast, if there is a Lasserre solution $y \in \text{LAS}_1(K(T))$, then there is also a feasible integral schedule $\sigma : J \rightarrow [T]$. The schedule is obtained by scheduling the jobs in order of their *fractional completion time*, which for a job j is defined as $C_j^* := \max\{t \mid y_{\{(j,t)\}}^* > 0\}$. To keep the notation simple, let us sort the jobs so that $C_1^* \leq C_2^* \leq \dots \leq C_n^*$. Next, let σ be the *list schedule* according to the ordering $1, \dots, n$. In other words, we go through the time slots, starting at the beginning and at each slot, we consider the set of jobs that are unprocessed and have all their predecessors already finished, and then pick the job with smallest index among those. We denote σ_j as the slot in which we process job j . For the 2nd slot at time t , it may happen that there is no such job available as all remaining jobs depend on the job that is processed in slot 1 in parallel. Moreover, it may happen that we process job $j + 1$ before job j as job j was dependent on other jobs that were not yet finished.

Let us first justify why the fractional completion time is a meaningful sorting criterion. Consider dependent jobs $i < j$. Then we claim that $C_i^* \leq C_j^* - 1$. Suppose for the sake of contradiction that $C_i^* \geq C_j^*$. Then we can *induce* on $x_{i,C_i^*} = 1$. In other words, we can extract from y a Lasserre solution $\tilde{y} \in \text{LAS}_0(K(T))$ with $\tilde{y}_{i,C_i^*} = 1$. But the way how the inducing operation is defined, the support of \tilde{y} is a subset of the support of y , that means $y_{jt} = 0 \Rightarrow \tilde{y}_{jt} = 0$. In particular $\tilde{y}_{jt} = 0$ for all $t > C_j^*$. So \tilde{y} schedules j fractionally in the interval $[0, C_j^*] \subseteq [0, C_i^*]$, so there must be a time $t^* \leq C_i^*$ with $\tilde{y}_{j,t^*} > 0$. But in particular the LP contains inequalities implying $\tilde{y}_{i,C_i^*} + \tilde{y}_{j,t^*} \leq 1$. This is a contradiction as \tilde{y} still gives a feasible LP solution.

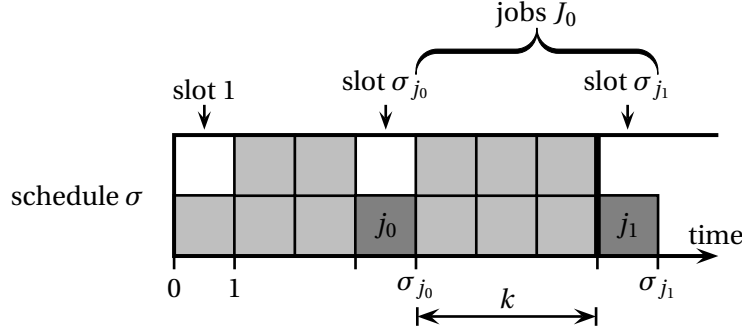
It remains to show the following lemma which then immediately implies that all jobs are finished by time T .

Lemma 10. *For any job $j \in J$ we have $\sigma_j \leq C_j^*$.*

Proof. Let us consider the job with the lowest index that doesn't satisfy the claim. Say

this is job j_1 . Let $j_0 \in \{1, \dots, j_1 - 1\}$ be the last job that was scheduled without any other job in $\{1, \dots, j_1\}$ in parallel⁴. In other words, the other slot at time σ_{j_0} was either empty or occupied with a job $j > j_1$.

Let $J_0 := \{j \in J \mid j \leq j_1 \text{ and } \sigma_j > \sigma_{j_0}\}$. First of all, all jobs in J_0 must be dependent on j_0 , i.e. $j_0 < j$ for $j \in J_0$, as otherwise we would have processed them at time σ_{j_0} (recall that at that time we had a slot that was either empty or occupied by a lower priority job). Moreover, by the choice of j_0 , the complete interval $[\sigma_{j_0}, \sigma_{j_1} - 1]$ of length $k := \sigma_{j_1} - 1 - \sigma_{j_0}$ is fully busy with $2k$ jobs from J_0 . But also the late job j_1 belongs to J_0 , thus $|J_0| > 2k$. By assumption $C_j^* \leq \sigma_{j_1} - 1$ for all $j \in J_0$.



