Active set algorithms for conic programming

John E. Mitchell[1] Kartik K. Sivaramakrishnan[2]

[1]Department of Mathematical Sciences
RPI, Troy, NY 12180 USA

[2]Department of Mathematics
North Carolina State University
Raleigh, NC 27695 USA

SIAM Conference on Optimization
Boston
May 13, 2008

Outline

Applications of Semidefinite Programming

- Combinatorial optimization: Obtain stronger relaxations than using linear programming for MaxCut, Satisfiability, Independent Set,...
- Control theory
- Structural optimization
- Electronic structure calculation and ground states. (Fukuda et al)
- Relaxations of other optimization problems: For example, nonconvex quadratically constrained quadratic programs. (Burer & Vandenbussche, Braun & M, Tseng, ...)
- Statistics: For example, principal component analysis. (d'Aspremont et al)

Applications of Semidefinite Programming

- Combinatorial optimization: Obtain stronger relaxations than using linear programming for MaxCut, Satisfiability, Independent Set,...
- Control theory
- Structural optimization

- Electronic structure calculation and ground states. (Fukuda et al)
- Relaxations of other optimization problems: For example, nonconvex quadratically constrained quadratic programs. (Burer & Vandenbussche, Braun & M, Tseng, ...)
- Statistics: For example, principal component analysis. (d'Aspremont et al)

Semidefinite Programming

A semidefinite program has a linear objective function and linear constraints.

The variables can be written in the form of a matrix.

This matrix of variables is constrained to be positive semidefinite.

SDPs can be solved in polynomial time using interior point methods.

SDPs generalize linear programming.

Can be generalized further to conic programming:

> *linear objective function, linear constraints, the variables belong to a convex cone.*

Outline

Why cutting planes and surfaces?

- Interior point algorithms solve SDPs in polynomial time.
- But, computational time becomes impractical for larger problems.

- Alternatively, use relaxations or approximations to get a good solution reasonably quickly.
- These relaxation approaches will also typically give an indication of a final duality gap.

- Other alternative approaches include:
  Spectral bundle method of Helmberg, Rendl, and Kiwiel.
  Augmented Lagrangian approaches of Burer, Monteiro, Zhang.
  First order methods based on Nesterov's smoothing scheme.
  Methods exploiting sparsity and symmetry by Fukuda et al and by De Klerk et al.

Why cutting planes and surfaces?

- Interior point algorithms solve SDPs in polynomial time.
- But, computational time becomes impractical for larger problems.

- Alternatively, use relaxations or approximations to get a good solution reasonably quickly.
- These relaxation approaches will also typically give an indication of a final duality gap.

- Other alternative approaches include:
  Spectral bundle method of Helmberg, Rendl, and Kiwiel.
  Augmented Lagrangian approaches of Burer, Monteiro, Zhang.
  First order methods based on Nesterov's smoothing scheme.
  Methods exploiting sparsity and symmetry by Fukuda et al and by De Klerk et al.

Why cutting planes and surfaces?

- Interior point algorithms solve SDPs in polynomial time.
- But, computational time becomes impractical for larger problems.

- Alternatively, use relaxations or approximations to get a good solution reasonably quickly.
- These relaxation approaches will also typically give an indication of a final duality gap.

- Other alternative approaches include:
  Spectral bundle method of Helmberg, Rendl, and Kiwiel.
  Augmented Lagrangian approaches of Burer, Monteiro, Zhang.
  First order methods based on Nesterov's smoothing scheme.
  Methods exploiting sparsity and symmetry by Fukuda et al and by De Klerk et al.

A relaxation of SDP

- Original primal-dual pair:

$$\begin{array}{llcl} \min & C \bullet X & & \\ \text{s.t.} & \mathcal{A}(X) & = & b \\ & X & \succeq & 0 \end{array} \qquad \begin{array}{llcll} \max & b^T y & & & \\ \text{s.t.} & \mathcal{A}^*(y) & + & S & = C \\ & & & S & \succeq 0 \end{array}$$

- Restricted primal, relaxed dual: For a given matrix P:

$$\begin{array}{llcl} \min_V & C \bullet PVP^T & & \\ \text{s.t.} & \mathcal{A}(PVP^T) & = & b \\ & V & \succeq & 0 \end{array} \qquad \begin{array}{llcll} \max_{y,S} & b^T y & & & \\ \text{s.t.} & \mathcal{A}^*(y) & + & S & = C \\ & & & P^T SP & \succeq 0 \end{array}$$

with possibly additional restrictions on V and corresponding weakening of the condition $P^T SP \succeq 0$.

The matrix P is updated over the course of the algorithm.

