# Directed Steiner Tree and the Lasserre Hierarchy

Thomas Rothvoß[*]

M.I.T.
rothvoss@math.mit.edu

June 13, 2012

### Abstract

The goal for the DIRECTED STEINER TREE problem is to find a minimum cost tree in a directed graph $G = (V, E)$ that connects all terminals $X$ to a given root $r$. It is well known that modulo a logarithmic factor it suffices to consider acyclic graphs where the nodes are arranged in $\ell \leq \log|X|$ levels. Unfortunately the natural LP formulation has a $\Omega(\sqrt{|X|})$ integrality gap already for 5 levels. We show that for every $\ell$, the $O(\ell)$-*round Lasserre Strengthening* of this LP has integrality gap $O(\ell \log|X|)$. This provides a polynomial time $|X|^{\varepsilon}$-approximation and a $O(\log^3|X|)$ approximation in $O(n^{\log|X|})$ time, matching the best known approximation guarantee obtained by a greedy algorithm of Charikar et al.

## 1 Introduction

Most optimization problems that appear in combinatorial optimization can be written as an integer linear problem, say in the form $\min\{c^T x \mid Ax \geq b; \ x \in \{0, 1\}^n\}$, where the system $Ax \geq b$ represents the problem structure. Since $c$ is linear, this is optimizing over the convex hull $K_I := \text{conv}(K \cap \{0, 1\}^n)$, where $K := \{x \in \mathbb{R}^n \mid Ax \geq b\}$ denotes a polyhedron.

Such optimization problems are **NP**-hard in general, thus a standard approach for obtaining approximate solutions is to optimize instead over the relaxation $K$ and then try to extract a close integral solution. This approach yields in many cases solutions whose quality matches the lower bound provided by the PCP Theorem or the Unique Games Conjecture (which is the case e.g. for SET COVER [Lov75, Fei98], VERTEX COVER [KR08] and FACILITY LO-CATION [GK99, Li11]. However, there is a significant number of problems, where the *integrality gap* between $K$ and $K_I$ appears to be far higher than the approximability of the problem, so that a stronger formulation is needed.

At least in the field of approximation algorithm, researchers have so far mostly preferred problem-specific inequalities to lower the integrality gap (a nice example is the $O(1)$-apx for MIN-SUM SET COVER [BGK10]). However there are very general techniques that can be used to strengthen the convex relaxation $K$.

---

Especially in the field of (computational) integer programming, the approach of cutting planes is very popular. In the *Gomory-Chvátal Closure* $CG(K) \subseteq K$ one adds simultaneously cuts $a^T x \leq \lfloor \beta \rfloor$ for all valid inequalities $a^T x \leq \beta$ with $a \in \mathbb{Z}^n$. On the positive side, after at most $O(n^2 \log n)$ iterative applications of the closure operation, one reaches $K_I$ [ES99] (assuming that $K \subseteq [0,1]^n$). But the drawback is that already optimizing over the first closure $CG(K)$ is **coNP**-hard [Eis99]. Singh and Talwar [ST10] studied the effect of Gomory-Chvátal cuts to the integrality gap of hypergraph matchings and other problems .

However, more promising for the sake of approximation algorithms are probably LP/SDP hierarchies like the ones of *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 $K$ (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. We refer to the survey of Laurent [Lau03a] for a detailed comparison.

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 $\gamma n$. After a constant number of rounds of Sherali-Adams, the integrality gap for the matching polytope reduces to $1 + \varepsilon$ [MS09]. The same is true for MAXCUT in dense graphs (i.e. graphs with $\Omega(n^2)$ edges) [dlVKM07]. The Sherali-Adams hierarchy is also 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 underling 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].

Moreover, integrality gap lower bounds exist for various problems [Lau03b, AAT05, Tou06, STT07a, STT07b, GMPT07, Sch08, CMM09, Tul09]. To name only few of these results, even a linear number of Lasserre rounds cannot refute unsatisfiable constraint satisfaction problems [Sch08] and the gap for GRAPH COLORING is still $k$ versus $2^{\Omega(k)}$ [Tul09]. In contrast, the LP-based hierarchies LS and SA cannot even reduce the MAXCUT gap below $2 - \varepsilon$ after $\Omega(n)$ [STT07b] and $n^\delta$ [CMM09] many rounds, respectively. Recall that already a single round of the SDP based hierarchies reduces the gap to 1.13.

In this paper, we apply the Lasserre relaxation to the flow-based linear programming relaxation of DIRECTED STEINER TREE. The input for this problem consists of a directed graph $G = (V, E)$ with edge cost $c : E \to \mathbb{R}_+$, a *root* $r \in V$ and *terminals* $X \subseteq V$. The goal is to
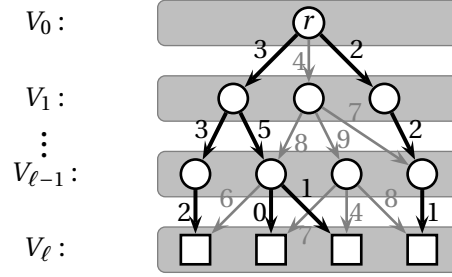
Figure 1: Illustration of layered graph $G$ with $\ell = 3$. Edges are labelled with their cost. Black edges denote an optimum DIRECTED STEINER TREE solution. Rectangles depict terminals.

compute a subset $T \subseteq E$ of edges such that there is an $r$-$s$ path in $T$ for each terminal $s$. Note that the cheapest such set always forms a tree (see Figure 1).

By a straightforward reduction from SET COVER one easily sees that the problem is $\Omega(\log n)$-hard [Fei98] to approximate. Zelikovsky [Zel97] obtained a $|X|^{\varepsilon}$-approximation for every constant $\varepsilon > 0$ using a greedy approach. In fact, he provided the following useful insight:

**Theorem 1** ([Zel97, CZ05])**.** *For every $\ell \geq 1$, there is a tree $T$ (potentially using edges in the metric closure) of cost $c(T) \leq \ell \cdot |X|^{1/\ell} \cdot OPT$ such that every $r$-$s$ path (with $s \in X$) in $T$ contains at most $\ell$ edges.*

In other words, at the cost of a factor[1] $\ell \cdot |X|^{1/\ell}$ in the approximation guarantee, one may assume that the graph is acyclic and the nodes are arranged in $\ell$ *levels*[2] (observe that for $\ell = \log |X|$, one has $\ell \cdot |X|^{1/\ell} = \ell \cdot (2^{\log |X|})^{1/\log |X|} = O(\log |X|)$). Later Charikar, Chekuri, Cheung, Goel, Guha and Li [CCC+99]) gave a $O(\log^3 |X|)$ approximation in time $n^{O(\log |X|)}$ again using a sophisticated greedy algorithm. So far, methods based on linear programming were less successful. In fact, Zosin and Khuller [ZK02] show that the natural flow based LP relaxation has an integrality gap of $\Omega(\sqrt{|X|})$. The importance of DIRECTED STEINER TREE lies in the fact that it generalizes a huge number of problems, e.g. SET COVER, (non-metric, multilevel) FACILITY LOCATION and GROUP STEINER TREE. For the latter problem, the input consists of an *undirected* (weighted) graph $G = (V, E)$, groups $G_1, \ldots, G_k \subseteq V$ of terminals and a root $r \in V$. The goal here is to find a tree $T \subseteq E$ of minimum cost $c(T)$ that connects at least one terminal from every group to the root. The state of the art for GROUP STEINER TREE is still the elegant approach of Garg, Konjevod and Ravi [GKR00]. Using $O(\log n)$-average distortion tree embeddings, one can assume that the input graph itself is a tree. Then [GKR00] solve a flow based LP and provide a rounding scheme, which gives a $O(\log^2 n)$-approximation w.r.t. the optimum solution in the tree graph. Surprisingly, [HKK+03] found a tree instance, which indeed has an integrality gap of $\Omega(\log^2 n)$. Later Halperin and Krauthgamer [HK03] even proved a $\Omega(\log^{2-\varepsilon} n)$ inapproximability for GROUP STEINER TREE (on tree graphs) and in turn also for DIRECTED STEINER TREE.

---

[1] The claim in [Zel97] was initially $|X|^{1/\ell}$, which is incorrect, though it was later on heavily used in the literature. In a later paper, Calinescu and Zelikovsky[CZ05] change the claim to $\ell \cdot |X|^{1/\ell}$. As a consequence, the $O(\log^2 |X|)$-apx in [CCC+99] has to be changed to a $O(\log^3 |X|)$-apx.

[2] This can easily be achieved by taking $\ell + 1$ copies of the node set in the original graph and insert cost-0 edges between copies of the same node.