Next, we induce on $x_{j_0, C_{j_0}^*} = 1$, that means we obtain a Lasserre solution $\tilde{y} \in \text{LAS}_0(K(T))$ with $\tilde{y}_{j_0, C_{j_0}^*} = 1$. Now, having this variable being 1 forces the fractional solution to schedule all the dependent jobs in J_0 later than $C_{j_0}^* \geq \sigma_{j_0}$ (by minimality of j_1). And again the support of \tilde{y} is a subset of the support of y . In other words, the fractional schedule \tilde{y} processes all $> 2k$ jobs in J_0 in the interval $[\sigma_{j_0}, \sigma_{j_1} - 1]$ that has only $2k$ slots. This violates the 2nd LP constraint. \square

We want to remark that one could replace the 3rd constraint

$$\sum_{t' \leq t} x_{it'} \geq \sum_{t' \leq t+1} x_{jt'} \quad \forall i < j \quad \forall t \in [T] \quad (5)$$

in K by the weaker one $x_{it} \leq x_{jt'}$ for $i < j$ and $t \geq t'$. If we call the weaker relaxation $K' \supseteq K$, then it is not difficult to argue that each $y \in \text{LAS}_2(K')$ again satisfies (5). To see this, fix a pair of dependent jobs $i < j$. Then via the Decomposition Theorem, we can write y as convex combination of solutions $\tilde{y} \in \text{LAS}_0(K)$ that are integral on all variables $U := \{x_{it}, x_{jt} \mid t \in [T]\}$, as never more than two of those variables can be 1 in any fractional solution. For each such solution \tilde{y} that is integral on U , the constraint (5) is satisfied. As (5) is a linear constraint, this must be the case as well for the convex combination y .

Whether the case of a constant number $m \geq 3$ of machines is solvable in polynomial time is wide open problem. Even a PTAS is not known, but it seems plausible that the Lasserre hierarchy might provide one.

⁴One might argue that it is possible that each time unit in $[0, \sigma_{j_1} - 1]$ both slots might be occupied by jobs $1, \dots, j_1$. This can be resolved by either introducing a simple extra case or arguing that one could add a dummy job j^* that every other job depends on, which then would be scheduled alone at time 1.

Open problem: Is there a PTAS for $m = 3$ machines based on a $f(\varepsilon)$ -round Lasserre solution for the LP in (4)?

3.2 Set Cover

One of the fundamental optimization problems is SET COVER, where the input consists of a family of sets $\mathcal{S} = \{S_1, \dots, S_m\}$ and each set $S \in \mathcal{S}$ has cost c_S . The goal is to cover the ground set $\{1, \dots, n\}$ at minimum cost. For any constant $\varepsilon > 0$, it is **NP**-hard to find a $(1 - \varepsilon) \ln(n)$ -approximation [DS13] and on the other side, already a simple LP rounding approach gives a $\ln(n) + 1$ approximation [Chv79]. So it does not seem that there is much room for any improvement. But more precisely, what the recent result of Dinur and Steurer provides is that there is a $n^{O(1/\varepsilon)}$ -time reduction from SAT to a $(1 - \varepsilon) \ln(n)$ gap instance of SETCOVER. This hardness result does not rule out a *subexponential time algorithm*, which gives a $(1 - \varepsilon) \ln(n)$ approximation say in time $2^{O(n^\varepsilon)}$. In fact, exactly this is the case as we demonstrate here, following [CFG12].

The key trick is that SET COVER can be better approximated if all sets are small. Recall that $H(k) := \sum_{i=1}^k \frac{1}{i} \leq \ln(k) + 1$ is the k th harmonic number. The following theorem is due to Chvátal [Chv79]:

Theorem 11. *Given any solution x to the linear system $\sum_{S:i \in S} x_S \geq 1 \forall i \in [n]$ $x_S \geq 0 \forall S \in \mathcal{S}$, the greedy algorithm provides a solution of cost at most $H(k) \cdot \sum_{S \in \mathcal{S}} c_S x_S$ where $k := \max\{|S_i| \mid x_i > 0\}$ is the size of the largest set used.*

Now, suppose for the sake of simplicity that we know the value of OPT (which can be done by slightly rounding the costs c_i and guessing). Then we choose the convex relaxation K as those fractional SET COVER solutions that cost at most OPT :

$$K := \left\{ x \in \mathbb{R}^m \mid \sum_{i:j \in S_i} x_i \geq 1 \forall j \in [n]; \sum_{i=1}^m c_i x_i \leq OPT; x_i \geq 0 \forall i \in [m] \right\}$$

Theorem 12. *Fix a $0 < \varepsilon < 1$. One can find a SET COVER solution of cost $((1 - \varepsilon) \ln(n) + O(1)) \cdot OPT$ in time $m^{O(n^\varepsilon)}$.*