A relaxation of SDP

- Original primal-dual pair:

$$\begin{array}{llcl} \min & C \bullet X & & \\ \text{s.t.} & \mathcal{A}(X) & = & b \\ & X & \succeq & 0 \end{array} \qquad \begin{array}{llcccl} \max & b^T y & & & & \\ \text{s.t.} & \mathcal{A}^*(y) & + & S & = & C \\ & & & S & \succeq & 0 \end{array}$$

- Restricted primal, relaxed dual: For a given matrix P:

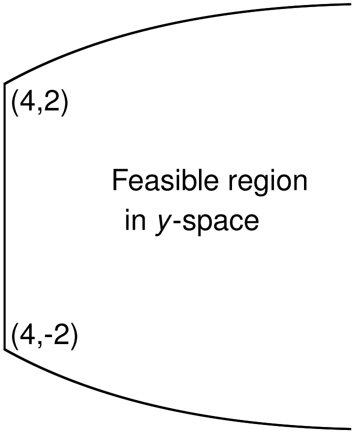
$$\begin{array}{llcl} \min_V & C \bullet PVP^T & & \\ \text{s.t.} & \mathcal{A}(PVP^T) & = & b \\ & V & \succeq & 0 \end{array} \qquad \begin{array}{llcccl} \max_{y,S} & b^T y & & & & \\ \text{s.t.} & \mathcal{A}^*(y) & + & S & = & C \\ & & & P^T SP & \succeq & 0 \end{array}$$

with possibly additional restrictions on V and corresponding weakening of the condition $P^T SP \succeq 0$.

The matrix P is updated over the course of the algorithm.

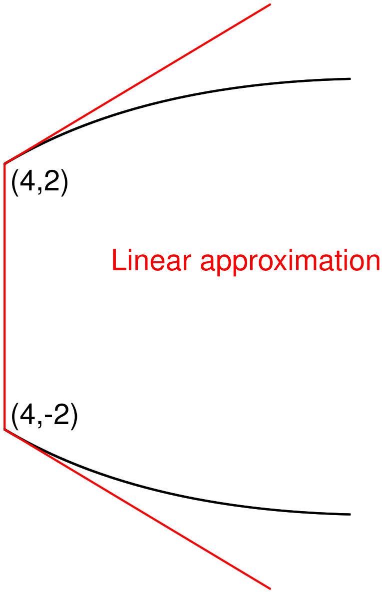
An example of an SDP feasible region

$$S = \begin{bmatrix} y_1 & y_2 & 0 \\ y_2 & y_1 - 3 & 0 \\ 0 & 0 & y_1 - 4 \end{bmatrix} \succeq 0$$



Tangent line approximation

$$S = \begin{bmatrix} y_1 & y_2 & 0 \\ y_2 & y_1 - 3 & 0 \\ 0 & 0 & y_1 - 4 \end{bmatrix} \succeq 0$$



Algorithmic framework

1. Approximately solve the modified problem, get $X = PVP^T$, S.
2. If S is not positive semidefinite:

   Add one or more columns to P corresponding to eigenvectors of S with negative eigenvalue.

   If desired, delete columns of P, or aggregate them into a linear term.

   Return to Step 1.

3. If duality gap too large, tighten the tolerance and return to Step 1.
4. STOP with an approximate solution to SDP.

Algorithmic framework

1. Approximately solve the modified problem, get $X = PVP^T$, S.
2. If S is not positive semidefinite:

   Add one or more columns to P corresponding to eigenvectors of S with negative eigenvalue.
   If desired, delete columns of P, or aggregate them into a linear term.
   Return to Step 1.
3. If duality gap too large, tighten the tolerance and return to Step 1.
4. STOP with an approximate solution to SDP.

Algorithmic framework

1. Approximately solve the modified problem, get $X = PVP^T$, S.
2. If S is not positive semidefinite:

   Add one or more columns to P corresponding to eigenvectors of S with negative eigenvalue.
   If desired, delete columns of P, or aggregate them into a linear term.
   Return to Step 1.

3. If duality gap too large, tighten the tolerance and return to Step 1.
4. STOP with an approximate solution to SDP.

Variant 1: Require V to be diagonal

$$V = \begin{bmatrix} * & & & \\ & * & & \\ & & \ddots & \\ & & & * \end{bmatrix}$$

$$\begin{array}{lrcl} \min & C \bullet PVP^T & & \\ \text{s.t.} & \mathcal{A}(PVP^T) & = & b \\ & V & \succeq & 0, \text{ diagonal} \end{array}$$

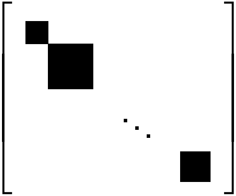
$$\begin{array}{lrcl} \max & b^T y & & \\ \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ \text{diagonal entries of } & P^T SP & \geq & 0 \end{array}$$

Gives a linear programming relaxation of the dual SDP.

