Solution methods for SDP arising from combinatorial optimization problems

Franz Rendl

http://www.math.uni-klu.ac.at

Alpen-Adria-Universität Klagenfurt

Austria

Overview

- Part 1:
  How Semidefinite Problems arise as relaxations of Combinatorial Optimization Problems
- Part 2:
  How SDP can be solved: from 'safe' techniques (Interior-Point Technology) to more advanced nonlinear techniques, suitable also for large scale problems (but with weaker convergence properties)

Overview (Part 1)

- Semidefinite Programming (SDP) Basics
- Modeling with SDP
  - Graph Partition Problems
  - Stable Sets, Cliques
  - Coloring
  - Quadratic Assignment Problems
- Tightening by Cutting Planes

Semidefinite Programs

$$\max\{\langle C, X\rangle : A(X) = b, X \succeq 0\} = \ \min\{b^T y : A^T(y) - C = Z \succeq 0\}$$

This is a linear optimization problem over the cone of semidefinite matrices. Such problems are called semidefinite optimization problems (SDP).

Increased interest since early 1990's, due to success of interior-point methods.
SDP are convex optimization problems. SDP are powerful tool in many areas of applied mathematics.
SDP-based models are often much stronger than purely polyhedral relaxations.
General Reference: Handbook on SDP (Kluwer 2000), by Wolkowicz et al.

Semidefinite Programs (2)

Some notation and assumptions:

$$X, Z \text{ symmetric } n \times n \text{ matrices}$$

The linear equations $A(X) = b$ read $\langle A_i, X \rangle = b_i$ for given symmetric matrices $A_i, i = 1, \ldots, m$.
The adjoint map A^T is given by $A^T(y) = \sum y_i A_i$.
It is defined through

$$\langle y, A(X) \rangle = \langle A^T(y), X \rangle \ \ \forall X, y.$$

How derive the dual ?

SDP duality

Use Lagrange dual and Minimax Inequality to get Weak duality :

$$\sup_{A(X)=b,\ X\succeq 0} \langle C, X\rangle = \sup_{X\succeq 0} \inf_{y} \langle C, X\rangle + y^T(b - A(X))$$

SDP duality

Use Lagrange dual and Minimax Inequality to get Weak duality :

$$\sup_{A(X)=b,\ X\succeq 0} \langle C, X\rangle = \sup_{X\succeq 0} \inf_{y} \langle C, X\rangle + y^T(b - A(X))$$

$$\leq \inf_{y} \sup_{X\succeq 0} b^T y + \langle C - A^T(y), X\rangle$$

SDP duality

Use Lagrange dual and Minimax Inequality to get Weak duality :

$$\sup_{A(X)=b,\ X\succeq 0} \langle C, X\rangle = \sup_{X\succeq 0} \inf_{y} \langle C, X\rangle + y^T(b - A(X))$$

$$\leq \inf_{y} \sup_{X\succeq 0} b^T y + \langle C - A^T(y), X\rangle$$

$$= \inf_{y,\ A^T(y)-C\succeq 0} b^T y.$$

In general, $\sup$ and $\inf$ need not be attained, there can be strict inequality after exchanging $\sup$ and $\inf$ and also a finite (nonzero) duality gap between primal and dual value.

Strong duality

Strong duality (primal=dual and optima are attained) holds if we assume that both the primal and the dual problem have strictly feasible points ($X, Z \succ 0$).
Then it follows from the general Karush-Kuhn-Tucker theory that (X, y, Z) is optimal if and only if

$$A(X) = b, \ X \succeq 0, \quad A^T(y) - Z = C, \ Z \succeq 0, \quad \langle X, Z \rangle = 0.$$

We have $m + \binom{n+1}{2} + 1$ equations, and $m + 2\binom{n+1}{2}$ variables.

Strong duality

Strong duality (primal=dual and optima are attained) holds if we assume that both the primal and the dual problem have strictly feasible points ($X, Z \succ 0$).
Then it follows from the general Karush-Kuhn-Tucker theory that (X, y, Z) is optimal if and only if

$$A(X) = b,\ X \succeq 0,\ \ A^T(y) - Z = C,\ Z \succeq 0,\ \ \langle X, Z\rangle = 0.$$

We have $m + \binom{n+1}{2} + 1$ equations, and $m + 2\binom{n+1}{2}$ variables.
$X,\ Z \succeq 0$ means $X = UU^T,\ Z = VV^T$, so we conclude that $0 = \langle X, Z\rangle = \|U^T V\|^2$ implies

$$ZX = UU^T VV^T = 0$$

Now too many equations as ZX need not be symmetric.

