Lasserre Hierarchy, Higher Eigenvalues, and Approximation Schemes for Quadratic Integer Programming with PSD Objectives

Venkatesan Guruswami, Ali Kemal Sinop

Preprocessed and presented by Pavel Veselý

Combinatorics and Graph Theory PhD Seminar, 2014/15, MFF UK

Positive semidefiniteness

D: A symmetric matrix $M \in \mathbb{R}^{n\times n}$ is positive semidefinite (PSD) $\equiv \forall z \in \mathbb{R}^n : z^T M z \geq 0$. We write $M \succeq 0$. Equivalently:

- M is PSD iff there exists U such that $U^T U = M$.
- M is PSD iff any principal submatrix U of M has $\det(U) \geq 0$ (a submatrix U is principal iff we get it by removing a subset of rows and the same subset of columns).

Graph Partitioning / Labeling Problems

Generally, we deal with problems of assigning each vertex a label $i \in [k] = \{1, 2, \ldots k\}$ (i.e., cut the graph into k parts). We can formulate the problem as a quadratic integer program (QIP) which has a binary variable $x_u(i)$ for each vertex u and each label i.

D(Quadratic Integer Program with PSD costs): Given a PSD matrix $L \in \mathbb{R}^{(V\times[k])\times(V\times[k])}$, consider the problem of finding $x \in \{0,1\}^{V\times[k]}$ minimizing $x^T L x$ subject to:

1. exactly one of $\{x_u(i)\}_{i\in[k]}$ equals 1 for each u and
2. some linear constraints $Ax \geq b$.

Result: In time $n^{\mathcal{O}(r/\varepsilon^2)}$ we find such an x with

$$x^T L x \leq \frac{1+\varepsilon}{\min\{1, \lambda_r(\mathcal{L})\}} OPT$$

where $\mathcal{L} = \text{diag}(L)^{-1/2} \cdot L \cdot diag(L)^{-1/2}$ is a normalization of L and $\text{diag}(L)$ is the diagonal of L and $\lambda_r(\mathcal{L})$ is the r-th smallest eigenvalue of $\mathcal{L}$.

For the purpose of this talk, we consider only the Minimum Bisection problem on d-regular unweighted graphs, but a similar algorithm and analysis can be done for any such QIP with PSD costs.

D(Minimum Bisection):

Input: undirected d-regular graph $G = (V, E)$

Goal: Find $U \subset V$ with $|U| = \mu$ that minimizes $|\Gamma(U)|$ (where $\Gamma(U)$ is the set of edges with exactly one endpoint in U).

Big Picture

For our QIP with PSD costs we do the following:

1. relax QIP to the r'-th level of Lasserre SDP hierarchy,
2. solve the resulting SDP in $n^{\mathcal{O}(r')}$ (with an additive error ε_0),
3. randomly round the optimal solution of SDP to an integer solution.

Lasserre Hierarchy

Intuition: Take a polytope $K = \{x \in \mathbb{R}^n | Ax \geq b\}$ such that $K_I := \text{conv}(K \cap \{0,1\})^n$ precisely corresponds to **solutions** of a given problem. Then every point $x \in K_I$ corresponds to a **probability distribution** of valid solutions, since it is a convex combination of integral vertices of K_I. The problem is that in general $K \neq K_I$, so $x \in K$ may not correspond to a distribution of valid solutions.

What is wrong with $x \in K \setminus K_I$? It only gives us valid *marginal* probabilities, but it might not satisfy e.g. $\mathbb{P}[X_i = X_j = 1] \in [\max\{x_i + x_j - 1, 0\}, \min\{x_i, x_j\}]$ for $i \neq j$. What about introducing variables for every $I \subseteq [n], |I| = 2$? By adding appropriate constraints we can force any vector x to be a convex combination of some vectors integral on 2 coordinates. That would ensure the solution to be sensible w.r.t. second moments. Now generalize to higher moments...

Notation: Let $\mathcal{P}_r([n]) := \{I \subseteq [n] \mid |I| \leq r\}$ be the set of all index sets of cardinality at most r and let $y \in \mathbb{R}^{\mathcal{P}_{2r+2}([n])}$ be a vector with entries y_I for all $I \subseteq [n]$ with $|I| \leq 2r+2$. Intuitively $y_{\{i\}}$ represents the original variable x_i and the new variables y_I represent $\prod_{i\in I} x_i$.

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

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

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

$$M_r^\ell(y)_{I,J} := (\sum_{i=1}^{n} A_{\ell i} y_{I\cup J\cup\{i\}}) - b_\ell y_{I\cup J}$$

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

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

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

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

T(Lasserre properties): Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ and $y \in \text{Las}_r(K)$. Then the following holds:

(a) $\text{conv}(K \cap \{0,1\}^n) = \text{Las}_n^{\text{proj}}(K) \subseteq \ldots \subseteq \text{Las}_0^{\text{proj}}(K) \subseteq K$.