Proof. We compute a solution $y \in \text{LAS}_{n^\varepsilon}(K)$ in time $m^{O(n^\varepsilon)}$. We pick the largest set S_i with $y_i > 0$ and after reindexing we assume this is S_1 and replace y with a solution $y^{(1)} \in \text{LAS}_{n^\varepsilon - 1}(K)$ by inducing on $x_i = 1$. More generally, after $k \in \{0, \dots, n^\varepsilon - 1\}$ steps, we have a solution $y^{(k)}$ and have already induced on taking S_1, \dots, S_k . Next, we pick the set S with $0 < y_S^{(k)} < 1$ that covers the most elements in $[n] \setminus (S_1 \cup \dots \cup S_k)$. We stop the procedure after having induced on n^ε many variables⁵ and let $y' := y^{(n^\varepsilon)}$. It is clear that we have paid at most OPT for the sets $S_1, \dots, S_{n^\varepsilon}$ so far. Let $X' := X \setminus (S_1 \cup \dots \cup S_{n^\varepsilon})$ be the ground set of not yet covered elements. Consider the set system $\mathcal{S}' := \{S \cap X' \mid S \in \mathcal{S} : y'_S > 0\}$ induced

⁵If we cannot find a fractional variable anymore, then we have a feasible solution of cost OPT already.

by X' . We claim that the remaining sets have size $|S| \leq n^{1-\varepsilon}$ for all $S \in \mathcal{S}'$. If some set S would have more elements than in any single conditioning step, we must have covered more than $n^{1-\varepsilon}$ many elements, which means that in total more than $n^\varepsilon \cdot n^{1-\varepsilon} = n$ would have been covered, which is impossible.

Note that y' defines a feasible LP solution for \mathcal{S}' of cost at most OPT . Thus covering the remaining elements costs at most $(\ln(n^{1-\varepsilon}) + 1)OPT$ using Theorem 11. This shows the claim. \square

3.3 Matching

In the following let $G = (V, E)$ be an undirected graph. A *matching* is a set $M \subseteq E$ of non-incident edges. In other words, the set of all matchings is the set of all integral solutions to

$$K = \left\{ x \in \mathbb{R}_{\geq 0}^E \mid \sum_{e \in \delta(v)} x_e \leq 1 \quad \forall v \in V \right\}$$

In contrast to our other applications, a maximum matching can be found in polynomial time. On the other hand, the straightforward relaxation K is not integral. It was shown by [Edm65] that one can obtain $\text{conv}(K \cap \{0, 1\})$ by adding the so-called *Blossom-inequalities* saying that in any induced subgraph with $2k + 1$ nodes, a matching can only pick at most k edges:

$$\sum_{e \in E \cap (U \times U)} x_e \leq \frac{|U| - 1}{2} \quad \forall U \subseteq V : |U| \text{ odd}$$

Note that separation over these exponentially many inequalities is doable in polynomial time, but still highly non-trivial. So it is still interesting how the Lasserre SDP behaves on them. In fact, the integrality gap for any cost function $c : E \rightarrow \mathbb{R}$ is bounded by $1 + \frac{1}{2t}$, or in a more polyhedral form:

Lemma 13. *One has $\text{LAS}_t^{\text{proj}}(K) \subseteq (1 + \frac{1}{2t}) \cdot \text{conv}(K \cap \{0, 1\}^n)$.*

Proof. Let $y \in \text{LAS}_t(K)$. It suffices to show that for any $k \in \mathbb{N}$ and $|U| = 2k + 1$ one has $\sum_{e \in E \cap (U \times U)} y_e \leq (1 + \frac{1}{2t}) \cdot k$. First of all, the degree constraints imply that $\sum_{e \in E \cap (U \times U)} y_e \leq k + \frac{1}{2}$ so there is nothing to show for large U 's with $k > t$. So assume $k \leq t$. Note that for $I \subseteq E(U)$ with $|I| > k$, one has $y_I = 0$. Thus we can write y as convex combination of solutions $z \in \text{LAS}_0(K)$ with $z_e \in \{0, 1\}$ for all $e \in E(U)$. Note that we want to show that a *linear constraint* (namely $\sum_{e \in E(U)} y_e \leq (1 + \frac{1}{2t}) \cdot k$) is satisfied and for this it suffices to prove that the same linear constraint holds for each of the vectors z used in that convex combination. But $\{e \in E(U) \mid z_e = 1\}$ is a matching, thus $\sum_{e \in E(U)} z_e \leq k$ holds and we are done. \square

Note that without the decomposition theorem we still could have obtained a bound of $1 + O(\frac{1}{\sqrt{t}})$.

