Lift-and-Project Techniques and SDP Hierarchies

MFO seminar on Semidefinite Programming

Monique Laurent - CWI - Amsterdam & Tilburg University

May 30, 2010

Typical combinatorial optimization problem:

$$\max \ c^T x \ \text{ s.t. } Ax \leq b, \ x \in \{0,1\}^n$$

$P := \{x \in \mathbb{R}^n \mid Ax \leq b\}$ $\rightsquigarrow$ LP relaxation

$P_I := \mathrm{conv}(K \cap \{0,1\}^n)$ $\rightsquigarrow$ Integral polytope to be found

Goal: Construct a new relaxation P' such that $P_I \subseteq P' \subseteq P$, leading to P_I after finitely many iterations.

Gomory-Chvatal closure:

$$P' = \{x \mid u^T Ax \leq \lfloor u^T b \rfloor \ \forall u \geq 0 \text{ with } u^T A \text{ integer}\}.$$

P_I is found after $O(n^2 \log n)$ iterations if $P \subseteq [0,1]^n$.
[Eisenbrand-Schulz 1999]

But optimization over P' is hard! [Eisenbrand 1999]

Plan of the lecture

Goal: We present several techniques to construct a *hierarchy* of **LP/SDP** relaxations:

$$P \supseteq P_1 \supseteq \ldots \supseteq P_n = P_I.$$

$\rightsquigarrow$ Lovász-Schrijver N / N_+ operators	LP / SDP
$\rightsquigarrow$ Sherali-Adams construction	LP
$\rightsquigarrow$ Lasserre construction	SDP

Great interest recently in such hierarchies:

- **Polyhedral combinatorics:** How many rounds are needed to find P_I? Which valid inequalities are satisfied after t rounds?
- **Complexity theory:** What is the integrality gap after t rounds? Link to hardness of the problem?
- **Proof systems:** Use hierarchies as a model to generate inequalities and show e.g. $P_I = \emptyset$.

Lift-and-project strategy

1. **Generate new constraints: Multiply** the system $Ax \leq b$ by products of the constraints $x_i \geq 0$ and $1 - x_i \geq 0$.

   $\rightsquigarrow$ Polynomial system in x.

2. **Linearize** (and **lift**) by introducing new variables y_I for products $\prod_{i \in I} x_i$ and setting $x_i^2 = x_i$.

   $\rightsquigarrow$ Linear system in (x, y).

3. **Project** back on the x-variable space.

   $\rightsquigarrow$ LP relaxation P' satisfying $P_I \subseteq P' \subseteq P$.

Some notation

$$\text{Write} \quad Ax \leq b \quad \text{as} \quad a_\ell^T x \leq b_\ell \quad (\ell = 1, \ldots, m)$$

$$\text{or as} \quad g_\ell^T \begin{pmatrix} 1 \\ x \end{pmatrix} \geq 0 \quad (\ell = 1, \ldots, m)$$

setting $g_\ell = \begin{pmatrix} b_\ell \\ -a_\ell \end{pmatrix}$.

Homogenization of P:

$$\tilde{P} = \left\{ \lambda \begin{pmatrix} 1 \\ x \end{pmatrix} \mid \lambda \geq 0,\ x \in P \right\} = \left\{ y \in \mathbb{R}^{n+1} \mid g_\ell^T y \geq 0 \ (\ell = 1, \ldots, m) \right\}$$

$$V = \{1, \ldots, n\}.$$

The Lovász-Schrijver construction

1. **Multiply** $Ax \leq b$ by $x_i,\ 1 - x_i\ \ \forall i \in V$.

   $\rightsquigarrow$ Quadratic system: $g_\ell^T \begin{pmatrix} 1 \\ x \end{pmatrix} x_i,\ g_\ell \begin{pmatrix} 1 \\ x \end{pmatrix} (1 - x_i) \geq 0\ \ \forall i$

2. **Linearize:** Introduce the matrix variable $Y = \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^T$,

   indexed by $\{0\} \cup V$. Then, Y belongs to

$$\mathcal{M}(P) = \{Y \in \mathcal{S}_{n+1} \mid Y_{0i} = Y_{ii},\ Ye_i, Y(e_0 - e_i) \in \tilde{P}\ \ \forall i\},$$

$$\mathcal{M}_+(P) = \mathcal{M}(P) \cap \mathcal{S}_{n+1}^+.$$

3. **Project:**

$$N(P) = \left\{ x \in \mathbb{R}^V \mid \exists Y \in \mathcal{M}(P) \text{ s.t. } \begin{pmatrix} 1 \\ x \end{pmatrix} = Ye_0 \right\}$$

$$N_+(P) = \left\{ x \in \mathbb{R}^V \mid \exists Y \in \mathcal{M}_+(P) \text{ s.t. } \begin{pmatrix} 1 \\ x \end{pmatrix} = Ye_0 \right\}$$