**Our contribution**

Many researchers have failed in designing stronger LP relaxations for DIRECTED STEINER TREE (see Alon, Moitra and Sudakov [AMS12] for a counterexample to a promising approach). We make partial progress by showing that in an $\ell$-level graph, already $O(\ell)$ rounds of the Lasserre hierarchy drastically reduce the integrality gap of the natural flow-based LP for DIRECTED STEINER TREE from $\Omega(\sqrt{|X|})$ (for $\ell \geq 5$) down to $O(\ell \log |X|)$. This gives an alternative polylogarithmic approximation in quasi-polynomial time (and the first one that is based on convex relaxations).

In this paper, we try to promote the application of hierarchies in approximation algorithms. For this sake, we demonstrate how the Lasserre relaxation can be used as a black box in order to obtain powerful (yet reasonably simple) approximation algorithms.

From a technical view point we adapt the rounding scheme of Garg, Konjevod and Ravi [GKR00]. Though their algorithm and analysis crucially relies on the fact that the input graph itself is a tree, it turns out that one can use instead the values of the auxiliary variables in the $O(\ell)$-round SDP to perform the rounding. Another ingredient for our analysis is the recent *Decomposition Theorem* of Karlin, Mathieu and Nguyen [KMN11] for the Lasserre hierarchy.

## 2  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 mainly follow the survey of Laurent [Lau03a]. Let $\mathscr{P}_t([n]) := \{I \subseteq [n] \mid |I| \leq t\}$ be the set of all index sets of cardinality at most $t$ and let $y \in \mathbb{R}^{\mathscr{P}_{2t+2}([n])}$ be a vector with entries $y_I$ for all $I \subseteq [n]$ with $|I| \leq 2t+2$. Intuitively $y_{\{i\}}$ represents the original variable $x_i$ and the new variables $y_I$ represent $\prod_{i \in I} x_i$. We define the *moment matrix* $M_{t+1}(y) \in \mathbb{R}^{\mathscr{P}_{t+1}([n]) \times \mathscr{P}_{t+1}([n])}$ by

$$(M_{t+1}(y))_{I,J} := y_{I \cup J} \quad \forall |I|, |J| \leq t+1.$$

For a linear constraint $a^T x \geq \beta$ with $a \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$ we define $\binom{a}{\beta} * y$ as the vector $z$ with $z_I := \sum_{i \in [n]} a_i y_{I \cup \{i\}} - \beta y_I$.[3]

**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 $y \in \mathbb{R}^{\mathscr{P}_{2t+2}([n])}$ that satisfy*

$$M_{t+1}(y) \succeq 0; \qquad M_t(\binom{A_\ell}{b_\ell} * y) \succeq 0 \quad \forall \ell \in [m]; \qquad y_\emptyset = 1.$$

*Furthermore, let $\text{LAS}_t^{proj} := \{(y_{\{1\}}, \ldots, y_{\{n\}}) \mid y \in \text{LAS}_t(K)\}$ be the projection on the original variables.*

Intuitively, the PSD-constraint $M_t(\binom{A_\ell}{b_\ell} * y) \succeq 0$ guarantees that $y$ satisfies the $\ell$-th linear constraint, while $M_{t+1}(y) \succeq 0$ takes care that the variables are consistent (e.g. it guarantees that $y_{\{1,2\}} \in [y_{\{1\}} + y_{\{2\}} - 1, \min\{y_{\{1\}}, y_{\{2\}}\}]$). 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.

---

[3]This notation was initially introduced for multivariate degree-one polynomials $g(x) = \sum_I g_I \prod_{i \in I} x_i$ which induce constraints of the form $g(x) \geq 0$. In this general case, one defines $(g * y)_I := \sum_{K \subseteq [n]} g_K \cdot y_{I \cup K}$. Note that for a linear constraint $a^T x \geq \beta$, one has $g_{\{i\}} = a_i, g_\emptyset = -\beta$ and $g_I = 0$ for $|I| > 1$. We stick to this notation to be consistent with the existing literature.

Fortunately, one can use the Lasserre relaxation conveniently as a black-box. We list all properties, that we need for our approximation algorithm:

**Theorem 2.** *Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ and $y \in \mathrm{LAS}_t(K)$. Then the following holds:*

*(a)* $\mathrm{conv}(K \cap \{0,1\}^n) = \mathrm{LAS}_n^{proj}(K) \subseteq \mathrm{LAS}_{n-1}^{proj}(K) \subseteq \ldots \subseteq \mathrm{LAS}_0^{proj}(K) \subseteq K.$

*(b)* *One has* $0 \leq y_I \leq y_J \leq 1$ *for all* $I \supseteq J$ *with* $0 \leq |J| \leq |I| \leq t$.

*(c)* *Let* $I \subseteq [n]$ *with* $|I| \leq t$. *Then* $K \cap \{x \in \mathbb{R}^n \mid x_i = 1\ \forall i \in I\} = \emptyset \implies y_I = 0$.

*(d)* *Let* $I \subseteq [n]$ *with* $|I| \leq t$. *Then* $y \in \mathrm{conv}(\{z \in \mathrm{LAS}_{t-|I|}(K) \mid z_{\{i\}} \in \{0,1\}\ \forall i \in I\})$.[4]

*(e)* *Let* $S \subseteq [n]$ *be a subset of variables such that* $\max\{|I| : I \subseteq S; x \in K; x_i = 1\ \forall i \in I\} \leq k < t$. *Then* $y \in \mathrm{conv}(\{z \in \mathrm{LAS}_{t-k}(K) \mid z_{\{i\}} \in \{0,1\}\ \forall i \in S\})$.

*(f)* *For any* $|I| \leq t$ *one has* $y_I = 1 \Leftrightarrow \bigwedge_{i \in I}(y_{\{i\}} = 1)$.

*(g)* *For* $|I| \leq t$: $(\forall i \in I : y_{\{i\}} \in \{0,1\}) \implies y_I = \prod_{i \in I} y_{\{i\}}$.

*(h)* *Let* $|I|, |J| \leq t$ *and* $y_I = 1$. *Then* $y_{I \cup J} = y_J$.

*Proof.* Proofs of (a),(b),(c) can be found in Laurent [Lau03a]. (e) is the Decomposition Theorem of [KMN11]. (f) and (g) follow easily from (b) and (h). For (h), consider the principal submatrix

$$M = \begin{pmatrix} 1 & 1 & y_J \\ 1 & 1 & y_{I \cup J} \\ y_J & y_{I \cup J} & y_J \end{pmatrix}$$

of $M_{t+1}(y)$ that is induced by indices $\{\emptyset, I, J\}$ (substituting $y_I$ with 1). Then $\det(M) = -(y_J - y_{I \cup J})^2 \geq 0$ implies that $y_J = y_{I \cup J}$. $\qquad\square$

Though all these properties are well known, to be fully self contained, we provide a complete introduction with proofs of the non-trivial statements (a),(d),(e) in the appendix.

Especially (e) is a remarkably strong property that does not hold for the Sherali-Adams or Lovász-Schrijver hierarchy (see [KMN11]). For example, it implies that after $t = O(\frac{1}{\varepsilon})$ rounds, the integrality gap for the KNAPSACK polytope is bounded by $1 + \varepsilon$ (taking $S$ as all items that have profit at least $\varepsilon \cdot OPT$). The same bound holds for the MATCHING polytope $\{x \in \mathbb{R}_+^E \mid x(\delta(v)) \leq 1\ \forall v \in V\}$ (since Property (e) implies all Blossom inequalities up to $2t + 1$ nodes). Another immediate consequence is that the INDEPENDENT SET polytope $\{x \in \mathbb{R}_+^V \mid x_u + x_v \leq 1\ \forall \{u, v\} \in E\}$ describes the integral hull after $\alpha(G)$ rounds of Lasserre (where $\alpha(G)$ is the stable set number of the considered graph).

## 3 The linear program

The natural LP formulation for DIRECTED STEINER TREE sends a unit flow from the root to each terminal $s \in X$ (represented by variables $f_{s,e}$). The amount of capacity that has to be paid on edge $e$ is $y_e = \max\{f_{s,e} \mid s \in X\}$. We abbreviate $\delta^+(v) := \{(v, u) \mid (v, u) \in E\}$ ($\delta^-(v) :=$

---

[4]Formally spoken, vectors in $\mathrm{LAS}_{t-|I|}$ have less dimensions than $y$. Thus it would be more correct to write $z_{|\mathscr{P}_{2(t-|I|)+2}([n])} \in \mathrm{LAS}_{t-|I|}(K)$ where $z_{|\mathscr{P}_{2t+2-2|I|}([n])}$ denotes the restriction of $z$ to all entries $I$ with $|I| \leq 2(t-|I|)+2$.

