Rounding Semidefinite Programming Hierarchies via Global Correlation

Boaz Barak, Prasad Raghavendra, David Steurer
Presented by Martin Böhm
Combinatorics and Graph Theory PhD Seminar, 2014, MFF UK

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 $y \in \mathbb{R}^{\mathcal{P}_{2t+2}([n])}$ be a vector with entries y_I for all $I \subseteq [n]$ with $|I| \leq 2t+2$.

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 $M_t^\ell(y) \in \mathbb{R}^{\mathcal{P}_t([n]) \times \mathcal{P}_t([n])}$:

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

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

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

Furthermore, let $\text{LAS}_t^{\text{proj}} := \{(y_{\{1\}}, \ldots, y_{\{n\}}) \mid y \in \text{LAS}_t(K)\}$ 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.

T(Lasserre properties from Martin K's lecture): Let $K = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ and $y \in \text{LAS}_t(K)$. 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 $0 \leq |J| \leq |I| \leq t$.

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

(e) Let $S \subseteq [n]$ be a subset of variables such that not many can be equal to 1 at the same time:
$$\max\{|I| : I \subseteq S; x \in K; x_i = 1 \, \forall i \in I\} \leq k < t.$$
Then we have
$$y \in \text{conv}(\{z \in \text{LAS}_{t-k}(K) \mid z_{\{i\}} \in \{0,1\} \, \forall i \in S\}).$$

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

(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 $y_{I \cup J} = y_J$.

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

L(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 $\|\mathbf{v}_\emptyset\|_2^2 = 1$.

From vectors to distributions

Binary setting

Solution in $x \in \text{conv}(K \cap \{0,1\}^n) \to$ a probability distribution over integral solutions in K. For t-round Lasserre we cannot have a globally feasible probability distribution, but instead one that is locally consistent.

L: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

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

General 2CSP setting

All 2CSP problems can be restated using SDPs with constraints hidden in the maximization clause, so we do not depend on the moment matrices.

D: Let $V = [n]$ be a set of vertices and $[k]$ the set of possible values. An *m-local distribution* is a distribution $\mathcal{D}^T$ over the set of assignments $[k]^T$ of the vertices of some set $T \subseteq V$ of size at most $m+2$. The choice $+2$ is for convenience.

D: A collection $\{\mathcal{D}^T | T \subseteq V, |T| \leq m+2\}$ of m-local distributions is *consistent* if all pairs of distributions $\mathcal{D}^T, \mathcal{D}^{T'}$ are consistent on their intersection $T \cap T'$. By this we mean that any event defined on $T \cap T'$ has the same probability in $\mathcal{D}^T$ and in $\mathcal{D}^{T'}$.

Notation trick: If we have n vertices and $|T| \leq m$, instead of the entire collection $\{\mathcal{D}^T | T \subseteq V, |T| \leq m+2\}$ we talk instead about a set of *m-local random variables* $X_1, X_2, \ldots, X_n$. We can think of those random variables as variables X_i coming from the distribution $\mathcal{D}^{\{i\}}$. Note that these variables are **not** jointly distributed random variables, but for each subset of at most $m+2$ of them, one can find a sample space $\mathcal{D}^T$ where the corresponding variables X_i^T are jointly distributed.

More notation.

- $\{X_i | X_S\} \equiv$ a random variable obtained by conditioning $X_i^{S \cup i}$ on variables $\{X_j^{(S \cup \{i\})} | j \in S\}$;
- $P[X_i = X_j | X_S] \equiv P[X_i^{S \cup i \cup j} = X_j^{S \cup i \cup j} | X_S^{S \cup i \cup j}]$.

D(Lasserre hierarchy in the prob. setting):

An m-round Lasserre solution of a 2CSP problem consists of m-local random variables $X_1, X_2, \ldots, X_n$ and vectors $v_{S,\alpha}$ for all $S \subseteq \binom{V}{m+2}$ and all local assignments $\alpha \in [k]^S$, if the following holds $\forall S, T \subseteq V, |S \cup T| \leq m+2, \forall \alpha \in [k]^S, \beta \in [k]^T$:

$$\langle v_{S,\alpha}, v_{T,\beta} \rangle = P[X_S = \alpha, X_T = \beta].$$

We usually want a solution for MAX 2CSP, so we add a maximization clause, for instance $\max P_{(i,j,\Pi) \in \mathcal{I}}[(x_i, x_j \in \Pi)]$.

O:A covariance matrix $E[(X - E[X])(X - E[X])^T]$ is always positive semidefinite for a random vector X.

C:For a fixed local assignment $x_S \in [k]^S$ (where $|S| \leq m$) and fixed a, b, it holds that the matrix $\left(\text{Cov}(X_{ia}, X_{jb} | X_S = x_S)\right)_{i,j \in V}$ is positive semidefinite for the m-th level of the Lasserre hierarchy.

Main results

D: The $\tau-$*threshold rank* of a regular graph G, denoted $\text{rank}_{\geq\tau}(G)$, is the number of eigenvalues of the normalized adjacency matrix of G that are larger than τ. We can define this for any MAX 2-CSP problem, by taking the adjacency graph of the predicates.