3.4 MaxCut

The problem that brought the breakthrough by Goemans and Williamson [GW95] for SDP based approximation algorithms is the MAXCUT problem, so it is worth studying how the Lasserre hierarchy behaves for it. Let $G = (V, E)$ be an undirected graph with positive weights $c : E \rightarrow \mathbb{R}_{\geq 0}$, then the goal is to find a cut $S \subseteq V$ maximizing the weight $\sum_{e \in \delta(S)} c_e$ of edges crossing the cut. In other words, we are looking at integral solutions solving the following optimization problem

$$\max \left\{ \sum_{e \in E} c_e z_e \mid \max\{x_i - x_j, x_j - x_i\} \leq z_{ij} \leq \min\{x_i + x_j, 2 - x_i - x_j\} \quad \forall (i, j) \in E \right\}$$

where x_i is the decision variable telling whether $i \in S$ and z_{ij} tells whether the edge (i, j) lies in the cut. It is not difficult to check that if all variables are integral, $z_{ij} = |x_i - x_j|$. Let K be the polytope of (x, z) satisfying the linear constraints.

Note that K itself is a quite useless relaxation. For any graph we can choose $x_i := \frac{1}{2}$ and $z_{ij} := 1$ and would obtain a feasible fractional solution with value $|E|$. On the other hand for a complete graph with unit weights, no integral solution is better than $\frac{1}{2}|E|$, showing an integrality gap of 2. In fact, no linear program of polynomial size is known that does better. So the magic lies in the SDP constraints.

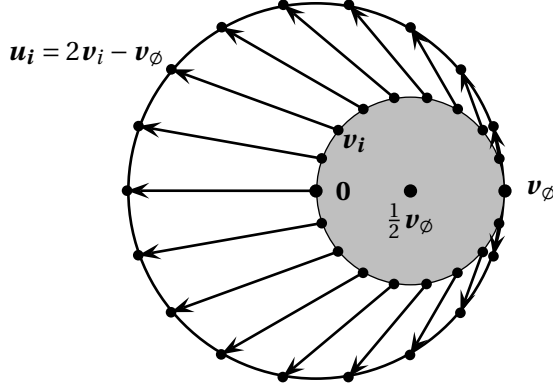
Let $y \in \text{LAS}_3(K)$ be a 3-round Lasserre lifting. As the original LP has two “sorts” of variables, namely x_i and z_e , we will write x_i and $x_{\{i,j\}}$ for the variables $\{x_i\}$ and $\{x_i, x_j\}$ in y (we will never need to consider joint variables that mix x and z variables). First let us understand, how the variables are related.

Lemma 14. *For any edge $e = (i, j) \in E$ we have $z_{ij} = x_i + x_j - 2x_{\{i,j\}}$.*

Proof. We can write y as convex combination of solutions $\tilde{y} \in \text{LAS}_0(K)$ that are integral on $\{x_i, x_j, z_e\}$. As $z_e = x_i + x_j - 2x_{\{i,j\}}$ is a linear constraint, it suffices to that it holds for each \tilde{y} , which is clearly the case (here we can use that $x_{\{i,j\}} = x_i \cdot x_j \in \{0, 1\}$ whenever $x_i, x_j \in \{0, 1\}$). \square

The seminal Goemans Williamson algorithm obtains an approximate maximum cut by first solving a vector SDP then choosing a random hyper plane and declaring all nodes whose vectors are on one side of the hyperplane as cut S . We know by Lemma 7 that there are vectors \mathbf{v}_i such that $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = x_{\{i,j\}}$ for all $i, j \in V$. So the obvious idea would be to perform the hyperplane rounding with these \mathbf{v}_i 's. But this is actually not a good idea. Because the vectors have angles between 0° and 90° , they are all on “one side” from the origin. It would be smarter to perform a *vector transformation* such that the angles between vectors are in the range of 0° to 180° .

In fact, if we choose $\mathbf{u}_i := 2\mathbf{v}_i - \mathbf{v}_\emptyset$ then the \mathbf{u}_i 's are unit vectors. Moreover for an edge $e = (i, j)$ with $z_e = 0$ the “strong will” of the Lasserre solution is to have i and j on different sides of the cut and indeed the vectors \mathbf{u}_i and \mathbf{u}_j are placed antipodal.



Let us show these claims formally.

Lemma 15. *The vectors $\mathbf{u}_i := 2\mathbf{v}_i - \mathbf{v}_\phi$ form a solution to the Goemans-Williamson SDP*

$$\max \left\{ \sum_{(i,j) \in E} c_{ij} \cdot \frac{1 - \mathbf{u}_i \mathbf{u}_j}{2} \mid \|\mathbf{u}_i\|_2 = 1 \ \forall i \in V \right\}$$

and $\frac{1 - \mathbf{u}_i \mathbf{u}_j}{2} = z_{ij}$.