$\{(u, v) \mid (u, v) \in E\}$, resp.) as the edges outgoing (ingoing, resp.) from $v$ and $y(E') := \sum_{e \in E'} y_e$. The LP is

$$\min \sum_{e \in E} c_e y_e$$

$$\sum_{e \in \delta^+(v)} f_{s,e} - \sum_{e \in \delta^-(v)} f_{s,e} = \begin{cases} 1 & v = r \\ -1 & v = s \\ 0 & \text{otherwise} \end{cases} \quad \forall s \in X \ \forall v \in V$$

$$f_{s,e} \leq y_e \quad \forall s \in X \ \forall e \in E$$

$$y(\delta^-(v)) \leq 1 \quad \forall v \in V$$

$$0 \leq y_e \leq 1 \quad \forall e \in E$$

$$0 \leq f_{s,e} \leq 1 \quad \forall s \in X \ \forall e \in E$$

Note that we have an additional constraint $y(\delta^-(v)) \leq 1$ (i.e. only one ingoing edge for each node) which is going to help us in the analysis. Let $K \subseteq \mathbb{R}^E \times \mathbb{R}^{X \times E}$ be the set of fractional solutions. This LP has an integrality gap[5] of $\Omega(\sqrt{|X|})$ even if the number of layers is 5 [ZK02]. From now on, we make the choice $t := 2\ell$, i.e. we consider the $2\ell$-round Lasserre strengthening of the above LP. Let $\mathcal{V}_t = \{(P, H) \mid P \subseteq E; H \subseteq X \times E; |P| + |H| \leq 2t + 2\}$ be the set of variable indices for the $t$-th level of Lasserre. In other words $\text{LAS}_t(K) \subseteq [0, 1]^{\mathcal{V}_t}$. Let $Y = (Y_{P,H})_{(P,H) \in \mathcal{V}_t} \in \text{LAS}_t(K)$ be an optimum solution for the Lasserre relaxation, which can be computed in time $n^{O(t)}$. We abbreviate $OPT_f := \sum_{e \in E} c_e y_{\{e\}}$ as the objective function value.

We will only address either groups of $y_e$ variables (then we write $y_H := Y_{H,\emptyset}$ for $H \subseteq E$), or we address groups of $f_{s,e}$ variables for the same terminal $s \in X$. Then we write $f_{s,H} := Y_{\emptyset, \{(s,e) \mid e \in H\}}$.

## 4 The rounding algorithm

By Theorem 1, we may assume that the node set is partitioned into *levels* $V_0 = \{r\}, V_1, \ldots, V_{\ell-1}, V_\ell = X$ and all edges are running between consecutive layers (i.e. $E \subseteq \bigcup_{j=1}^{\ell} (V_{j-1} \times V_j)$). See Figure 1 for an illustration. In the following, we will present an adaptation of the [GKR00] rounding scheme to sample a set $T$ of paths from a distribution that depends on $Y$. For this sake, starting at layer 0, we will go through all layers and for each path $P$ (ending in node $u$) that is sampled so far, we will extend it to $P \cup \{(u, v)\}$ with probability $\frac{y_{P \cup \{(u,v)\}}}{y_P}$.

(1) $T := \emptyset$

(2) FOR ALL $e \in \delta^+(r)$ DO

    (3) independently, with prob. $y_{\{e\}}$, add path $\{e\}$ to $T$

(4) FOR $j = 1, \ldots, \ell - 1$ DO

    (5) FOR ALL $u \in V_j$ and all $r$-$u$ paths $P \in T$ DO

        (6) FOR ALL $e \in \delta^+(u)$ DO

---

[5]Unfortunately, the instance has a number of nodes which is exponential in the number of terminals. Of course, the instance of [HKK+03] provides a $\Omega(\log^2 n)$ gap. To the best of our knowledge, there is no known instance with a $\omega(\log^2 n)$ integrality gap.

(7) independently with prob. $\frac{y_{P \cup \{e\}}}{y_P}$ add $P \cup \{e\}$ to $T$

(8) return $E(T)$.

With $V(P)$ we denote the set of vertices on path $P$. Furthermore, let $E(T) := \bigcup_{P \in T} P$ be the set of all edges on any path of $T$. Also let $V(T) := \bigcup_{P \in T} V(P)$. Note that we did not remove partial paths (i.e. paths from $r$ to some layer $j < \ell$), which will turn out to be convenient later.

## 5 The analysis

The analysis consists of two parts:

(i) We show that for each edge $e$ the probability to be included is $\Pr[e \in E(T)] \leq y_{\{e\}}$.

(ii) We prove that for each terminal $s \in X$, the probability to be connected by a path satisfies $\Pr[s \in V(T)] \geq \Omega(\frac{1}{\ell})$.

Part (*i*) provides that the expected cost for the sampled paths is at most $OPT_f$, while part (*ii*) implies that after repeating the sampling procedure $O(\ell \log |X|)$ times, each terminal will be connected to the root with high probability. Let us begin with part (*i*).

### Upper bounding the expected cost

For each node $v \in V$, let $Q(v) := \{P \mid P \text{ is } r\text{-}v \text{ path}\}$ be the set of paths from the root to $v$. For an edge $e = (u, v) \in E$ we denote $Q(e)$ as the set of $r$-$v$ paths that have $e$ as last edge.

**Lemma 3.** *Let $P$ be an $r$-$v$ path with $v \in V$. Then $\Pr[P \in T] = y_P$.*

*Proof.* Let $P = (e_1, \ldots, e_j)$ be the path with $e_i \in V_{i-1} \times V_i$. Then the probability that the algorithm samples path $P$ is

$$\Pr[P \in T] = y_{\{e_1\}} \cdot \frac{y_{\{e_1, e_2\}}}{y_{\{e_1\}}} \cdot \frac{y_{\{e_1, e_2, e_3\}}}{y_{\{e_1, e_2\}}} \cdot \ldots \cdot \frac{y_P}{y_{P \setminus \{e_j\}}} = y_P.$$

$\square$

The next lemma will imply that each edge $e$ is sampled with probability at most its fractional value $y_{\{e\}}$:

**Lemma 4.** *For any edge $e \in E$, one has $\sum_{P \in Q(e)} y_P \leq y_{\{e\}}$.*

*Proof.* We prove the following claim by induction over $j = 0, \ldots, \ell - 1$: *For any edge $e \in V_j \times V_{j+1}$ and any solution $\bar{Y} \in \text{LAS}_{t'}(K)$ with $t' \geq j$ one has $\sum_{P \in Q(e)} \bar{y}_P \leq \bar{y}_{\{e\}}$.*[6]

The claim is clear for $j = 0$, thus consider an edge $e = (u, v) \in V_j \times V_{j+1}$ between the $j$th and the $(j+1)$th level. Applying Thm. 2.(d) with $I := \{e\}$ we write $Y = y_{\{e\}} \cdot Y^{(1)} + (1 - y_{\{e\}}) \cdot Y^{(0)}$ such that $Y^{(0)}, Y^{(1)} \in \text{LAS}_{t'-1}(K)$ and $y_{\{e\}}^{(1)} = 1$ as well as $y_{\{e\}}^{(0)} = 0$. For edges $e' \in \delta^-(u)$ ingoing

---

[6]We abbreviate $\bar{y}_P := \bar{Y}_{P, \emptyset}$.

to $u$, we apply the induction hypothesis and get $\sum_{P\in Q(e')} y_P^{(1)} \le y_{\{e'\}}^{(1)}$. Since $y_{\{e\}}^{(1)} = 1$, we know that $y_{\{e\}\cup P}^{(1)} = y_P^{(1)}$ (see Theorem 2.(h)). It follows that

$$\sum_{P\in Q(e)} y_P \;=\; y_{\{e\}}\cdot\sum_{P\in Q(e)} y_P^{(1)} = y_{\{e\}}\sum_{e'\in\delta^-(u)}\sum_{P\in Q(e')}\underbrace{y_{P\cup\{e\}}^{(1)}}_{=y_P^{(1)}} \le y_{\{e\}}\sum_{e\in\delta^-(u)}\underbrace{y_{\{e'\}}^{(1)}}_{\le 1} \le y_{\{e\}}.$$

$$\underbrace{\phantom{y_{P\cup\{e\}}^{(1)}}}_{\le y_{\{e'\}}^{(1)}}$$

$\square$

Combining both Lemmas 3 and 4, we obtain $\Pr[e\in E(T)] \le y_{\{e\}}$ and consequently $E[c(E(T))] \le \sum_{e\in E} c_e y_{\{e\}}$ by linearity of expectation.

