

Lower bounds on the size of SDP relaxations

James R. Lee, Prasad Raghavendra, David Steurer
Presented by Martin Böhm
Combinatorics PhD Seminar, 2014/15, MFF UK

Some notation

1. A $\mathcal{P}_{2t}(n)$ is a set of all subsets of $[n]$ of size at most $2t$.
2. $\mathbb{N}_{\leq t}^n$ is a set of all monomials in n variables of degree at most t . (Here $0 \in \mathbb{N}$.)
3. $\|A\|_F$ is the Frobenius norm, defined as $\sqrt{\text{Tr}(A^T A)}$.
4. For a polynomial $p(x)$ and a monomial α , $[\alpha]p$ is the coefficient of the monomial α .
5. $\text{vec}(p)$ is a vectorization of a polynomial p – $\text{vec}(p)$ is thus a vector with each monomial α of p contributing a real number $[\alpha]p$.

Semidefinite programming

D: A symmetric matrix $M \in \mathbb{R}^{n \times n}$ is positive semidefinite $\equiv \forall z \in \mathbb{R}^n : z^T M z \geq 0$. We write $M \succeq 0$. Equivalently, M is positive semidefinite iff it has a square root, i.e. there exists U such that $U^T U = M$. We write $U = \sqrt{M}$.

D(S_+^k): The set (cone) of all positive semidefinite matrices in $\mathbb{R}^{k \times k}$ will be denoted S_+^k .

D: A linear operator $\bullet : \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is defined as $A \bullet B = \text{Tr}(A^T B)$.

D: A semidefinite program is a convex optimization program of the form:

$\max C \bullet X$ subject to constraints $A_i \bullet X = b_i$ and $X \succeq 0$.

D: For a program of the form $\max C \bullet X, A_i \bullet X = b_i, X \succeq 0$, we can define a *dual program* $\min b^T y$ subject to $\sum_{i=1}^m y_i A_i - C \succeq 0$.

Under mild conditions (if the primal is feasible, with finite value, and it has at least one positive definite solution) it holds that the values of the primal and dual semidefinite programs coincide.

Extended formulations

D(Linear lift): Consider some polytope P . We say that a polytope Q is a *linear lift* of P if P is an image of Q under some linear map. We measure the size of the lift as the number of facets. The polytope Q is also called *extension* of P .

D(PSD lift): We say that a polytope P of dimension n admits a *positive semidefinite lift* of size k if there exists an affine subspace $\mathcal{L} \subseteq \mathbb{R}^{k \times k}$ and a projection $\pi : \mathbb{R}^{k \times k} \rightarrow \mathbb{R}^n$ such that $P = \pi(\mathcal{L} \cap S_+^k)$.

In other words, we say that P has a positive semidefinite lift of size k if we can get P as some projection of a structure that is both inside an affine space ($A \bullet X = b$) and inside the set of all psd matrices ($X \succeq 0$).

D: We define a polytope $\text{CORR}_n \equiv \text{conv}(x \cdot x^T | x \in \{0, 1\}^n)$. CORR_n played an important role in the paper of Fiorini, Massar, Pokutta, Tiwary and de Wolf; bounds on extension complexity of CORR_n imply bounds on the cut polytope, TSP polytope and independent set polytope, among others.

D(PSD rank): We say that $M \in \mathbb{R}_+^{p \times q}$ admits a *rank- r psd factorization* if there exist two lists of positive semidefinite matrices $\{A_i \in S_+^r | i \in [p]\}$ and $\{B_j \in S_+^r | j \in [q]\}$ such that $M_{ij} = \sum_{x=1}^r \langle (A_i)_x, (B_j)_x \rangle = \text{Tr}(A_i B_j)$.

D: We say that $M \in \mathbb{R}_+^{p \times q}$ has psd rank r if r is the smallest number such that M admits a rank- r psd factorization. We say that a polytope P has psd rank r if an associated slack matrix S has psd rank r .

T(Bound on psd rank):

$$\text{rk}_{\text{psd}}(\text{CORR}_n) \geq 2^{\Omega(n^{2/13})}.$$

D(From a function to a matrix): For a function $f : \{0, 1\}^m \rightarrow \mathbb{R}$ and a number $n \geq m$, we define the following matrix $M_n^f \in \mathbb{R}_+^{\binom{n}{m} \times 2^n}$:

$$M_n^f(S, x) = f(x_S) = f(x \text{ restricted to the positions in } S)$$

Clm: If $f : \{0, 1\}^m \rightarrow \mathbb{R}_+$ is a non-negative quadratic polynomial over $\{0, 1\}^m$, then for any $n \geq m$ it holds that M_n^f is a submatrix of some slack matrix associated with CORR_n .

T(Large PSD rank $\Leftrightarrow$ large lift): For every $n, k \geq 1$ every polytope $P \subseteq \mathbb{R}^n$ and every slack matrix S associated to P , it holds that $\text{rk}_{\text{psd}}(S) \leq k$ if and only if P admits a positive semidefinite lift of size k .

From the previous Theorem and Claim, we know that to finish the proof of Theorem *Bound on a psd rank*, we need a good lower bound on $\text{rk}_{\text{psd}}(M_n^f)$.

Lasserre hierarchy

Notation: Let $\mathcal{P}_t([n]) := \{I \subseteq [n] \mid |I| \leq t\}$ be the set of all index sets of cardinality at most t and let $y \in \mathbb{R}^{\mathcal{P}_{2t}([n])}$ be a vector with entries y_I for all $I \subseteq [n]$ with $|I| \leq 2t$.