Properties of the N- and N_+-operators

- $P_I \subseteq N_+(P) \subseteq N(P) \subseteq P$.
- $N(P) \subseteq \mathrm{conv}(P \cap \{x \mid x_i = 0, 1\})$ for all $i \in V$.
- $N^n(P) = P_I$.
- Assume one can optimize in polynomial time over P. Then the same holds for $N^t(P)$ and for $N_+^t(P)$ for any **fixed** t.

Example: Consider the ℓ_1-ball centered at $e/2$:

$$P = \left\{ x \in \mathbb{R}^V \mid \sum_{i \in I} x_i + \sum_{i \in V \setminus I} 1 - x_i \geq \frac{1}{2} \;\; \forall I \subseteq V \right\}.$$

Then: $P_I = \emptyset$, but $\frac{1}{2} e \in N_+^{n-1}(P)$.

$\rightsquigarrow$ n iterations of the N_+ operator are needed to find P_I

Application to stable sets [Lovász-Schrijver 1991]

$P = \mathrm{FRAC}(G) = \{x \in \mathbb{R}_+^V \mid x_i + x_j \leq 1 \ (ij \in E)\}$
$P_I = \mathrm{STAB}(G)$: stable set polytope of $G = (V, E)$.

- $N(\mathrm{FRAC}(G)) = \mathrm{FRAC}(G)$ intersected by the constraints:

  $\sum_{i \in V(C)} x_i \leq \frac{|C|-1}{2}$ for all odd circuits C.

- $Y \in \mathcal{M}(\mathrm{FRAC}(G)) \Longrightarrow y_{ij} = 0$ for edges $ij \in E$.

$\rightsquigarrow$ $N_+(\mathrm{FRAC}(G)) \subseteq \mathrm{TH}(G)$.

$\rightsquigarrow$ Any clique inequality $\sum_{i \in Q} x_i \leq 1$ is valid for $N_+(P)$, while its N-rank is $|Q| - 2$. $\rightsquigarrow$ **The N_+ operator helps!**

- $\frac{n}{\alpha(G)} - 2 \leq N\text{-rank} \leq n - \alpha(G) - 1$.

- N_+-rank $\leq \alpha(G)$ [equality if G = line graph of K_{2p+1}]

The Sherali-Adams construction

1. **Multiply** $Ax \leq b$ by $\prod_{i \in I} x_i \prod_{j \in J} (1 - x_j)$ for all disjoint $I, J \subseteq V$ with $|I \cup J| = t$.

2. **Linearize & lift:** Introduce new variables y_U for all $U \in \mathcal{P}_t(V)$, setting $x_i^2 = x_i$ and $y_i = x_i$.

3. **Project** back on x-variables space.

$\rightsquigarrow$ **Relaxation:** $\mathrm{SA}_t(P)$.

- **Then:** $\mathrm{SA}_1(K) = N(P), \quad \mathrm{SA}_t(P) \subseteq N(\mathrm{SA}_{t-1}(P))$.

  Thus: $\mathrm{SA}_t(P) \subseteq N^t(P)$.

Application to the matching polytope

For $G = (V, E)$, let $P = \{x \in \mathbb{R}^E_+ \mid x(\delta(v)) \leq 1 \ \forall v \in V\}$.

Then: P_I is the matching polytope (= stable set polytope of the line graph of G).