Proof. First of all, the \mathbf{u}_i 's are indeed unit vectors as $\|\mathbf{u}_i\|_2^2 = \|2\mathbf{v}_i - \mathbf{v}_\phi\|_2^2 = 4\|\mathbf{v}_i\|_2^2 - 2 \cdot 2\mathbf{v}_i \mathbf{v}_\phi + \|\mathbf{v}_\phi\|_2^2 = 1$. We verify that

$$\mathbf{u}_i \mathbf{u}_j = (2\mathbf{v}_i - \mathbf{v}_\phi)(2\mathbf{v}_j - \mathbf{v}_\phi) = 4\mathbf{v}_i \mathbf{v}_j - \mathbf{v}_\phi \mathbf{v}_i - \mathbf{v}_\phi \mathbf{v}_j + \|\mathbf{v}_\phi\|_2^2 = 1 - 2z_{ij}$$

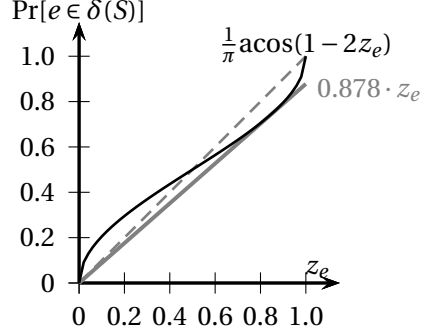
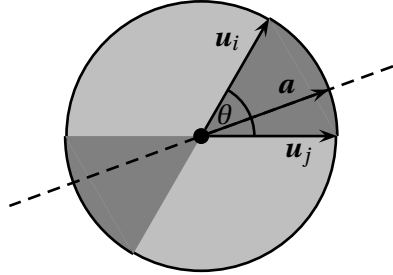
and rearranging yields the claim $z_{ij} = \frac{1 - \mathbf{u}_i \mathbf{u}_j}{2}$. \square

For the rounding, we will need a random vector. First, recall that a *1-dimensional Gaussian* with mean 0 and variance 1 is a continuous random variable $g_i \in \mathbb{R}$ with density function $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. We can use this to define an *m-dimensional Gaussian* by choosing independently 1-dimensional Gaussians g_1, \dots, g_m and put them together in a vector $\mathbf{g} = (g_1, \dots, g_m)$. The nice property of Gaussians is that they are *rotationally symmetric*. This means, we could have picked any other orthonormal basis $\mathbf{b}_1, \dots, \mathbf{b}_m$, then the distribution of $\mathbf{g} = \sum_{i=1}^m \mathbf{b}_i g_i$ would not have changed. In particular, for any unit vector \mathbf{v} , the scalar product $\langle \mathbf{g}, \mathbf{v} \rangle$ is again distribution like a 1-dimensional Gaussian. In the following let $SDP = \sum_{e \in E} c_e z_e$ be the value of the Lasserre SDP.

Lemma 16. *Take a random Gaussian $\mathbf{g} \sim N^m(0, 1)$ and define $S := \{i \in V \mid \mathbf{g} \mathbf{u}_i \geq 0\}$. Then $\Pr[e \in \delta(S)] \geq 0.87 \cdot z_e$. In particular $\mathbb{E}[\sum_{e \in \delta(S)} c_e] \geq 0.87 \cdot SDP$.*

Proof. Consider an edge $(i, j) \in E$ and the 2-dimensional plane containing \mathbf{u}_i and \mathbf{u}_j and denote the angle between \mathbf{u}_i and \mathbf{u}_j by θ . The projection of \mathbf{g} into this plane is just a 2-dimensional random Gaussian inducing a hyperplane through the origin with normal vector \mathbf{g} . Alternatively we can imagine that this hyperplane is just a line spanned by a

random unit vector \mathbf{a} and $(i, j) \in \delta(S)$ if and only if the random direction \mathbf{a} separates \mathbf{u}_i and \mathbf{u}_j .