Polyhedral versus Semidefinite approach

Polyhedral approach to IP:

• study convex hull of (characteristic vectors x_F) of feasible solutions

$$conv\{x_F : F \text{ feasible}\}.$$

Semidefinite Programming (SDP) approach:

• Move from $x_F \in \mathbb{R}^n$ to symmetric matrices $x_F x_F^T$ and study

$$conv\{x_F x_F^T : F \text{ feasible}\}.$$

• Fact 1: Contained in the cone of semidefinite matrices.
• Fact 2: Anything quadratic in x will be linear in the matrix space.

The Max-Cut Problem

Unconstrained quadratic 1/-1 optimization:

$$\max \ x^T L x \text{ such that } x \in \{-1,1\}^n$$

This is Max-Cut as a binary quadratic problem. Graph interpretation: $G = (V, E)$ edge-weighted graph, with weighted adjacency matrix A. Define Laplacian $L = L_A$ as $L = \text{Diag}(Ae) - A$. $S = \{i : \ x_i = 1\}, \ T = \{i : \ x_i = -1\}$ gives bisection. Total weight of edges joining S and T is to be maximized.
Same as unconstrained quadratic 0/1 minimization:

$$\min \ x^T Q x + c^T x \text{ such that } x \in \{0,1\}^n$$

Q upper triangular, or symmetric with zero diagonal.

SDP relaxation for Max-Cut

Linearize (and simplify) to get tractable relaxation
$x^T Lx = \langle L, xx^T \rangle$. New variable is $X = xx^T$.
Basic SDP relaxation: (e ... all-ones vector)

$$\max\{\langle L, X\rangle : \text{diag}(X) = e,\ X \succeq 0\}$$

This model goes back to Schrijver.
See also Poljak, R. (1995) primal-dual formulation, and Goemans, Williamson (1995) for the hyperplane rounding analysis.

SDP relaxation for Max-Cut

Linearize (and simplify) to get tractable relaxation
$x^T Lx = \langle L, xx^T \rangle$. New variable is $X = xx^T$.
Basic SDP relaxation:

$$\max\{\langle L, X\rangle : \text{diag}(X) = e,\ X \succeq 0\}$$

This model goes back to Schrijver.
See also Poljak, R. (1995) primal-dual formulation, and Goemans, Williamson (1995) for the hyperplane rounding analysis.

0/1 version:
Relax $X - xx^T = 0,\ \text{diag}(X) = x$ by (convex) constraint:
$X - xx \succeq 0,\ \text{diag}(X) = x$. Resulting SDP relaxation equivalent to Max-Cut, see Helmberg (1997).

Bisection and Equicut

In Max-Cut, the number of $+1$ in x is not constrained.

If $|S|, |T|$ (cardinalities of partition blocks) are specified, this can be modeled as a constraint on $\sum x_i$.

If $\sum_i x_i = 0$ there are as many $+1$ as -1 in x. In terms of partition, there is an equal number of nodes in each bisection block.

Equicut:

$$\min \ x^T L x \text{ such that } x \in \{-1, 1\}^n, \ e^T x = 0.$$

This problem has been investigated by Kernighan, Lin 1972

SDP relaxation of Equicut

$$e^T x = 0 \iff (e^T x)^2 = 0 \iff \langle ee^T, xx^T \rangle = 0.$$

This translates into

$$\langle J, X \rangle = 0 \text{ with } J = ee^T.$$

SDP relaxation for Equicut:

$$\min\{\langle L, X \rangle : \text{diag}(X) = e,\ \langle J, X \rangle = 0,\ \ X \succeq 0\}$$

Note that $\langle J, X \rangle = 0,\ X \succeq 0$ implies $\lambda_{\min}(X) = 0$, hence there is no strictly feasible solution X to this SDP.

k-equipartition

Partitioning the nodes of a graph into $k > 2$ sets of equal cardinality can be modeled more easily with 0-1 variables. $X \ldots n \times k$ models incidence vectors of the partition blocks $1, \ldots, k$.

$$\min\{\text{tr} X^T L X : X \text{ partition matrix }\}$$

Linearization idea: $XX^T = Y \succeq 0$. The following conditions must hold for Y, if $n = km$ and we partition into k sets of cardinality m:

$$\text{diag}(Y) = e, \;\; Ye = me$$

This leads to SDP relaxation

$$\min\langle L, Y\rangle \text{ s.t. } \text{diag}(Y) = e, \; Ye = me, \; Y \succeq 0.$$

Similar to Max-Cut but with eigenvector condition $Ye = me$.