(b) We have $0 \leq y_I \leq y_J \leq 1$ for all $I \supseteq J$ with $0 \leq |J| \leq |I| \leq r$.

(c) Let $I \subseteq [n]$ with $|I| \leq r$. 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 $|I| \leq r$. Then

$$y \in \text{conv}(\{z \in \text{Las}_{r-|I|}(K) \mid z_{\{i\}} \in \{0,1\}\, \forall i \in I\}).$$

Vector representation: For each event $\bigcap_{i\in I}(x_i = 1)$ with $|I| \leq r$ there is a vector $\mathbf{x}_I$ representing it in a consistent way:

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

Lasserre for Labeling

In QIP for a graph labeling problem we have a binary variable for each vertex v and each labeling of v. In its Lasserre hierarchy relaxation instead of having a variable for each (small) subset of original variables it makes more sense to have a variable for each (small) subset S of vertices and each labeling f of S, denoted by $x_S(f)$. Furthermore, we will work in the vector representation of SDPs, i.e., $\mathbf{x}_S(f) \in \mathbb{R}^Y$ is a vector (in some dimension Y).

Notation: We use $\mathbf{x}_u(i)$ for singletons $u \in V$ and $i \in [k]$. For $f \in [k]^S$ and $v \in S$, let $f(v)$ be the label that v receives from f. For sets S with labeling $f \in [k]^S$ and T with labeling $g \in [k]^T$ such that f and g agree on $S \cap T$, we use $f \circ g$ to denote the labeling of $S \cup T$ consistent with f and g.

D: Given a set V of variables, a set $[k]$ of labels and $r \geq 0$, a **set of vectors x** is said to satisfy the r-level of the Lasserre hierarchy on k labels, denoted by

$$\mathbf{x} \in \text{LabelLas}_r(V \times [k])$$

if it satisfies the following conditions:

1. For each $S \in \binom{V}{\leq r+1}$ and $f \in [k]^S$ there is a vector $\mathbf{x}_S(f) \in \mathbb{R}^Y$.
2. $\|\mathbf{x}_\emptyset\|^2 = 1$.
3. $\langle \mathbf{x}_S(f), \mathbf{x}_T(g)\rangle = 0$ if there exists $u \in S \cap T$ s.t. $f(u) \neq g(u)$.
4. $\langle \mathbf{x}_S(f), \mathbf{x}_T(g)\rangle = \langle \mathbf{x}_{S'}(f'), \mathbf{x}_{T'}(g')\rangle$ if $S \cup T = S' \cup T'$ and $f \circ g = f' \circ g'$.
5. For any $u \in V$, $\sum_{j\in[k]} \|\mathbf{x}_u(j)\|^2 = 1$.
6. For any $S \in \binom{V}{\leq r+1}$, $u \in S$ and $f \in [k]^{S\setminus\{u\}}$ it holds $\sum_{j\in[k]} \mathbf{x}_S(f \circ j) = \mathbf{x}_{S\setminus\{u\}}(f)$.

We say that $\mathbf{x} \in \text{LabelLas}_r(V \times [k])$ satisfies m linear constraints $Ax \geq b$ where $A \in \mathbb{R}^{(V\times[k])\times m}$ if the following holds for all $\ell \in [m]$, all subsets $S \in \binom{V}{\leq r}$ and all $f \in [k]^S$:

$$\sum_{u\in V, j\in[k]} \langle \mathbf{x}_S(f), \mathbf{x}_u(j)\rangle A_{\ell,(u,j)} \geq b_\ell \|\mathbf{x}_S(f)\|^2 .$$

Approximating Minimum Bisection

For simplicity we assume that the graph is d-regular and unweighted. We relax the following QIP with $k = 2$:

$$\min \textstyle\sum_{\{u,v\}\in E} (x_u(1) - x_v(1))^2$$
$$\text{s.t. } \textstyle\sum_u x_u(1) = \mu, \forall u : x_u(1) + x_u(2) = 1, x \in \{0,1\}^{V\times[2]}$$

The corresponding Lasserre relaxation on level r': find a vector set $\mathbf{x} \in \text{LabelLas}_{r'}(V \times [2])$ minimizing

$$\sum_{\{u,v\}\in E} \|\mathbf{x}_u(1) - \mathbf{x}_v(1)\|^2 \qquad (1)$$

subject to

$$\forall S \in \binom{V}{\leq r'} \forall f \in [k]^S : \sum_u \langle \mathbf{x}_S(f), \mathbf{x}_u(1)\rangle = \mu \|\mathbf{x}_S(f)\|^2$$

D:If A is adjacency matrix of a graph, the Laplacian matrix is $L := D - A$ where D is a diagonal matrix with degrees on diagonal. The normalized Laplacian matrix is $\mathcal{L} := D^{-1/2} L D^{-1/2}$. Both L and $\mathcal{L}$ are PSD.