For $G = K_{2p+1}$:

- N_+-rank $= p$ [Stephen-Tunçel 1999]
- N-rank $\in [2p, p^2]$ [LS 1991] [Goemans-Tunçel 2001]
- SA-rank $= 2p - 1$ [Mathieu-Sinclair 2009]

  Detailed analysis of the integrality gap:

$$g_t = \frac{\max_{x \in \mathrm{SA}_t(P)} e^t x}{\max_{x \in P} e^T x} = \frac{\max_{x \in \mathrm{SA}_t(P)} e^t x}{p}.$$

$$g_t = \begin{cases} 1 + \frac{1}{2p} & \text{if } t \leq p - 1 \\ 1 & \text{if } t \geq 2p - 1 \\ \exists \text{ phase transition} & \text{at } 2p - \Theta(\sqrt{p}) \end{cases}$$

A canonical lifting lemma

- $$x \in \{0,1\}^n \rightsquigarrow \quad y^x = (\prod_{i\in I} x_i)_{I\subseteq V} \in \{0,1\}^{\mathcal{P}(V)}$$
  $$= (1, x_1, .., x_n, x_1x_2, .., x_{n-1}x_n, .., \prod_{i\in V} x_i)$$
- Z: matrix with columns y^x for $x \in \{0,1\}^n$.
- **Equivalently:** Z is the 0/1 matrix indexed by $\mathcal{P}(V)$ with
  $$Z(I,J) = 1 \text{ if } I \subseteq V, \ \ 0 \text{ else.}$$
  $$Z^{-1}(I,J) = (-1)^{|J\setminus I|} \text{ if } I \subseteq J, \ \ 0 \text{ else.}$$
- If $x \in P \cap \{0,1\}^n$, then $Y = y^x(y^x)^T$ satisfies:
  - $Y \succeq 0$
  - $Y_\ell = g_\ell(x) Y \succeq 0$ $\rightsquigarrow$ *localizing matrix*
  - $Y(I,J)$ depends only on $I \cup J$ $\rightsquigarrow$ *moment matrix*

$$y \in \mathbb{R}^{\mathcal{P}(V)} \rightsquigarrow Y = M_V(y) = (y_{I\cup J}), \ \ Y_\ell = M_V(g_\ell y)$$

Lemma: P_I is equal to the projection on the x-variable space of

$$\{y \in \mathbb{R}^{\mathcal{P}(V)} \mid y_0 = 1,\ M_V(y) \succeq 0,\ M_V(g_\ell y) \succeq 0\ \forall \ell\}.$$

Sketch of proof:

1. Verify that $M_V(y) = Z\ \mathrm{diag}(Z^{-1}y)Z^T$.
2. $M_V(y) \succeq 0 \Longrightarrow \lambda := Z^{-1}y \geq 0 \Longrightarrow y = Z\lambda = \sum_{x \in \{0,1\}^n} \lambda_x y^x$

   where $\sum_x \lambda_x = y_0 = 1$.
3. Use $M_V(g_\ell y) \succeq 0$ to show that

$$\lambda_x > 0 \Longrightarrow x \in P\ (\Longrightarrow x \in P_I).$$

$\rightsquigarrow$ **Each 0/1 polytope is projection of a simplex.**

Case $n = 2$

$$Z = \begin{array}{c} \\ \emptyset \\ 1 \\ 2 \\ 12 \end{array} \begin{array}{c} \begin{array}{cccc} \emptyset & 1 & 2 & 12 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \rightsquigarrow Z^{-1} = \begin{array}{c} \\ \emptyset \\ 1 \\ 2 \\ 12 \end{array} \begin{array}{c} \begin{array}{cccc} \emptyset & 1 & 2 & 12 \end{array} \\ \begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$M_V(y) = \begin{pmatrix} y_0 & y_1 & y_2 & y_{12} \\ y_1 & y_1 & y_{12} & y_{12} \\ y_2 & y_{12} & y_2 & y_{12} \\ y_{12} & y_{12} & y_{12} & y_{12} \end{pmatrix} \succeq 0 \Longleftrightarrow \left\{ \begin{array}{r} y_0 - y_1 - y_2 + y_{12} \geq 0 \\ y_1 - y_{12} \geq 0 \\ y_2 - y_{12} \geq 0 \\ y_{12} \geq 0 \end{array} \right.$$

SDP hierarchies

Idea: Get SDP hierarchies by **truncating** $M_V(y)$ and $M_V(g_\ell y)$:

- Consider $M_U(y) = (y_{I\cup J})_{I,J\subseteq U}$, indexed by $\mathcal{P}(U)$ for $U \subseteq V$,
- or $M_t(y) = (y_{I\cup J})_{|I|,|J|\le t}$, indexed by $\mathcal{P}_t(V)$ for some $t \le n$.