Stable sets

$e\ldots$ all-ones vector, and $J = ee^T \ldots$ all-ones matrix.

$$\alpha(G) = \max \sum_i x_i \text{ such that } x_i x_j = 0 \; ij \in E, \; x_i \in \{0,1\}$$

Linearization trick: Consider $X = \frac{1}{x^T x} x x^T$. X satisfies:
tr$(X) = 1$ and $e^T x = x^T x$, so $e^T x = \langle J, X \rangle$. Hence,

$$\alpha(G) = \max \langle J, X \rangle \text{ such that } x_{ij} = 0 \; \forall ij \in E,$$

$$X = \frac{1}{x^T x} x x^T, \; x \in \{0,1\}^n$$

Stable sets

$e \dots$ all-ones vector, and $J = ee^T \dots$ all-ones matrix.
Linearization trick: Consider $X = \frac{1}{x^T x} x x^T$. X satisfies:
tr$(X) = 1$ and $e^T x = x^T x$, so $e^T x = \langle J, X \rangle$. Hence,

$$\alpha(G) = \max \langle J, X \rangle \text{ such that } x_{ij} = 0 \ \forall ij \in E,$$

$$X = \frac{1}{x^T x} x x^T, \ x \in \{0,1\}^n$$

After eliminating x it is easy to see

$$\alpha(G) = \max \langle J, X \rangle \text{ such that } x_{ij} = 0 \ \forall ij \in E,$$

$$\text{tr}(X) = 1, \ X \in PSD, \ rank(X) = 1.$$

Proof:

- $X = vv^T$ for some vector v.
- $x_{ij} = 0$ on edges ij implies that support of v is nonzero only on some stable set S of G.
- Looking at nonzero part v_S of v, the maximization of $(v^T e)^2$ forces v_S to be parallel to e.
- Therefore X is multiple of vv^T where v is characteristic vector of some stable set.

Proof:

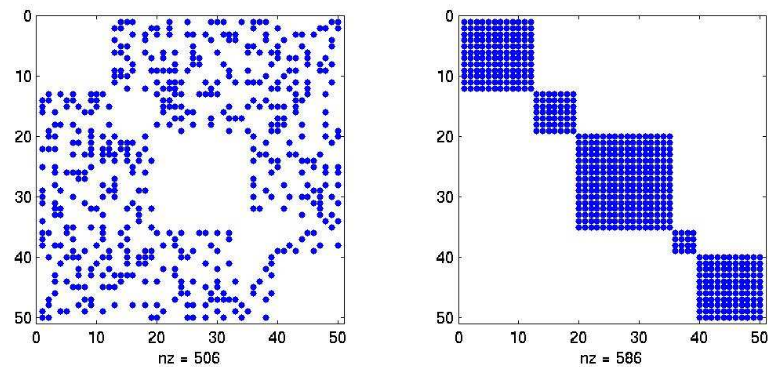
• $X = vv^T$ for some vector v.
• $x_{ij} = 0$ on edges ij implies that support of v is nonzero only on some stable set S of G.
• Looking at nonzero part v_S of v, the maximization of $(v^T e)^2$ forces v_S to be parallel to e.
• Therefore X is multiple of vv^T where v is characteristic vector of some stable set.

Leaving out the rank condition on X, we get the Theta number of Lovasz (1979):

$$\vartheta(G) := \max\{\langle J, X\rangle : X \succeq 0,\ \text{tr}(X) = 1,\ \ x_{ij} = 0\ (ij) \in E\}$$

This SDP has $m + 1$ equations, if $|E| = m$.

More SDP modeling: Graph Coloring



Adjacency matrix A of a graph (left), associated Coloring Matrix (right). The graph can be colored with 5 colors.

- M is coloring matrix if $\exists P \in \Pi$ such that $P^T MP$ is direct sum of all-ones blocks and $m_{ij} = 0$ if $[ij] \in E(G)$.
- Number of colors = number of all-ones blocks = rank of M.

Chromatic number

- M is coloring matrix if $\exists P \in \Pi$ such that $P^T M P$ is direct sum of all-ones blocks and $m_{ij} = 0$ if $[ij] \in E(G)$.
- Number of colors = number of all-ones blocks = rank of M.

Therefore chromatic number $\chi(G)$ of graph G can be defined as follows:

$$\chi(G) = \min\{rank(M) : M \text{ is coloring matrix of } G\}.$$

We need a 'better' description of coloring matrices.

More on Coloring Matrices

Lemma: M is coloring matrix if and only if

$$M = M^T,\ m_{ij} \in \{0,1\}, m_{ij} = 0\ (ij) \in E,$$

$$(tM - J \succeq 0\ \Leftrightarrow\ t \geq rank(M)).$$

Proof:

$\Rightarrow$: Nonzero principal minor of $tM - J$ has form $tI_s - J_s$ and $s \leq rank(M)$. Hence $tM - J \succeq 0$ iff $t \geq rank(M)$.

$\Leftarrow$: $m_{ii} = 1$ (so each vertex in one color class).