**Lower bounding the success probability**

In the following Lemma, we relate the "path variables" $y_P$ with the edge capacities $y_{\{e\}}$. In fact, (*a*) will imply that each terminal is connected once in expectation and (*b*) bounds the probability that $s$ is connected by a path containing a fixed subpath $P'$:

**Lemma 5.** *Fix a terminal $s\in X$ and an $r$-$v$ path $P'$ for some $v\in V$. Then*

a) $\sum_{P\in Q(s)} y_P = 1$

b) $\sum_{P\in Q(s): P'\subseteq P} y_P \le y_{P'}$.

*Proof.* Consider the set of variables $S := \{f_{s,e} \mid e\in E\}$. If more than $\ell$ of those variables are set to 1, they cannot define a feasible unit flow. In other words, we can apply Theorem 2.(e) in order to write $Y = \sum_{H\subseteq E} \lambda_H Y^H$ as a convex combination of vectors $Y^H$ such that for all $H$ with $\lambda_H > 0$ one has: (i) $Y^H \in \text{LAS}_\ell(K)$; (ii) $f_{s,e}^H \in \{0,1\}$ for all $e\in E$ and (iii) $f_{s,e}^H = 1 \Leftrightarrow e\in H$ (again we abbreviate $f_{s,e}^H := Y_{\emptyset,(s,e)}^H$ and $y_e^H := Y_{\{e\},\emptyset}^H$).

But the variables $\{f_{s,e}^H \mid e\in E\}$ can only represent a unit $r$-$s$ flow if $H$ is an $r$-$s$ path as well. Thus we have $\lambda_H = 0$, whenever this is not the case. In other words, our convex combination is of the form $Y = \sum_{P\in Q(s)} \lambda_P Y^P$.

Using Theorem 2.(b) we obtain $y_e^P \le 1$ for all $e\in P$ and the LP constraints imply $y_e^P \ge f_{s,e}^P = 1$. We conclude that $y_e^P = 1$ for all $e\in P$. Then Theorem 2.(f) provides $y_P^P = 1$.

Conversely, consider any $r$-$s$ path $P'\in Q(s)$ with $P'\ne P$ and let $v\in V(P)$ be a vertex, where path $P'$ enters $P$, i.e. $P\cap\delta^-(v) = \{e\}$ and $P'\cap\delta^-(v) = \{e'\}$ with $e\ne e'$. Since $y_e^P = 1$ and $\sum_{e''\in\delta^-(v)} y_{e''}^P \le 1$ (by LP constraint), we have $y_{e'}^P = 0$ and thus $y_{P'}^P = 0$. We conclude Claim *a*), since

$$y_P = \sum_{\bar P\in Q(s)} \lambda_{\bar P}\;\underbrace{y_P^{\bar P}}_{=0\text{ if }P\ne\bar P} = \lambda_P$$

and $\sum_{P\in Q(s)} \lambda_P = 1$.

To see *b*) note that $y_{P'}^P = 1$, whenever $P'\subseteq P$. Thus

$$y_{P'} = \sum_{P\in Q(s)} \lambda_P y_{P'}^P \ge \sum_{P\in Q(s): P'\subseteq P} \lambda_P = \sum_{P\in Q(s): P'\subseteq P} y_P.$$

$\square$

8

In the following, we fix a terminal $s \in X$ and define $Z := |T \cap Q(s)|$ as the random variable that yields the number of sampled paths that end in $s$. Our goal is to show that $\Pr[Z \geq 1] \geq \Omega(\frac{1}{\ell})$. Recall that by Lemma 5.(a) and Lemma 3, we know already that

**Corollary 6.** $E[Z] = 1$.

Interestingly, the key insight of Garg, Konjevod and Ravi [GKR00] is to prove an *upper* bound on $Z$ in order to *lower* bound $\Pr[Z \geq 1]$.

**Lemma 7.** $E[Z \mid Z \geq 1] \leq \ell + 1$.

*Proof.* Fix a path $P = (e_1, \ldots, e_\ell) \in Q(s)$. It suffices to show $E[Z \mid P \in T] \leq \ell + 1$.[7] Let $P_i = (e_1, \ldots, e_i) \subseteq P$ be the $r$-subpath of $P$ containing the first $i$ edges.

Consider any path $P' \in Q(s)$ and say it contains $P_i$, but not $P_{i+1}$. Since the probability distribution depends only on the "joint history" of $P$ and $P'$, we have $\Pr[P' \in T \mid P \in T] = \Pr[P' \in T \mid P_i \in T]$. We use this to bound

$$
E[|\{P' \in T \cap Q(s) \mid P_i \subseteq P'; P_{i+1} \nsubseteq P'\}|] \quad \leq \quad \sum_{P' \in Q(s): P' \supseteq P_i} \Pr[P' \in T \mid P_i \in T]
$$

$$
\stackrel{\text{cond. prob.}}{=} \sum_{P' \in Q(s): P' \supseteq P_i} \frac{\Pr[P' \in T \text{ and } P_i \in T]}{\Pr[P_i \in T]}
$$

$$
\stackrel{\text{Lemma 3}}{=} \sum_{P' \in Q(s): P' \supseteq P_i} \frac{y_{P'}}{y_{P_i}} \stackrel{\text{Lemma 5}}{\leq} 1
$$

The claim follows since there are only $\ell + 1$ such paths $P_i$. $\qquad \square$

Garg-Konjevod-Ravi [GKR00] make use of a sophisticated probabilistic result, the *Janson Inequality* (see e.g. [AS08]). However, the desired bound can be achieved much easier:

**Lemma 8.** $\Pr[Z \geq 1] \geq \frac{1}{\ell+1}$.

*Proof.* By the law of total probability

$$
1 = E[Z] = \Pr[Z = 0] \cdot \underbrace{E[Z \mid Z = 0]}_{=0} + \Pr[Z \geq 1] \cdot \underbrace{E[Z \mid Z \geq 1]}_{\leq \ell+1 \text{ by Lem. 7}}
$$

thus $\Pr[Z \geq 1] \geq \frac{1}{\ell+1}$. $\qquad \square$

Finally, we show the $O(\ell \log |X|)$ integrality gap. Interestingly, though the Lasserre solution $Y$ has $n^{\Theta(\ell)}$ entries, we only query a polynomial number of entries $y_P$. In other words, if we could evaluate each single entry $y_P$ in polynomial time, the algorithm would be polynomial as well.

**Theorem 9.** *Let $Y \in \text{LAS}_t(K)$ be a given $t = 2\ell$ round Lasserre solution. Then one can compute a feasible solution $H \subseteq E$ with $E[c(H)] \leq O(\ell \log |X|) \cdot \sum_{e \in E} y_{\{e\}}$. The expected number of Lasserre queries and the expected overhead running time are both polynomial in $n$.*

---

[7] The formal argument works as follows: Let $A_1, \ldots, A_m$ be any events (in our application, $A_P$ is the event "$P \in T$, *conditioned on $Z \geq 1$*") and $Z := |\{i \mid A_i\}|$ the number of occurring events. We claim that $E[Z] \leq \max_{i \in [m]} E[Z \mid A_i] =: \rho$. Proof: Using Jensen's inequality $E[Z]^2 \leq E[Z^2] = \sum_{i,j} \Pr[A_i \cap A_j] = \sum_i \Pr[A_i] \cdot \sum_j \Pr[A_j \mid A_i] = \sum_i \Pr[A_i] \cdot E[Z \mid A_i] \leq \rho \sum_i \Pr[A_i] = \rho E[Z]$. Rearranging yields the claim.

*Proof.* Repeat the sampling algorithm for $2\ell \log|X|$ many times and let $H$ be the union of the sampled paths. The probability that a fixed terminal $s \in X$ is not connected is bounded by $(1 - \frac{1}{\ell+1})^{2\ell \log|X|} \leq \frac{1}{|X|}$. For any terminal $s$ that remains unconnected, we buy the cheapest $r$-$s$ path. The expected cost for this repair step is bounded by $|X| \cdot \frac{1}{|X|} \cdot \sum_{e \in E} c_e y_{\{e\}}$.

Consider a single sample $T$. For every node $v \in V$ (no matter whether terminal or not), the expected number of paths $P \in T$ connecting $v$ is upper bounded by one. Thus the expected number of sampled partial paths is bounded by $n$ (if we denote the total number of nodes on all layers by $n$). After a partial path $P$ is sampled, the algorithm queries at most $n$ values of the form $y_{P \cup \{e\}}$. Thus the total expected number of queries for a single sample is upper bounded by $n^2$. $\qquad\square$

Together with Lemma 1 and the fact that $Y$ can be computed in time $n^{O(\ell)}$, this provides a polynomial time $|X|^\varepsilon$-approximation algorithm for any constant $\varepsilon > 0$. If we choose $\ell = \log|X|$, then we obtain a $O(\log^3|X|)$ approximation in time $n^{O(\log|X|)}$.