The probability for this is exactly $\frac{\theta}{\pi}$. It remains to express the angle as $\theta = \text{acos}(\mathbf{u}_i \mathbf{u}_j) = \text{acos}(1 - 2z_{ij})$, hence $\Pr[(i, j) \in \delta(S)] = \frac{1}{\pi} \text{acos}(1 - 2z_{ij})$. Then the expected integrality gap is

$$\frac{\Pr[(i, j) \in \delta(S)]}{z_{ij}} \geq \min_{0 \leq z_{ij} \leq 1} \frac{1}{\pi} \frac{\text{acos}(1 - 2z_{ij})}{z_{ij}} \approx 0.878$$

□

3.5 Global Correlation Rounding

In the following we want to show a PTAS for MAXCUT in dense graphs, based on the Lasserre hierarchy. More precisely, in any undirected graph with at least εn^2 edges, we will find a cut of value at least $(1 - \varepsilon) \cdot \text{OPT}$. Such a result was first obtained using a combinatorial algorithm chooses randomly a constant number of nodes and guesses their partition in the optimum [AKK95, dIV96]. Also for hierarchies such a result is known due to de la Vega and Kenyon-Mathieu [dIVKm07].

However, we want to use the opportunity to demonstrate the important technique of *Global Correlation Rounding* to prove a PTAS. This technique was introduced by Barak, Raghavendra and Steurer [BRS11] and independently by Guruswami and Sinop [GS11] to solve constraint satisfaction problems like MAXCUT or UNIQUEGAMES via SDP hierarchies. Here the performance depends crucially on the Eigenvalue spectrum of the underlying constraint graph. In our exposition, we use parts of the analysis of [AG11].

For the moment, let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ an arbitrary polytope and $y \in \text{LAS}_t(K)$ with $t \geq 3$. Consider two variables $i, j \in [n]$ and let $(X_i, X_j) \sim \mathcal{D}_{\{i, j\}}(y)$ be the corresponding jointly distributed 0/1 random variables, see Lemma 5⁶. The following fact is well known: conditioning on X_j ($X_j = 1$ with probability y_j and $X_j = 0$ otherwise) will decrease the variance of X_i (or leave it invariant) and the decrease is growing with the correlation of X_i and X_j .

This leads to the following approach for a rounding algorithm: first we induce on a constant number of variables so that most pairs of variables i, j will later be uncorre-

⁶Recall that this means $\Pr[X_i = 1] = y_i$, $\Pr[X_j = 1] = y_j$ and $\Pr[X_i = X_j = 1] = y_{\{i, j\}}$

lated. Then a simple randomized rounding of an uncorrelated instance can provide a solution which is almost as good as the SDP relaxation.

Let us make this more formal. We want to remind the reader that the *variance* of a 0/1 variable is $\text{Var}[X_j] = \mathbb{E}[X_j^2] - \mathbb{E}[X_j]^2 = y_j(1 - y_j)$ and their *covariance* is $\text{Cov}[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] = y_{ij} - y_i y_j$. The *law of total variance* says that

$$\mathbb{E}_{X_j} [\text{Var}[X_i | X_j]] = \text{Var}[X_i] - \text{Var}_{X_j}[\mathbb{E}[X_i | X_j]]$$

We want to examine the quantity $\text{Var}_{X_j}[\mathbb{E}[X_i | X_j]]$ a little closer. First of all $\mathbb{E}[X_i | X_j = 1] = \frac{y_{i,j}}{y_j}$ and $\mathbb{E}[X_i | X_j = 0] = \frac{y_i - y_{i,j}}{1 - y_j}$ and $\mathbb{E}_{X_j}[\mathbb{E}[X_i | X_j]] = \mathbb{E}[X_i] = y_i$, thus

$$\begin{aligned} \text{Var}_{X_j}[\mathbb{E}[X_i | X_j]] &= \underbrace{y_j \cdot \left(\frac{y_{i,j}}{y_j}\right)^2 + (1 - y_j) \cdot \left(\frac{y_i - y_{i,j}}{1 - y_j}\right)^2}_{\mathbb{E}[\dots]^2} - \underbrace{y_i^2}_{=\mathbb{E}[\dots]^2} \\ &= \frac{(y_i y_j - y_{i,j})^2}{y_j(1 - y_j)} = \frac{\text{Cov}[X_i, X_j]^2}{\text{Var}[X_j]} \geq 4 \cdot \text{Cov}[X_i, X_j]^2 \end{aligned} \quad (6)$$

(7)

This motivates two important definitions. Let $R \subseteq [n]$ be a subset of **relevant variables** for which we would like to have a low correlation. We define the *variance of y w.r.t. R* as $V_R(y) := \sum_{i \in R} y_i(1 - y_i)$. Moreover, we define the *correlation of y w.r.t. variables R* as

$$C_R(y) = \sum_{i \in R} \sum_{j \in R} (y_i y_j - y_{i,j})^2$$

We obtain the following general theorem.