T: There is a constant c such that for every $\varepsilon > 0$, and every MAX 2-CSP instance $\mathcal{I}$ with objective value v and alphabet size k, the following holds:

The objective value $\text{sdpopt}(\mathcal{I})$ of the r-round Lasserre hierarchy for $r \geq k \cdot \text{rank}_{\geq\tau}(\mathcal{I})/\varepsilon^c$ is within ϵ of the objective value v of $\mathcal{I}$, i.e., $\text{sdpopt}(\mathcal{I}) \leq v + \varepsilon$.

Moreover, there exists a polynomial time rounding scheme that finds an assignment x satisfying $\text{val}_{\mathcal{I}}(x) > v - \varepsilon$ given optimal SDP solution as input.

T: There is an algorithm, based on rounding r rounds of the Lasserre hierarchy and a constant c, such that for every $\varepsilon > 0$ and input instance $\mathcal{I}$ of UNIQUE GAMES with objective value v, alphabet size k, satisfying $\text{rank}_{\geq\tau}(\mathcal{I}) \leq \varepsilon^c r/k$, where $\tau = \varepsilon^c$, the algorithm outputs an assignment x satisfying $\text{val}_{\mathcal{I}}(x) > v - \varepsilon$.

T: There is an algorithm, based on rounding r rounds of the Lasserre hierarchy and a constant c, such that for every $\varepsilon > 0$ and input UNIQUE GAMES instance $\mathcal{I}$ with objective value $1 - \varepsilon$ and alphabet size k, satisfying $r \geq ck \cdot \min\{n^{c\varepsilon^{1/3}}, \text{rank}_{\geq 1-c\varepsilon}(\mathcal{I})\}$, the algorithm outputs an assignment x satisfying $\text{val}_{\mathcal{I}}(x) > 1/2$.

A sample 2CSP: MaxCut

D:SDP relaxation of MAXCUT:

$$\text{maximize} \mathop{\mathbb{E}}_{i,j \in E} \|v_i - v_j\|^2 \text{ subject to } \|v\|_i^2 = 1 \, \forall i \in V.$$

Step 1. Use an m-round Lasserre to get a collection of m-local variables $X_1, X_2, \ldots, X_n$. For an edge ij, its contribution to the SDP objective is:

$$\mathop{\mathbb{P}}_{\mathcal{D}^{ij}}[X_i \neq X_j] = \|v_i - v_j\|^2.$$

Step 2. Our goal is sampling that is close to sampling $\mathcal{D}^{ij}$. Try first independent sampling from marginals $\mathcal{D}^i$.

O(Local correlation):On an edge (i,j), the local distribution $\mathcal{D}^{ij}$ is *far* from the independent sampling distribution $\mathcal{D}^i \times \mathcal{D}^j$ only if the random variables X_i, X_j are *correlated.*

O(Correlation helps): If two variables X_i, X_j are correlated, then sampling/fixing the value of X_i reduces the uncertainty in the value of X_j. More precisely:

$$\mathop{\mathbb{E}}_{\{X_i\}} \mathrm{Var}[X_j|X_i] = \mathrm{Var}[X_j] - \frac{1}{\mathrm{Var}[X_i]} \left[\mathrm{Cov}(X_i, X_j)\right]^2 .$$

The reduction in uncertainty is actually related to the global expected correlation:

$$\mathop{\mathbb{E}}_{j\in V} \mathrm{Var}[X_j] - \mathop{\mathbb{E}}_{i\in V} \mathop{\mathbb{E}}_{\{X_i\}} \left[\mathop{\mathbb{E}}_{j\in V} \mathrm{Var}[X_j|X_i]\right] \geq \mathop{\mathbb{E}}_{i,j\in V} |\,\mathrm{Cov}(X_i, X_j)|^2.$$

Step 3. Assume that average local correlation is at least ε, that is

$$\mathop{\mathbb{E}}_{ij\sim G} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \geq \varepsilon\,.$$

Use PSD of correlations, apply the following Lemma for vectors $v_i \equiv u_i^{\otimes 2}$:

L(Local Correlation vs. Global Correlation on Low-Rank Graphs): Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be vectors in the unit ball. Suppose that the vectors are correlated across the edges of a regular n-vertex graph G,

$$\mathop{\mathbb{E}}_{ij\sim G} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \geq \rho\,.$$

Then, the global correlation of the vectors is lower bounded by

$$\mathop{\mathbb{E}}_{i,j\in V} |\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle| \geq \Omega(\rho)/\mathrm{rank}_{\geq\Omega(\rho)}(G)\,.$$

where $\mathrm{rank}_{\geq\rho}(G)$ is the number of eigenvalues of adjacency matrix of G that are larger than ρ.

Step 4. If the independent sampling is at least $\varepsilon-$far from correlated sampling over the edges, we can use the previous Lemma and reduce the average variance. Therefore, after $rank_{\geq\varepsilon^2}(G)/\varepsilon^2$ steps, we are done.