D(Moment matrix): $M_{t+1}(y) \in \mathbb{R}^{\mathcal{P}_{t+1}([n]) \times \mathcal{P}_{t+1}([n])}$:

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

D(Moment matrix of slacks): For the ℓ -th ($\ell \in [m]$) constraint of the LP $A^T x \geq b$, we create $M_t^\ell(y) \in \mathbb{R}^{\mathcal{P}_t([n]) \times \mathcal{P}_t([n])}$:

$$M_t^\ell(y)_{I,J} := \left(\sum_{i=1}^n A_{li} y_{I \cup J \cup \{i\}} \right) - b_l y_{I \cup J}$$

D(t -th level of the Lasserre hierarchy): Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$. Then $\text{LAS}_t(K)$ is the set of vectors $y \in \mathbb{R}^{\mathcal{P}_{2t}([n])}$ that satisfy

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

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

Intuition: $M_{t+1}(y) \succeq 0$ ensures *consistency* (y behaves *locally* as a distribution) while $M_t^\ell(y) \succeq 0$ guarantees that y satisfies the ℓ -th linear constraint.

We call any solution y a *pseudo-distribution* or a *pseudo-density* of vertices of the polytope. In our $\{0, 1\}^n$ setting, it is a pseudo-density on $\{0, 1\}^n$.

Note: In our case, we only deal with problems of the form $\max f(x)$ subject to $x \in \{0, 1\}^n$, and so we can simplify our Lasserre system to:

$$\max f(y_{\{1\}}, y_{\{2\}}, \dots, y_{\{n\}}) \quad \text{s.t.} \quad M_{t+1}(y) \succeq 0; \quad y_\emptyset = 1.$$

Sum of Squares upper bounds

D: For a polynomial $f, f : \{0, 1\}^n \rightarrow \mathbb{R}_+$, a *sum of squares* program of degree d is a program of the form:

$$\min \rho \quad \text{s.t.} \quad \forall x \in \{0, 1\}^n : \rho - f(x) = \sum_{i=1}^k g_i(x)^2;$$

$$\forall i \in [k] : \deg(g_i) \leq d/2.$$

The number ρ is called the *sum of squares upper bound* of degree d .

Original idea: verify that $\forall x : \rho - f(x) \geq 0$ using a sum of squares (which is always non-negative).

Clm: We can compute the sum of squares upper bound using a semidefinite program of size $n^{O(d)}$.

The semidefinite program is as follows: the variable matrix X is indexed by a pair of monomials α, β of degree at most $d/2$ each. The program itself is:

$$\min \rho \quad \text{s.t.} \quad \forall \gamma \in \mathbb{N}_{\leq d}^n : \sum_{\alpha, \beta \mid \alpha + \beta = \gamma} X_{\alpha, \beta} = [\gamma](\rho - f); \quad X \succeq 0.$$

L(Eye-opening lemma): A sum of squares semidefinite program of degree d is dual to the $t/2$ -th level of the Lasserre hierarchy.

Sum of squares vs. PSD rank

D: A *degree* of a function $\{0, 1\}^m \rightarrow \mathbb{R}$ will be the degree of the unique multilinear polynomial agreeing with f on every point of $\{0, 1\}^m$.

D(sos degree): Consider a non-negative function $f : \{0, 1\}^m \rightarrow \mathbb{R}_+$. We say that f has a *sum-of-squares certificate* of degree d if there exist functions $g_1, \dots, g_k : \{0, 1\}^m \rightarrow \mathbb{R}$ such that $\deg(g_i) \leq d/2$ and $f(x) = \sum_{i=1}^k g_i(x)^2$ for all $x \in \{0, 1\}^n$.

We say that f has *sos degree* d' and write $\deg_{\text{sos}}(f) = d'$ if d' is the minimal degree of a sum-of-squares certificate for f .

T(Main theorem): For every $m \geq 1$ and $f : \{0, 1\}^m \rightarrow \mathbb{R}_+$, there exists a constant $C > 0$ such that the following holds: For $n \geq 2m$, if $\deg_{\text{sos}}(f) = d + 2$, then

$$1 + n^{1+d/2} \geq \text{rk}_{\text{psd}}(M_n^f) \geq C \left(\frac{n}{\log n} \right)^{d/4}.$$

$f : \{0, 1\}^m \rightarrow \mathbb{R}$ has $\deg_{\text{sos}}(f) \geq m + 1$:

$$f(x) = \left(\frac{m}{2} - \sum_{i=1}^m x_i \right)^2 - 1/4.$$

The *Main theorem* gives us a good lower bound on f if $\deg_{\text{sos}}(f) = d + 2$. We finalize the proof of Theorem *Bound on psd rank* by applying the following:

T(Grigoriev): For every odd integer $m \geq 1$, the following function

Further reading

1. Bernd Gärtner and Jiří Matoušek: *Approximation Algorithms and Semidefinite Programming*, Springer, 2012
2. Monique Laurent: *Sums of squares, moment matrices and optimization over polynomials. Emerging applications of algebraic geometry*. Springer New York, 2009. 157-270.
3. Thomas Rothvoß: *The Lasserre hierarchy in approximation algorithms. Lecture Notes for MAPSP 2013, 2013*.