$m_{ij} = m_{jk} = 1$ implies $m_{ik} = 1$ because $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \not\succeq 0.$

Therefore M is direct sum of all-ones blocks.

Chromatic number

Hence

$$\chi(G) = \min\{rank(M) : M \text{ is coloring matrix of } G\} =$$

$$\min\{t : M = M^T, m_{ij} \in \{0,1\}, m_{ij} = 0 \forall ij \in E(G), tM - J \succeq 0\},$$

using the previous lemma.

Chromatic number

Hence

$$\chi(G) = \min\{rank(M) : M \text{ is coloring matrix of } G\} =$$

$$\min\{t : M = M^T, m_{ij} \in \{0,1\}, m_{ij} = 0 \forall ij \in E(G), tM - J \succeq 0\},$$

using the previous lemma.
Leaving out $m_{ij} \in \{0,1\}$ gives SDP lower bound:

$$\chi(G) \geq \min\{t : Y - J \succeq 0,\ y_{ii} = t\ \forall i,\ y_{ij} = 0\ ij \in E(G)\} = \vartheta(G).$$

This gives the second inequality in the Lovasz sandwich theorem, Lovasz (1979):
$\omega(G) \leq \vartheta(G) \leq \chi(G)$.
The first inequality can be derived in a similar way, using the dual SDP.

Quadratic Assignment Problem (QAP)

(QAP) $\min\langle AXB + C, X\rangle$ such that X is permutation matrix

Using $x =$ vec$(X),\; x \circ x = x$ we get

$$\langle AXB + C, X\rangle = \langle B \otimes A + \text{Diag}(vec(C)), xx^T\rangle$$

Now linearize $Y = xx^T$ to get SDP or COP relaxations.

A technical problem:
How translate permutation properties from x to Y?

$$X = (x_1, \ldots, x_n), Y = \begin{pmatrix} Y^{11} & \ldots & Y^{1n} \\ \vdots & & \vdots \\ Y^{n1} & \ldots & Y^{nn} \end{pmatrix}, Y^{ij} = x_i x_j^T$$

QAP (2)

$$\sum_i Y^{ii} = \sum_i x_i x_i^T = I, \quad \text{tr}(Y^{ij}) = x_i^T x_j = \delta_{ij}$$

$$\langle J, Y \rangle = (e^T x)(x^T e) = n^2$$

X is orthogonal, sums of all elements =n.

$$\mathcal{F} := \{Y \succeq 0, \sum_i Y^{ii} = I, \text{tr}(Y^{ij}) = \delta_{ij}, \langle J, Y \rangle = n^2\}$$

The last condition can also be written out for each block Y^{ij} as

$$\langle J, Y^{ij} \rangle = 1.$$

Note that Y is $n^2 \times n^2$.

SDP relaxation of QAP

$$L = B \otimes A + \text{Diag}(vec(C)), \; Y \text{ as before.}$$

$$\min \langle L, Y \rangle : Y \in \mathcal{F}$$

Further constraints possible:

$$Y \geq 0 \quad O(n^4) \text{ sign constraints !!}$$

$$Y_{ij,ik} = Y_{ik,jk} = 0 \text{ for all } i, j \neq k \quad O(n^3) \text{ equations}$$

This leads to SDP which could only be solved very recently using refined nonlinear techniques, see Sun, Toh, Zhang (working paper, 2008).

Tightening with Cutting Planes

Recall SDP relaxation of Max-Cut:
$\max\langle L, X\rangle : \text{ diag}(X) = e, \ \ X \succeq 0$.
Can be further tightened by Combinatorial Cutting Planes:

A simple observation:
Barahona, Mahjoub (1986): Cut Polytope, Deza, Laurent (1997): Hypermetric Inequalities

$$x \in \{-1, 1\}^n, \ \ f = (1, 1, 1, 0, \ldots, 0)^T \ \ \Rightarrow |f^T x| \geq 1.$$

Results in $x^T f \ f^T x = \langle (xx^T), (ff^T)\rangle = \langle \mathbf{X}, \mathbf{ff^T}\rangle \geq \mathbf{1}$.

Can be applied to any **triangle** $i < j < k$.
Nonzeros of f can also be -1.

Max-Cut and Triangles

There are $4\binom{n}{3}$ such triangle inequality constraints, which we collect in $B(X) \leq b$.

SDP relaxation of Max-Cut with triangles:

$$\max\langle L, X\rangle : \quad \text{diag}(X) = e, \quad B(X) \leq b, \quad X \succeq 0$$

Direct application of standard methods not possible for $n \approx 100$, because there are about $\frac{2}{3}n^3$ inequalities.

Can also be applied to other partition problems, and to relaxatons for stable sets and coloring.

This gives tighter relaxation, but the SDP becomes much more difficult to solve.