*Remark* 1. Observe that we explicitly used the Decomposition Theorem of [KMN11] in the proof of Lemma 5. Since the Decomposition Theorem does not hold for the Sherali-Adams or Lovász-Schrijver hierarchy, it is not clear whether the same integrality gap bound is true for those weaker relaxations.

However, there is a well-known reduction from a level-$\ell$ instance of DIRECTED STEINER TREE to a tree instance $F$ of GROUP STEINER TREE such that the produced tree $F$ has size $n^{O(\ell)}$ and contains all possible integral DST solutions as subtree. Of course, the corresponding GROUP STEINER TREE LP for this instance $F$ has only a polylogarithmic integrality gap [GKR00] and can also be interpreted as an LP for DIRECTED STEINER TREE.

It remains a challenging open problem, whether there is a convex relaxation with a polylog$(|X|)$ integrality gap that can be solved in polynomial time. Note that it would in fact suffice, to have a polynomial time oracle that takes a single path $P \subseteq E$ as input and outputs the Lasserre entry $y_P$.

# References

[AAT05]    M. Alekhnovich, S. Arora, and I. Tourlakis. Towards strong nonapproximability results in the Lovasz-Schrijver hierarchy. In *STOC*, pages 294–303, 2005.

[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.

[AMS12]    N. Alon, A. Moitra, and B. Sudakov. Nearly complete graphs decomposable into large induced matchings and their applications. In *STOC*, pages 1079–1090, 2012.

[ARV04]    S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *STOC*, pages 222–231, 2004.

[AS08]     N. Alon and J. H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.

[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.

[BGK10]    N. Bansal, A. Gupta, and R. Krishnaswamy. A constant factor approximation algorithm for generalized min-sum set cover. In *SODA*, pages 1539–1545, 2010.

[BRS11]    B. Barak, P. Raghavendra, and D. Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS*, volume abs/1104.4680, 2011.

[CCC+99]   M. Charikar, C. Chekuri, T. Cheung, Z. Dai, A. Goel, S. Guha, and M. Li. Approximation algorithms for directed steiner problems. *J. Algorithms*, 33(1):73–91, 1999.

[Chl07]    E. Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *FOCS*, pages 691–701, 2007.

[CMM09]    M. Charikar, K. Makarychev, and Y. Makarychev. Integrality gaps for Sherali-Adams relaxations. In *STOC*, pages 283–292, 2009.

[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.

[CZ05]     G. Calinescu and A. Zelikovsky. The polymatroid steiner problems. *J. Comb. Optim.*, 9(3):281–294, 2005.

[dlVKM07]  W. Fernandez de la Vega and C. Kenyon-Mathieu. Linear programming relaxations of maxcut. In *SODA*, pages 53–61, 2007.

[Eis99]    F. Eisenbrand. On the membership problem for the elementary closure of a polyhedron. *Combinatorica*, 19(2):297–300, 1999.

[ES99]     F. Eisenbrand and A. Schulz. Bounds on the Chvátal rank of polytopes in the 0/1-cube. In *IPCO*, pages 137–150, 1999.

[Fei98]    U. Feige. A threshold of ln $n$ for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.

[GK99]      S. Guha and S. Khuller. Greedy strikes back: Improved facility location algorithms. *J. Algorithms*, 31(1):228–248, 1999.

[GKR00]     N. Garg, G. Konjevod, and R. Ravi. A polylogarithmic approximation algorithm for the group Steiner tree problem. *J. Algorithms*, 37(1):66–84, 2000. Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (San Francisco, CA, 1998).

[GMPT07]    K. Georgiou, A. Magen, T. Pitassi, and I. Tourlakis. Integrality gaps of 2 - o(1) for vertex cover SDPs in the Lovász-Schrijver hierarchy. In *FOCS*, pages 702–712, 2007.

[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.

[HK03]      E. Halperin and R. Krauthgamer. Polylogarithmic inapproximability. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, STOC '03, pages 585–594, New York, NY, USA, 2003. ACM.

[HKK$^+$03] E. Halperin, G. Kortsarz, R. Krauthgamer, A. Srinivasan, and N. Wang. Integrality ratio for group steiner trees and directed steiner trees. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '03, pages 275–284, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.

[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.

[KR08]      S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2-epsilon. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008.

[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.

[Lau03a]    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.

[Lau03b]    M. Laurent. Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. *Math. Oper. Res.*, 28(4):871–883, 2003.

[Li11]      S. Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. In *ICALP (2)*, pages 77–88, 2011.

[Lov75]     L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4):383–390, 1975.

[LS91]      L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[MS09]    C. Mathieu and A. Sinclair. Sherali-Adams relaxations of the matching polytope. In *STOC*, pages 293–302, 2009.

[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.

[Sch08]   G. Schoenebeck. Linear level lasserre lower bounds for certain k-CSPs. In *FOCS*, pages 593–602, 2008.

[ST10]    M. Singh and K. Talwar. Improving integrality gaps via Chvátal-Gomory rounding. In *APPROX-RANDOM*, pages 366–379, 2010.

[STT07a]  G. Schoenebeck, L. Trevisan, and M. Tulsiani. A linear round lower bound for Lovasz-Schrijver SDP relaxations of vertex cover. In *IEEE Conference on Computational Complexity*, pages 205–216, 2007.

[STT07b]  G. Schoenebeck, L. Trevisan, and M. Tulsiani. Tight integrality gaps for Lovasz-Schrijver LP relaxations of vertex cover and max cut. In *STOC*, pages 302–310, 2007.

[Tou06]   I. Tourlakis. New lower bounds for vertex cover in the Lovasz-Schrijver hierarchy. In *Computational Complexity, 2006. CCC 2006. Twenty-First Annual IEEE Conference on*, pages 170 –182, 0-0 2006.

[Tul09]   M. Tulsiani. CSP gaps and reductions in the lasserre hierarchy. In *STOC*, pages 303–312, 2009.

[Zel97]   A. Zelikovsky. A series of approximation algorithms for the acyclic directed steiner tree problem. *Algorithmica*, 18:99–110, 1997.

[ZK02]    L. Zosin and S. Khuller. On directed steiner trees. In *SODA*, pages 59–63, 2002.

# A   Properties of the Lasserre Hierarchy

The main goal of this section is to present a complete proof of the convergence of the Lasserre hierarchy (Theorem 2.(a)); the feasibility of "conditioned" solutions (Theorem 2.(d)) and the Karlin-Mathieu-Nguyen Decomposition Theorem [KMN11] (Theorem 2.(e)). In the following, we are going to reproduce the proof in [KMN11]. However, we will use a different notation and try to put the emphasis on an intuitive exposition instead of a space efficient one.

Let $\mathscr{P}([n]) := \mathscr{P}_n([n]) = \{I \mid I \subseteq [n]\}$ be the family of all subsets of $[n]$. Recall that $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ is the set of relaxed solutions with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

## A.1   The Inclusion-Exclusion Formula

Suppose for the moment, that $y \in \mathbb{R}^{\mathscr{P}([n])}$ indeed is *consistent*, i.e. there exists a random variable $Z \in \{0, 1\}^n$ such that $\Pr[\bigwedge_{i \in I}(Z_i = 1)] = y_I$ (thus $E[Z_i] = y_i$). Here, $Z$ can be taken from any distribution — especially $Z_i$ and $Z_{i'}$ do not need to be independent.

Initially, the Lasserre relaxation contains only variables for "positive" events of the form $\bigwedge_{i \in I}(Z_i = 1)$. But using the *inclusion-exclusion formula,* one can also obtain probabilities for all other events.