1. **(local)** Get the Sherali-Adams relaxation $\mathrm{SA}_t(P)$ when considering

   $$M_U(y) \succeq 0,\ M_W(g_\ell y) \succeq 0\ \ \forall U \in \mathcal{P}_t(V),\ W \in \mathcal{P}_{t-1}(V).$$

   $\rightsquigarrow$ LP with variables y_I for all $I \in \mathcal{P}_t(V)$

2. **(global)** Get the Lasserre relaxation $\mathrm{L}_t(P)$ when considering

   $$M_t(y) \succeq 0,\ M_{t-1}(g_\ell y) \succeq 0.$$

   $\rightsquigarrow$ SDP with variables y_I for all $I \in \mathcal{P}_{2t}(V)$

Obviously: $\mathrm{L}_t(P) \subseteq \mathrm{SA}_t(P)$.

Link to the Lovász-Schrijver construction

- $\mathrm{L}_1(P) \subseteq P, \quad \mathrm{L}_t(P) \subseteq N_+(\mathrm{L}_{t-1}(P))$.
- **Thus:** $\mathrm{L}_t(P) \subseteq N_+^{t-1}(P)$.

  $\mathrm{L}_t(P)$ is tighter but more expensive to compute!

  • The SDP for $\mathrm{L}_t(P)$ involves one matrix of order $O(n^t)$, $O(n^{2t})$ variables.

  • The SDP for $N_+^{t-1}(P)$ involves $O(n^{t-1})$ matrices of order $n+1$, $O(n^{t+1})$ variables.
- **Note:** One can define a (block-diagonal) hierarchy, in-between and cheaper than both $\mathrm{L}_t(P)$ and $N_+^{t-1}(P)$; roughly,
  • 'unfold' the recursive definition of the LS hierarchy, and
  • consider suitably defined principal submatrices of $M_t(y)$ (which can be block-diagonalized to blocks of order $n+1$).
  [Gvozdenovic-L-Vallentin 2009]

Application of the Lasserre construction to stable sets

- The localizing conditions in $L_t(\mathrm{FRAC}(G))$ boil down to the edge conditions: $y_{ij} = 0$ $(ij \in E)$ (for $t \geq 2$).

  $\rightsquigarrow$ Natural generalization of the theta body $\mathrm{TH}(G)$.

  $\rightsquigarrow$ Get the bound $\mathrm{las}^{(t)}(G)$.

- **Convergence in $\alpha(G)$ steps:**

$$L_t(\mathrm{FRAC}(G)) = \mathrm{STAB}(G) \quad \text{for} \quad t \geq \alpha(G).$$

Open: Exist graphs G for which $\alpha(G)$ steps are needed?

Question: What is the Lasserre rank of the matching polytope?

Application of the Lasserre construction to Max-Cut

Max-Cut: $\max \sum_{ij \in E} w_{ij} \frac{1 - x_i x_j}{2}$ s.t. $x \in \{\pm 1\}^V$.

Consider $P = [-1, 1]^V$, write $x_i^2 = 1$, and project onto the subspace $\mathbb{R}^{\binom{n}{2}}$ indexed by edges.

- The order 1 relaxation is the basic GW relaxation:

$$\max \sum_{ij \in E} w_{ij} \frac{1 - X_{ij}}{2} \text{ s.t. } X \in \mathcal{S}_n^+, \ \mathrm{diag}(X) = e.$$

- The Lasserre rank of $\mathrm{CUT}(K_n)$ is at least $n/2$. [La 2003]

  (First time when $\sum_{ij \in E(K_n)} x_{ij} \geq -\lfloor n/2 \rfloor$ becomes valid).

  Question: Does equality hold? (**Yes** for $n \leq 7$).

The Lasserre relaxation of order 2 relaxation satisfies the triangle inequalities:

$$
Y = \begin{array}{c} \\ \emptyset \\ 12 \\ 13 \\ 23 \end{array}
\begin{array}{c} \begin{array}{cccc} \emptyset & 12 & 13 & 23 \end{array} \\
\begin{pmatrix} 1 & y_{12} & y_{13} & y_{23} \\ y_{12} & 1 & y_{23} & y_{13} \\ y_{13} & y_{23} & 1 & y_{12} \\ y_{23} & y_{13} & y_{12} & 1 \end{pmatrix} \end{array} \succeq 0
$$

$$\Longrightarrow e^T Y e \geq 0$$

$$\Longrightarrow y_{12} + y_{13} + y_{23} \geq 1.$$

Some negative results about integrality gaps of hierarchies for max-cut

Consider the basic LP relaxation of max-cut defined by the triangle inequalities.

$\rightsquigarrow$ Its integrality gap is $1/2$.