T:For any $r \geq 1$ and $\varepsilon > 0$ there exists $r' = \mathcal{O}(r/\varepsilon^2)$ such that given $\mathbf{x} \in \textsc{LabelLas}_{r'}(V \times [2])$ satisfying linear constraints $\sum_u \mathbf{x}_u(1) = \mu$ and with objective value OPT, we can find in randomized $n^{\mathcal{O}(1)}$ time a set $U \subset V$ satisfying with high probability:

1. $|\Gamma(U)| \leq \frac{1+\varepsilon}{min\{1,\lambda_r\}} OPT$ where λ_r is the r-th smallest eigenvalue of $\mathcal{L}$,
2. $\mu - \mathcal{O}(\sqrt{\mu \log(1/\varepsilon)}) \leq |U| \leq \mu + \mathcal{O}(\sqrt{\mu \log(1/\varepsilon)})$

Randomized Rounding

1. Fix a set S of r' vertices (to be chosen later).
2. Choose randomly a labeling $f \in [2]^S$ with probability $\|\mathbf{x}_S(f)\|^2$.
3. Independently for all $u \in V$ choose $i \in [2]$ and assign $x_u(i) \leftarrow 1$ with probability

$$\Pr[x_u(i) = 1] = \frac{\|\mathbf{x}_{S \cup \{u\}}(f \circ i)\|^2}{\|\mathbf{x}_S(f)\|^2}$$

.

4. Return $U = \{u \mid x_u(1) = 1\}$.

For convenience we use traces in the analysis:

D:Trace of a matrix M is the sum of elements on the main diagonal, i.e., $\text{tr}(M) := \sum_{i=0}^{n} M_{i,i}$. For matrices A, B it holds that $\text{tr}(A^T B) = \sum_{i,j} A_{i,j} B_{i,j}$.

Let $X \in \mathbb{R}^{Y \times V}$ be the matrix with columns corresponding to vectors $\mathbf{x}_u(1)$ for all $u \in V$. Then the objective value in Equation 1 can be rewritten as $\text{tr}(X^T X L)$ where L is the Laplacian matrix.

We also use projections on span of some vectors from the solution:

D:For a subset $S \in \binom{V}{\leq r'}$ we use the projection operator $\Pi_S \in \mathbb{R}^{Y \times Y}$ that projects vectors on the $\text{span}(\{\mathbf{x}_S(f)\}_{f \in [2]^S})$, i.e., $\Pi_S := \sum_{f \in [2]^S} \overline{\mathbf{x}_S(f)} \cdot \overline{\mathbf{x}_S(f)}^T$ where $\overline{\mathbf{x}_S(f)}$ is a unit vector in the direction of $\mathbf{x}_S(f)$ if $\mathbf{x}_S(f)$ is non-zero; otherwise $\overline{\mathbf{x}_S(f)} = 0$. Let $\Pi_S^{\perp} := I - \Pi_S$ be the orthogonal projection to Π_S.

D:Let P_S be the projection operator on $\text{span}(\{\mathbf{x}_v(j)\}_{v \in S, j \in [2]})$, i.e., $P_S := \sum_{v \in S, j \in [2]} \overline{\mathbf{x}_v(j)} \cdot \overline{\mathbf{x}_v(j)}^T$, and let $P_S^{\perp} := I - P_S$ be the orthogonal projection.

D:Finally, let X_S^{Π} be the projection operator on span of subset S of columns of X, i.e. $\text{span}(\{\mathbf{x}_v(1)\}_{v \in S})$ and $X_S^{\Pi} := \sum_{v \in S} \overline{\mathbf{x}_v(1)} \cdot \overline{\mathbf{x}_v(1)}^T$. Let $X_S^{\perp} := I - X_S^{\Pi}$ be the orthogonal projection.

Acknowledgments

The speaker thanks Martin Böhm for discussions about the Lasserre hierarchy (and also for this template).

Further reading

1. *Bernd Gärtner and Jiří Matoušek: Approximation Algorithms and Semidefinite Programming, Springer, 2012*
2. *Thomas Rothvoß: The Lasserre hierarchy in approximation algorithms. Lecture Notes for MAPSP 2013, 2013.* `http://math.mit.edu/~rothvoss/lecturenotes/lasserresurvey.pdf`

Appendix

T(Other Lasserre properties): Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ and $y \in \textsc{Las}_r(K)$. Then the following holds:

(a) Let $S \subseteq [n]$ be a subset of variables such that not many can be equal to 1 at the same time:

$$\max\{|I| : I \subseteq S; x \in K; x_i = 1 \ \forall i \in I\} \leq k < r.$$

Then we have

$$y \in \text{conv}(\{z \in \textsc{Las}_{r-k}(K) \mid z_{\{i\}} \in \{0,1\} \ \forall i \in S\}).$$

(b) For any $|I| \leq r$ we have $y_I = 1 \Leftrightarrow \bigwedge_{i \in I}(y_{\{i\}} = 1)$.

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

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

Remark: Property (a) is very strong and does not hold for the Sherali-Adams or Lovász-Schrijver hierarchy. 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 $\varepsilon \cdot OPT$). Another example 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.