Recall that for any index set $J \subseteq [n]$ the inclusion-exclusion formula says that

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

Negating this event yields

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

Next, let $I \subseteq [n]$ be another index set (not necessarily disjoint to $J$). Observe that Equation (1) remains valid if all events are intersected with the same event $\bigwedge_{i \in I}(Z_i = 1)$. In other words we arrive at the *generalized inclusion exclusion formula* (sometimes called *Möbius inversion*)

$$\Pr\left[\bigwedge_{i \in I}(Z_i = 1) \wedge \bigwedge_{i \in J}(Z_i = 0)\right] = \sum_{H \subseteq J}(-1)^{|H|}\Pr\left[\bigwedge_{i \in I \cup H}(Z_i = 1)\right]. \quad (2)$$

Thus for any $I, J \subseteq [n]$ we define

$$y_{I,-J} := \sum_{H \subseteq J}(-1)^{|H|}y_{I \cup H} = \Pr\left[\bigwedge_{i \in I}(Z_i = 1), \bigwedge_{i \in J}(Z_i = 0)\right] \quad (3)$$

(for example $y_{\emptyset,-\{1\}} = y_\emptyset - y_{\{1\}}$). If $I \dot\cup J = [n]$, then we abbreviate $y_{I,-J} =: y_x$ as the probability for the *atomic event* $x \in \{0,1\}^n$ with

$$x_i = \begin{cases} 1 & i \in I \\ 0 & i \in J \end{cases}$$

Furthermore we denote $\mathrm{supp}(x) := \{i \in [n] \mid x_i = 1\}$ and $\overline{\mathrm{supp}}(x) := \{i \in [n] \mid x_i = 0\}$. We saw so far, that the $2^n$ many probabilities $y_I$ uniquely define the $2^n$ many probabilities $y_x$ for atomic events. Conversely, one can obtain the values $y_I$ and $y_{I,-J}$ by summing over all atomic events that are consistent with the events, i.e.

$$y_I = \sum_{x \in \{0,1\}^n : I \subseteq \mathrm{supp}(x)} y_x \quad (4)$$

$$y_{I,-J} = \sum_{x \in \{0,1\}^n : I \subseteq \mathrm{supp}(x), J \subseteq \overline{\mathrm{supp}}(x)} y_x. \quad (5)$$

Let us make a couple of observations:

- Equations (3) and (4) are both linear, thus they define an isomorphism between $(y_I)_{I \subseteq [n]}$ and $(y_x)_{x \in \{0,1\}^n}$. This isomorphism is well defined even if $y$ is not consistent (i.e. even if some $y_x$ are negative or $\sum_{x \in \{0,1\}^n} y_x \neq 1$).

- If $I \cap J \neq \emptyset$, then by definition one has $y_{I,-J} = 0$ since the sum in Eq. (3) can be grouped into pairs that have the same absolute value – but different signs (for example $y_{\{1\},-\{1,2\}} = y_{\{1\} \cup \emptyset} - y_{\{1\} \cup \{1\}} - y_{\{1\} \cup \{2\}} + y_{\{1\} \cup \{1,2\}} = 0$).

14

*Remark* 2. Lemma 2 in the survey of Laurent [Lau03a] states the following equivalences for $y \in \mathbb{R}^{\mathscr{P}([n])}$:

- $M_n(y) \succeq 0 \iff \forall x \in \{0,1\}^n : y_x \geq 0$

- $M_n(\binom{A_\ell}{b_\ell} * y) \succeq 0 \iff \forall x \in \{0,1\}^n : y_x \cdot (A_\ell x - b_\ell) \geq 0$

If both conditions hold and additionally $y_\emptyset = 1$, then $\sum_{x \in \{0,1\}^n} y_x = y_\emptyset = 1$ and the $y_x$ define a probability distribution over $\{0,1\}^n$ with $y_x = 0$ for all $x$ with $Ax \not\geq b$. In other words

$$(y_{\{1\}}, \ldots, y_{\{n\}}) = \sum_{x \in \{0,1\}^n : Ax \geq b} y_x \cdot x$$

is a convex combination of feasible points. However, we will show the convergence proof following [KMN11], which has more synergy effects with the decomposition theorem.

## A.2  Partial assignments and the inversion formula

Let $\mathscr{T} \subseteq \mathscr{P}([n])$ be a family of index sets and let $y \in \mathbb{R}^{\mathscr{T}}$ be a corresponding vector. For subsets $X \subseteq S \subseteq [n]$, we define the *conditioning on $X$ and $-S \setminus X$* (sometimes also called *partial assignment*) as the vector $z = \{y\}_{X, -S \setminus X} \in \mathbb{R}^{\mathscr{T} \ominus S}$ with $z_I := y_{I \cup X, -S \setminus X} = \sum_{H \subseteq S \setminus X} (-1)^{|H|} y_{I \cup X \cup H}$ and $\mathscr{T} \ominus S := \{I \in \mathscr{T} \mid \forall J \subseteq S : I \cup J \in \mathscr{T}\}$. The definition of entry $z_I$ only makes sense, if $I \cup J \in \mathscr{T}$ for all $J \subseteq S$, thus $\mathscr{T} \ominus S$ is the maximum family of sets, for which this is satisfied. Note that the set $\mathscr{T} \ominus S$ is not necessarily smaller than $\mathscr{T}$. For example $\mathscr{P}([n]) \ominus S = \mathscr{P}([n])$ and $\mathscr{P}_t([n]) \ominus S \subseteq \mathscr{P}_{t-|S|}([n])$.

We define the *normalized conditioning on $X$ and $-S \setminus X$* as $w := \frac{z}{z_\emptyset}$ (say $w := \mathbf{0}$ if $z_\emptyset = 0$ to be well-defined). The intuition is that if $Z \in \{0,1\}^n$ again is a random variable with $y_{I,-J} = \Pr[\bigwedge_{i \in I}(Z_i = 1), \bigwedge_{i \in J}(Z_i = 0)]$, then the (normalized) conditioned solution reflects conditional probabilities, i.e.

$$w_{I,-J} = \Pr\left[ \bigwedge_{i \in I}(Z_i = 1), \bigwedge_{i \in J}(Z_i = 0) \mid \bigwedge_{i \in X}(Z_i = 1), \bigwedge_{i \in S \setminus X}(Z_i = 0) \right].$$

The events $(X, -S \setminus X)$ obviously partition the probability space, if $X$ runs over all subsets of $S$. This remains valid for conditioned Lasserre solutions.

**Lemma 10** (Inversion formula)**.** *Let $y \in \mathbb{R}^{\mathscr{P}([n])}$ and $S \subseteq [n]$. Then $y = \sum_{X \subseteq S} \{y\}_{X, -S \setminus X}$.*

*Proof.* We verify the equation for entry $I \subseteq V$:

$$\begin{aligned}
\sum_{X \subseteq S} y_{I \cup X, -S \setminus X} &= \sum_{X \subseteq S} \sum_{x \in \{0,1\}^n : I \cup X \subseteq \text{supp}(x), S \setminus X \subseteq \overline{\text{supp}}(x)} y_x \\
&= \sum_{x \in \{0,1\}^n : I \subseteq \text{supp}(x)} y_x \cdot \underbrace{|\{X \subseteq S : X \subseteq \text{supp}(x), S \setminus X \subseteq \overline{\text{supp}}(x)\}|}_{=1} = y_I
\end{aligned}$$

$\qquad\square$

## A.3   Feasibility of conditioned solutions

Next, we will see that conditioned solutions are still feasible on a smaller family of index sets:

**Lemma 11.** *Let* $X \subseteq S \subseteq V$, *and* $y \in \mathbb{R}^{\mathscr{P}([n])}$ *with* $\mathscr{T} \subseteq \mathscr{P}(V)$. *Then*

$$M_{\mathscr{T}}(y) \succeq 0 \quad \Longrightarrow \quad M_{\mathscr{T} \ominus S}(\{y\}_{X,-S\setminus X}) \succeq 0$$

*Proof.* Abbreviate $z := \{y_{X,-S\setminus X}\} \in \mathbb{R}^{\mathscr{P}([n])}$. Equivalently to $M_{\mathscr{T}}(y) \succeq 0$, there must be vectors $v_I$ with $v_I \cdot v_J = y_{I \cup J}$ for $I, J \in \mathscr{T}$. For each $I \in \mathscr{T} \ominus S$, we define another vector

$$w_I := \sum_{H \subseteq S\setminus X} (-1)^{|H|} v_{I \cup X \cup H}.$$

We claim that those vectors provide a factorization of $M_{\mathscr{T} \ominus S}(z)$ (which proves the positive semi-definiteness of $M_{\mathscr{T} \ominus S}(z)$)

$$
\begin{aligned}
w_I \cdot w_J &= \sum_{H \subseteq S\setminus X} \sum_{L \subseteq S\setminus X} (-1)^{|H|+|L|} v_{I \cup X \cup H} \cdot v_{J \cup X \cup L} \\
&= \sum_{H \subseteq S\setminus X} \underbrace{\sum_{L \subseteq S\setminus X} (-1)^{|H|+|L|} y_{I \cup J \cup X \cup H \cup L}}_{(*)=0 \text{ if } H \neq \emptyset} \\
&= \sum_{L \subseteq S\setminus X} (-1)^{|L|} y_{I \cup J \cup X \cup L} = y_{I \cup J \cup X, -S\setminus X} = z_{I \cup J}.
\end{aligned}
$$