- [Schoenebeck-Trevisan-Tulsiani 2006] For the Lovász-Schrijver construction:

  • The integrality gap remains $1/2 + \epsilon$ after $c_\epsilon n$ rounds of the N operator.

  • But the integrality gap is 0.878 after one round of the N_+ operator.

- [Charikar-Makarychev-Makarychev 2009] For the Sherali-Adams construction:

  • The integrality gap remains $1/2 + \epsilon$ after n^{γ_ϵ} iterations.

Some positive results

Chlamtac-Singh [2008] give (for the first time) an approximation algorithm whose approximation guarantee improves indefinitely as one uses higher order relaxations in the SDP hierarchy:

$\rightsquigarrow$ For the maximum independent set problem in a 3-uniform hypergraph G.

Namely: Given $\gamma > 0$, assuming G contains an independent set of cardinality γn, then one can find an independent set of cardinality $n^{\Omega(\gamma^2)}$ using the relaxation of order $\Theta(1/\gamma^2)$.

Extensions to optimization over polynomials

- Minimize $p(x)$ over $\{x \mid g_j(x) \geq 0\}$.

  • Linearize $p = \sum_\alpha p_\alpha x^\alpha$ by $\sum_\alpha p_\alpha y_\alpha$.

  • Impose SDP conditions on the moment matrix:
  $M_t(y) = (y_{\alpha+\beta}) \succeq 0$.

  $\rightsquigarrow$ hierarchy of SDP relaxations with asymptotic convergence (due to some SOS representation results).

- Exploit equations: $h_j(x) = 0$.

  • We saw how to exploit $x_i^2 = x_i$.

  • The 'canonical lifting' lemma extends to the finite variety case: when the equations $h_j = 0$ have finitely many roots.

  • Finite convergence of the hierarchy when the equations $h_j = 0$ have finitely many real roots.

Another hierarchy construction via copositive programming

Reformulation: $\alpha(G) = \min \ \lambda \ \text{ s.t. } \lambda(I + A_G) - J \in \mathcal{C}_n,$ where

$\mathcal{C}_n = \{M \in \mathcal{S}_n \mid x^T M x \geq 0 \ \forall x \in \mathbb{R}^n_+\}$ is **the copositive cone**.

Idea [Parrilo 2000]: Replace $\mathcal{C}_n$ by the subcones

$$\mathcal{L}_n^{(t)} = \{M \in \mathcal{S}_n \mid (x^T M x)\Big(\sum_{i=1}^n x_i\Big)^r \text{ has non-negative coefficients}\},$$

$$\mathcal{K}_n^{(t)} = \{M \in \mathcal{S}_n \mid \Big(\sum_{i,j=1}^n M_{ij} x_i^2 x_j^2\Big)\Big(\sum_{i=1}^n x_i^2\Big)^t \text{ is SOS}\},$$

$$\mathcal{L}_n^{(t)} \subseteq \mathcal{K}_n^{(t)} \subseteq \mathcal{C}_n.$$

[Pólya] If M is strictly copositive then $M \in \bigcup_{t \geq 0} \mathcal{L}_n^{(t)}$.

$\rightsquigarrow$ **LP bound:** $\nu^{(t)}(G) = \min\ \lambda\ \text{ s.t. } \lambda(I + A_G) - J \in \mathcal{L}_n^{(t)}$,

$\rightsquigarrow$ **SDP bound:** $\vartheta^{(t)}(G) = \min\ \lambda\ \text{ s.t. } \lambda(I + A_G) - J \in \mathcal{K}_n^{(t)}$.

- $\nu^{(t)}(G) < \infty \iff t \geq \alpha(G) - 1$.
- $\lfloor \nu^{(t)}(G) \rfloor = \alpha(G)$ if $t \geq \alpha(G)^2$.
- $\vartheta^{(0)}(G) = \vartheta'(G)$
- **Conjecture:** [de Klerk-Pasechnik 2002]

$$\vartheta^{(t)}(G) = \alpha(G) \ \text{ for } t \geq \alpha(G) - 1.$$

Yes: For graphs with $\alpha(G) \leq 8$ [Gvozdenovic-La 2007]

- The Lasserre hierarchy refines the copositive hierarchy:

$$\text{las}^{(t+1)}(G) \ \leq \ \vartheta^{(t)}(G).$$

Note: The convergence in $\alpha(G)$ steps was easy for the Lasserre hierarchy!