Theorem 17 (Global Correlation Theorem). *Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ be any polytope, $y \in \text{LAS}_t(K)$ with $t \geq \frac{1}{\varepsilon^3} + 2$ and $R \subseteq [n]$. Then one can induce on at most $\frac{1}{\varepsilon^3}$ variables in R to obtain $y' \in \text{LAS}_{t-1/\varepsilon^3}(K)$ such that $C_R(y') \leq \frac{\varepsilon^3}{4} |R|^2$ and in particular $\Pr_{i,j \in R}[|y_i y_j - y_{i,j}| \geq \varepsilon] \leq \varepsilon$.*

Proof. We begin with studying the effect of a single random conditioning. Choose an index $j \in R$ uniform at random and choose $a \in \{0, 1\}$ so that $\Pr[a = 1] = y_j$. Let $y' \in \text{LAS}_{t-1}(K)$ be the solution y induced on $y'_j = a$. Then the variance of y' behaves as

$$\mathbb{E}_{j,a} [V_R(y')] = \mathbb{E}_{j,a} \left[\sum_{i \in R} y'_i(1 - y'_i) \right] \stackrel{\text{Eq (6)}}{\leq} \frac{1}{|R|} \sum_{j \in R} \sum_{i \in R} \left(y_i(1 - y_i) - 4(y_i y_j - y_{i,j})^2 \right) = V_R(y) - \frac{4C_R(y)}{|R|}$$

Of course, there exists at least one choice of j and a such that indeed $V_R(y') \leq V_R(y) - \frac{4}{|R|} C_R(y)$. We run the following procedure (starting with $y' := y$).

(1) REPEAT

- (2) IF $C_R(y') \leq \frac{\varepsilon^3}{4} \cdot |R|^2$ THEN return the current y'
- (3) Find that $j \in R$ and $a \in \{0, 1\}$ maximizing the decrease in variance and induce on it.

It remains to show that the procedure stops after at most $\frac{1}{\varepsilon^3}$ iterations. Initially $V_R(y) = \sum_{i \in R} y_i(1 - y_i) \leq \frac{|R|}{4}$ and in each iteration the term $V_R(y')$ decreases by $\varepsilon^3|R|$. Thus we stop after at most $\frac{1}{\varepsilon^3}$ many iterations.

Finally, if and the end more than $\varepsilon|R|^2$ pairs i, j would have $|y_i y_j - y_{ij}| \geq \varepsilon$ then $C_R(y') \geq \varepsilon|R|^2 \cdot \varepsilon^2$ which is a contradiction. \square

We want to emphasize that so far we have not used anything problem specific.

3.5.1 Application to MaxCut in dense graphs

Now we are ready to demonstrate why uncorrelated solutions can be so useful. Let $G = (V, E)$ be a MAXCUT instance with $|E| \geq \varepsilon n^2$ many unweighted edges. Let $K(\alpha)$ be the convex set of solutions (x, z) satisfying

$$\begin{aligned} \max\{x_i - x_j, x_j - x_i\} \leq z_{ij} &\leq \min\{x_i + x_j, 2 - x_i - x_j\} \quad \forall (i, j) \in E \\ \sum_{(i,j) \in E} z_{ij} &= \alpha \end{aligned}$$

where $\alpha \in [0, |E|]$ is the value of the objective function. Now the PTAS is easy to obtain:

Lemma 18. *For $\varepsilon > 0$ let $G = (V, E)$ be a MAXCUT instance with $|E| \geq \varepsilon n^2$ and $\tilde{y} \in \text{LAS}_{(1/\varepsilon)^4+3}(K(\alpha))$. Then there is a cut $S \subseteq V$ with $|\delta(S)| \geq (1 - O(\varepsilon)) \cdot \alpha$.*

Proof. Using Theorem 17 we can extract an uncorrelated solution $y \in \text{LAS}_3(K)$ from \tilde{y} such that $C_R(y) \leq \varepsilon^4 n^2$ (we choose R as the set of x -variables). In particular, we have $\Pr_{i,j \in V}[|x_i x_j - x_{\{i,j\}}| \geq \varepsilon] \leq \varepsilon^2$. We call an edge $(i, j) \in E$ *bad* if either $|x_i x_j - x_{ij}| > \varepsilon^2$ or $x_{ij} \leq \varepsilon$ (or both). Other edges are called *good*. Note that the contribution of bad edges to the SDP value is at most $2\varepsilon^2 n^2 \leq 4\varepsilon \cdot \alpha$ (here we can assume that $\alpha \geq \frac{|E|}{2}$), thus it suffices to consider good edges in our rounding scheme. We determine the cut S as follows: for every node $i \in V$ we flip an independent coin and put it into S with probability x_i . Consider a good edge (i, j) and recall that its contribution to the SDP value is $z_{ij} = x_i + x_j - 2x_{\{i,j\}} \geq \varepsilon$, see Section 3.4. The chance that this good edge is included in our cut is