To see $(*)$, observe that if there is any $i \in H \subseteq S\setminus X$, then the sum contains the same term for $L \subseteq (S\setminus X)\setminus\{i\}$ and $L \cup \{i\}$ – just with different sign, so that the sum evaluates to 0. $\qquad\square$

The following lemma is usually called *commutativity of the shift operator* (see e.g. [Lau03a]). Recall that for $y \in \mathbb{R}^{\mathscr{T}}$, we interpret $w = \binom{a}{\beta} * y$ as the vector with $w_I := (\sum_{i \in [n]} a_i y_{I \cup \{i\}} - \beta y_I)$.[8]

**Lemma 12.** *Let* $y \in \mathbb{R}^{\mathscr{T}}$; $X \subseteq S \subseteq [n]$; *and* $a^T x \geq \beta$ *be a linear constraint. Then*

$$\binom{a}{\beta} * \underbrace{\{y\}_{X,-S\setminus X}}_{=:z} = \underbrace{\{\binom{a}{\beta} * y\}_{X,-S\setminus X}}_{=:u} \quad (\in \mathbb{R}^{\mathscr{T}'})$$

*with* $\mathscr{T}' := \{I \subseteq [n] \mid I \cup J \cup \{i\} \in \mathscr{T} \; \forall J \subseteq S \; \forall i \in [n]\}$.

*Proof.* Let $z_I = y_{I \cup X, -S\setminus X}$ and $u_I = \sum_{i \in [n]} a_i y_{I \cup \{i\}} - y_I \beta$. Evaluating the left hand side vector at entry $I \in \mathscr{T}'$ gives

$$
\begin{aligned}
\left(\binom{a}{\beta} * \{y\}_{X,-S\setminus X}\right)_I &= \sum_{i \in [n]} a_i z_{I \cup \{i\}} - \beta z_I = \sum_{i \in [n]} a_i y_{I \cup X \cup \{i\}, -S\setminus X} - \beta y_{I \cup X, -S\setminus X} \\
&= \sum_{H \subseteq S\setminus X} (-1)^{|H|} \left[ \sum_{i \in [n]} a_i y_{I \cup X \cup \{i\} \cup H} - \beta y_{I \cup X \cup H} \right]
\end{aligned}
$$

The right hand side entry for $I$ is

$$
\begin{aligned}
u_{I \cup X, -S\setminus X} &= \sum_{H \subseteq S\setminus X} (-1)^{|H|} u_{X \cup I \cup H} \\
&= \sum_{H \subseteq S\setminus X} (-1)^{|H|} \left[ \sum_{i \in [n]} a_i y_{X \cup I \cup \{i\} \cup H} - \beta y_{I \cup X \cup H} \right]
\end{aligned}
$$

Both expressions are identical and the claim follows. $\qquad\square$

---

[8]In fact, we were sloppy concerning the dimension of $w$ so far. Formally, one should define $w \in \mathbb{R}^{\mathscr{T}'}$ with $\mathscr{T}' := \{I \mid I \cup \{i\} \in \mathscr{T} \; \forall i \in [n]\}$.

Now we will see that normalized conditioned solutions are feasible and integral on variables in $S$:

**Lemma 13.** *Let $X \subseteq S \subseteq [n]$, $\mathcal{T}_1, \mathcal{T}_2 \subseteq \mathcal{P}([n])$ and $y \in \mathbb{R}^{\mathcal{P}([n])}$ with*

$$M_{\mathcal{T}_1}(y) \succeq 0; \quad M_{\mathcal{T}_2}(\left(\begin{smallmatrix} A_\ell \\ b_\ell \end{smallmatrix}\right) * y) \succeq 0 \; \forall \ell \in [m]; \quad y_\emptyset = 1.$$

*Define $z := \{y\}_{X, -S \setminus X} \in \mathbb{R}^{\mathcal{P}([n])}$. If $z_\emptyset > 0$, then for $w := \frac{z}{z_\emptyset}$ one has*

- $M_{\mathcal{T}_1 \ominus S}(w) \succeq 0; \quad M_{\mathcal{T}_2 \ominus S}(\left(\begin{smallmatrix} A_\ell \\ b_\ell \end{smallmatrix}\right) * w) \succeq 0 \; \forall \ell \in [m]; \quad w_\emptyset = 1.$
- $w_I = 1$ *if $I \subseteq X$*
- $w_I = 0$ *if $I \subseteq S$ but $I \cap (S \setminus X) \neq \emptyset$.*

*Proof.* Lemma 11 directly implies that $M_{\mathcal{T}_1 \ominus S}(w) = \frac{1}{z_\emptyset} M_{\mathcal{T}_1 \ominus S}(z) \succeq 0$. Applying the commutativity rule,

$$
\begin{aligned}
M_{\mathcal{T}_2 \ominus S}(\left(\begin{smallmatrix} A_\ell \\ b_\ell \end{smallmatrix}\right) * w) \quad &= \quad \frac{1}{z_\emptyset} M_{\mathcal{T}_2 \ominus S}(\left(\begin{smallmatrix} A_\ell \\ b_\ell \end{smallmatrix}\right) * \{y\}_{X, -S \setminus X}) \\
&\overset{\text{Lemma 12}}{=} \frac{1}{z_\emptyset} M_{\mathcal{T}_2 \ominus S}(\{\left(\begin{smallmatrix} A_\ell \\ b_\ell \end{smallmatrix}\right) * y\}_{X, -S \setminus X}) \overset{\text{Lemma 11}}{\succeq} 0
\end{aligned}
$$

using that $M_{\mathcal{T}_2}(\left(\begin{smallmatrix} A_\ell \\ b_\ell \end{smallmatrix}\right) * y) \succeq 0$. Furthermore for $I \subseteq S$ one has

$$w_I = \frac{z_I}{z_\emptyset} = \frac{y_{X \cup I, -S \setminus X}}{y_{X, -S \setminus X}} = \begin{cases} 1 & I \subseteq X \\ 0 & I \cap (S \setminus X) \neq \emptyset \end{cases}.$$

$\square$

## A.4 Convergence

With the last lemma at hand, the convergence of the Lasserre hierarchy follows quickly.

**Lemma 14.** $K \supseteq \text{LAS}_0^{proj}(K) \supseteq \ldots \supseteq \text{LAS}_n^{proj}(K) \supseteq conv(K \cap \{0,1\}^n)$.

*Proof.* By definition $\text{LAS}_t^{proj}(K) \supseteq \text{LAS}_{t+1}^{proj}(K)$. Let $y \in \text{LAS}_0(K)$. Then $0 \leq (M_0(\left(\begin{smallmatrix} A_\ell \\ b_\ell \end{smallmatrix}\right) * y))_{\emptyset, \emptyset} = \sum_{i \in [n]} A_{\ell i} y_{\{i\}} - b_i$, thus $A(y_{\{1\}}, \ldots, y_{\{n\}}) \geq b$ and $K \supseteq \text{LAS}_0^{proj}(K)$.

Finally let $x \in K \cap \{0,1\}^n$ and define $y \in \mathbb{R}^{\mathcal{P}([n])}$ with $y_I := \prod_{i \in I} x_i$. Then one has $M_n(y) = yy^T \succeq 0$. Furthermore $M_n(\left(\begin{smallmatrix} A_\ell \\ b_\ell \end{smallmatrix}\right) * y) = (A_\ell x - b_\ell) \cdot yy^T \succeq 0$, thus $y \in \text{LAS}_n(K)$. $\square$

The following statement implies that $\text{LAS}_n^{proj}(K) = conv(K \cap \{0,1\}^n)$:

**Lemma 15.** $\text{LAS}_n(K) \subseteq conv\{(\prod_{i \in I} x_i)_{I \subseteq [n]} \mid x \in \{0,1\}^n : Ax \geq b\}$.

*Proof.* For all $X \subseteq [n]$, define $z^X := \{y\}_{X, -[n] \setminus X} \in \mathbb{R}^{\mathcal{P}([n])}$ and $w^X := \frac{z^X}{z_\emptyset^X}$ if $z_\emptyset^X > 0$. Then

$$y \overset{\text{Lemma 10}}{=} \sum_{X \subseteq [n]} z^X = \sum_{X \subseteq [n] : z_\emptyset^X > 0} z_\emptyset^X \cdot w^X.$$

This is a convex combination since $\sum_{X \subseteq [n]} z_\emptyset^X = y_\emptyset = 1$ (again by Lemma 10). For a fixed $X$, we abbreviate $x_i := w_{\{i\}}^X$. Then Lemma 13 provides that $w^X \in \text{LAS}_n(K)$ (thus $x \in K$); $x \in \{0,1\}^n$ and $w_I^X = \prod_{i \in I} x_i$. $\square$

17

## A.5 Local consistency

Let $y \in \mathbb{R}^{\mathscr{P}_t(V)}$ and $t' < t$. Then $y_{|\mathscr{P}_{t'}(V)} \in \mathbb{R}^{\mathscr{P}_{t'}(V)}$ is the vector that emerges from $y$ after deletion of all entries $I$ with $|I| > t'$. Moreover, for any vector $y \in \mathbb{R}^{\mathscr{T}}$ with $\mathscr{T} \subseteq \mathscr{P}([n])$, we define the *extension* $y' \in \mathbb{R}^{\mathscr{P}([n])}$ as the vector $y$ where missing entries are filled with zeros.

**Lemma 16.** *Let $y \in \mathrm{LAS}_t(K)$ and $S \subseteq [n]$ with $|S| \leq t$. Then*

$$y \in conv\{w \mid w_{|\mathscr{P}_{2(t-|S|)+2}} \in \mathrm{LAS}_{t-|S|}(K); w_{\{i\}} \in \{0,1\} \ \forall i \in S\}$$

*Proof.* Again we can write a convex combination $y' = \sum_{X \subseteq S} z_\emptyset^X \cdot w^X$ (with $w^X \in \mathbb{R}^{\mathscr{P}(n)}$) according to Lemma 10. Recall that $\mathscr{P}_{2t+2}([n]) \ominus S \supseteq \mathscr{P}_{2(t-|S|)+2}([n])$, thus Lemma 13 provides that $w_{|\mathscr{P}_{2(t-|S|)+2}([n])} \in \mathrm{LAS}_{t-|S|}(K)$ and that $w$ is integral on $S$. $\qquad\square$

## A.6 The decomposition theorem

Imagine for a second that $y \in \mathrm{LAS}_t(K)$ and $y_I = 0$ for all $|I| \geq t$. Then we can just fill the matrices $M_{t+1}(y)$ and $M_t(\binom{A_\ell}{b_\ell} * y)$ with zeros to obtain $M_n(y)$ and $M_n(\binom{A_\ell}{b_\ell} * y)$ without destroying positive semi-definiteness or their consistency. Consequently, $y$ would even be in the convex hull of feasible integral vectors, even though we only assumed $y$ to be a $t$-round solution. With a bit more care, this approach applies more generally to any subset of variables.

**Theorem 17** (Decomposition Theorem [KMN11])**.** *Let $0 \leq k \leq t$, $y \in \mathrm{LAS}_t(K)$, $X \subseteq S \subseteq V$ so that $|I \cap S| > k \Rightarrow y_I = 0$. Then*

$$y \in conv\{w \mid w_{|2(t-k)+2} \in \mathrm{LAS}_{t-k}(K); w_{\{i\}} \in \{0,1\} \ \forall i \in S\}$$

*Proof.* Again extend $y$ to $y' \in \mathbb{R}^{\mathscr{P}(n)}$. Define $\mathscr{T}_1 := \{A \subseteq [n] \mid |A \backslash S| \leq t + 1 - k\}$ as the set of indices that have at most $t + 1 - k$ indices outside of $S$. After sorting the rows and columns by increasing size of index sets, we can write

$$M_{\mathscr{T}_1}(y') = \left(\begin{array}{cc} \boxed{\begin{array}{c} J : |J| \leq t+1 \\ \hline M \end{array}} & \boxed{\begin{array}{c} J : |J| > t+1 \\ \hline \mathbf{0} \end{array}} \\ \boxed{\mathbf{0}} & \boxed{\mathbf{0}} \end{array}\right) \succeq 0$$

where $M$ is a principal submatrix of $M_{t+1}(y) \succeq 0$. Next, observe that there are entries $y'_J$ that may appear inside of $M$ and outside. If they appear outside, say at entry $(J_1, J_2)$, then $J = J_1 \cup J_2$ and $|J_1| > t + 1$, but $|J_1 \backslash S| \leq t + 1 - k$ (so that $J_1 \in \mathscr{T}_1$). Thus $|J \cap S| \geq |J_1 \cap S| = |J_1| - |J_1 \backslash S| > k$, thus $y'_J = 0$ by assumption[9].

Analogously for $\mathscr{T}_2 := \{A \subseteq [n] \mid |A \backslash S| \leq t - k\}$ one can write

$$M_{\mathscr{T}_2}\left(\binom{A_\ell}{b_\ell} * y'\right) = \left(\begin{array}{cc} \boxed{\begin{array}{c} J : |J| \leq t \\ \hline N \end{array}} & \boxed{\begin{array}{c} J : |J| > t \\ \hline \mathbf{0} \end{array}} \\ \boxed{\mathbf{0}} & \boxed{\mathbf{0}} \end{array}\right) \succeq 0$$

---

[9] In other words, the matrix $M_{\mathscr{T}_1}(y')$ is *consistent* in the sense that entries $(J_1, J_2)$ and $(J_3, J_4)$ are identical whenever $J_1 \cup J_2 = J_3 \cup J_4$.

with $N \succeq 0$ being a principal submatrix of $M_t(\binom{A_\ell}{b_\ell} * y) \succeq 0$. Again, consider an entry $J$ that appears at least once outside of $N$ thus we can write $J = J_1 \cup J_2$ with $|J_1| > t$ and $J_1 \in \mathcal{T}_2$, hence $|J_1 \setminus S| \le t - k$. Then for any $J' \supseteq J_1$ one has $|J' \cap S| \ge |J_1 \cap S| = |J_1| - |J_1 \setminus S| \ge (t+1) - (t-k) > k$ and $y'_{J'} = 0$ by assumption. The matrix entry at position $J$ is $\sum_{i \in [n]} A_{\ell i} y_{J \cup \{i\}} - \beta y_J = 0$, hence the matrix is defined consistently.

Define $z^X := \{y\}_{X, -S \setminus X} \in \mathbb{R}^{\mathscr{P}([n])}$ and $w^X := \frac{z^X}{z_\emptyset^X}$ ($w^X := \mathbf{0}$, if $z_\emptyset^X = 0$), then as in Lemma 15, we can express $y'$ as convex combination $y' = \sum_{X \subseteq S} z_\emptyset^X w^X$. Since $\mathcal{T}_1 \ominus S = \mathcal{T}_1$ and $\mathcal{T}_2 \ominus S = \mathcal{T}_2$, we apply Lemma 13 and obtain $M_{\mathcal{T}_1}(w^X) \succeq 0$, $M_{\mathcal{T}_2}(\binom{A_\ell}{b_\ell} * w^X) \succeq 0$ for all $\ell \in [m]$ and $w_\emptyset^X = 1$. Observe that $\mathscr{P}_{t+1-k}([n]) \subseteq \mathcal{T}_1$ and $\mathscr{P}_{t-k}([n]) \subseteq \mathcal{T}_2$. Consequently $w_{|2(t-k)+2} \in \text{LAS}_{t-k}(K)$ as claimed. $\qquad\square$

Now we argue, why the Decomposition Theorem implies Property 2.(e).

**Lemma** (Property 2.(e))**.** *Let $K = \{x \in \mathbb{R}^n \mid Ax \ge b\}$ and $y \in \text{LAS}_t(K)$ and assume that for a subset of variables $S \subseteq [n]$ one has $\max\{|I| : I \subseteq S; x \in K; x_i = 1 \; \forall i \in I\} \le k < t$. Then $y \in conv(\{z \in \text{LAS}_{t-k}(K) \mid z_{\{i\}} \in \{0, 1\} \; \forall i \in S\})$.*

*Proof.* Consider an index set $I \subseteq S$ with $|I| = k + 1 \le t$. Then Property 2.(c) implies that $y_I = 0$. For all $|I|, |J| \le t + 1$ with $y_I = 0$, inspecting the determinant of the principal submatrix of $M_{t+1}(y)$ induced by indices $I$ and $J$, we see that

$$0 \le \det \begin{pmatrix} 0 & y_{I \cup J} \\ y_{I \cup J} & y_J \end{pmatrix} = -y_{I \cup J}^2,$$

thus $y_{I \cup J} = 0$ (i.e., all entries are monotone). We summarize: All entries $I \subseteq [n]$ with $|I| \le 2(t+1)$ and $|I \cap S| > k$ have $y_I = 0$. Then by the Decomposition Theorem we have

$$y \in \text{conv}\{w \mid w_{|2(t-k)+2} \in \text{LAS}_{t-k}(K); \; w_{\{i\}} \in \{0, 1\} \; \forall i \in S\}.$$

$\qquad\square$