$$\Pr[(i, j) \in \delta(S)] = x_i(1 - x_j) + (1 - x_i)x_j = x_i + x_j - 2x_{\{i,j\}} \stackrel{(i,j) \text{ good}}{\geq} \underbrace{x_i + x_j - 2(x_{ij} + \underbrace{\varepsilon^2}_{\leq \varepsilon z_{ij}})}_{=z_{ij}} \geq (1 - 2\varepsilon) \cdot z_{ij}$$

and the claim follows. \square

In Section 3.4, we somewhat got the impression that hyperplane rounding is “better” for MAXCUT, as independent rounding. But this is not the case here anymore. If we have

a completely uncorrelated edge $(i, j) \in E$, i.e. $x_{ij} = x_i \cdot x_j$ and for example $x_i = 0.1$ and $x_j = 0.9$, then the contribution to the SDP objective function is $z_{ij} = 0.82$, while a cut S provided by the hyperplane rounding yields $\Pr[(i, j) \in \delta(S)] = \frac{1}{\pi} \arccos(1 - 2z_{ij}) \approx 0.72 \approx 0.88 \cdot z_{ij}$.

References

- [ACC06] S. Arora, M. Charikar, and E. Chlamtac. New approximation guarantee for chromatic number. In *STOC*, pages 215–224, 2006.
- [AG11] S. Arora and R. Ge. New tools for graph coloring. In *APPROX-RANDOM*, pages 1–12, 2011.
- [AKK95] Sanjeev Arora, David R. Karger, and Marek Karpinski. Polynomial time approximation schemes for dense instances of np -hard problems. In *STOC*, pages 284–293, 1995.
- [ARV04] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *STOC*, pages 222–231, 2004.
- [BCC93] E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Math. Program.*, 58:295–324, 1993.
- [BCG09] M. Bateni, M. Charikar, and V. Guruswami. Maxmin allocation via degree lower-bounded arborescences. In *In STOC '09: Proceedings of the 41st annual ACM Symposium on Theory of computing*, pages 543–552, 2009.
- [BRS11] B. Barak, P. Raghavendra, and D. Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS*, volume abs/1104.4680, 2011.
- [CFG12] Eden Chlamtac, Zac Friggstad, and Konstantinos Georgiou. Understanding set cover: Sub-exponential time approximations and lift-and-project methods. *CoRR*, abs/1204.5489, 2012.
- [Chl07] E. Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *FOCS*, pages 691–701, 2007.
- [Chv79] V. Chvátal. A greedy heuristic for the set-covering problem. *Math. Oper. Res.*, 4(3):233–235, 1979.
- [CS08] E. Chlamtac and G. Singh. Improved approximation guarantees through higher levels of SDP hierarchies. In *APPROX-RANDOM*, pages 49–62, 2008.
- [CT11] E. Chlamtac and M. Tulsiani. Convex relaxations and integrality gaps. In *Handbook on Semidefinite, Cone and Polynomial Optimization*, 2011.

- [dlV96] Wenceslas Fernandez de la Vega. Max-cut has a randomized approximation scheme in dense graphs. *Random Struct. Algorithms*, 8(3):187–198, 1996.
- [dlVKm07] Wenceslas Fernandez de la Vega and Claire Kenyon-mathieu. Linear programming relaxations of maxcut. In *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms*, pages 53–61, 2007.
- [DS13] Irit Dinur and David Steurer. Analytical approach to parallel repetition. *CoRR*, abs/1305.1979, 2013.
- [Edm65] Jack Edmonds. Maximum matching and a polyhedron with 0,1-vertices. *J. Res. Nat. Bur. Standards Sect. B*, 69B:125–130, 1965.
- [GS11] V. Guruswami and A. Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with psd objectives. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:66, 2011.
- [GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995.
- [JG72] Edward G. Coffman Jr. and Ronald L. Graham. Optimal scheduling for two-processor systems. *Acta Inf.*, 1:200–213, 1972.
- [KMN11] A. Karlin, C. Mathieu, and C. Nguyen. Integrality gaps of linear and semi-definite programming relaxations for knapsack. In *IPCO*, pages 301–314, 2011.
- [Las01a] J. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *IPCO*, pages 293–303, 2001.
- [Las01b] J. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.
- [Lau03] M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, 28(3):470–496, 2003.
- [LS91] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.
- [SA90] H. Sherali and W. Adams. A hierarchy of relaxation between the continuous and convex hull representations. *SIAM J. Discret. Math.*, 3:411–430, May 1990.
- [Sve11] O. Svensson. Personal communication, 2011.