Each column p of P gives a constraint $p^T Sp \geq 0$, a linear constraint on the entries in S.

(Sivaramakrishnan and M, and others.)

Variant 2: Require V to be block-diagonal

$V =$ 

$$\begin{array}{llll} \min & C \bullet PVP^T & & \\ \text{s.t.} & \mathcal{A}(PVP^T) & = & b \\ & V & \succeq & 0, \text{ block diagonal} \end{array}$$

$$\begin{array}{llllll} \max & b^T y & & & & \\ \text{s.t.} & \mathcal{A}^*(y) & + & S & = & C \\ \multicolumn{4}{l}{\text{diagonal blocks of } P^T SP} & \succeq & 0 \end{array}$$

Approximate the original SDP constraint by several lower-dimensional SDP constraints.

For example, diagonal blocks of size two correspond to second order cone constraints. $M = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0 \Leftrightarrow a \geq 0, c \geq 0, ac \geq b^2$.

(Oskoorouchi et al.)

Variant 3: No additional restrictions on V

$$\begin{array}{llcl} \min & C \bullet PVP^T & & \\ \text{s.t.} & \mathcal{A}(PVP^T) & = & b \\ & V & \succeq & 0 \end{array} \qquad \begin{array}{llcl} \max & b^T y & & \\ \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ & P^T S P & \succeq & 0 \end{array}$$

$$X = \boxed{P}\,\boxed{V}\,\boxed{P^T}$$

Approximate the original SDP constraint by a single lower-dimensional SDP constraint.

We refer to this as an active set algorithm for SDP.
The columns of P constitute the active set.
This primal-dual pair is our active set subproblem.

Variant 3: No additional restrictions on V

$$\begin{array}{lrcl} \min & C \bullet PVP^T & & \\ \text{s.t.} & \mathcal{A}(PVP^T) & = & b \\ & V & \succeq & 0 \end{array} \qquad \begin{array}{lrcl} \max & b^T y & & \\ \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ & P^T S P & \succeq & 0 \end{array}$$

$$X = \boxed{P}\,\boxed{V}\,\boxed{P^T}$$

Approximate the original SDP constraint by a single lower-dimensional SDP constraint.

We refer to this as an active set algorithm for SDP.
The columns of P constitute the active set.
This primal-dual pair is our active set subproblem.

Outline

A simplex-style approach

Sivaramakrishnan, Pataki, Zhang:

Solve the primal active set SDP subproblem with the current active set P, get V as low rank as possible.

Find a basis Q for the columns of PVP^T.

Find S to satisfy complementary slackness in the constrained SDP, so $Q^TSQ = 0$.

If S has a negative eigenvector, modify P and repeat.

Working with interior solutions

For the theoretical analysis, work with the **dual feasibility problem:**

Find y so that $C - \mathcal{A}^*(y) \succeq 0$.

$$\begin{array}{llcl} \min_{V,\alpha} & C \bullet (\alpha W + PVP^T) & & \\ \text{s.t.} & \mathcal{A}(\alpha W + PVP^T) & = & 0 \\ & V & \succeq & 0 \\ & \alpha & \geq & 0 \end{array} \qquad \begin{array}{llcl} \max_{y,S} & 0 & & \\ \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ & P^T SP & \succeq & 0 \\ & W \bullet S & \geq & 0 \end{array}$$

Include an extra aggregation matrix W in the problem.
As columns of P are dropped, they are incorporated into W.

Look for analytic center: $P^T SPV = I$, $\alpha W \bullet S = 1$.

How to restart effectively after modifying P?

Working with interior solutions

For the theoretical analysis, work with the **dual feasibility problem:**

Find y so that $C - \mathcal{A}^*(y) \succeq 0$.

$$\begin{array}{llcl} \min_{V,\alpha} & C \bullet (\alpha W + PVP^T) & & \\ \text{s.t.} & \mathcal{A}(\alpha W + PVP^T) & = & 0 \\ & V & \succeq & 0 \\ & \alpha & \geq & 0 \end{array} \qquad \begin{array}{llcl} \max_{y,S} & 0 & & \\ \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ & P^T SP & \succeq & 0 \\ & W \bullet S & \geq & 0 \end{array}$$

Include an extra aggregation matrix W in the problem.
As columns of P are dropped, they are incorporated into W.

Look for analytic center: $P^T SPV = I$, $\alpha W \bullet S = 1$.

How to restart effectively after modifying P?

Working with interior solutions

For the theoretical analysis, work with the **dual feasibility problem:**

Find y so that $C - \mathcal{A}^*(y) \succeq 0$.

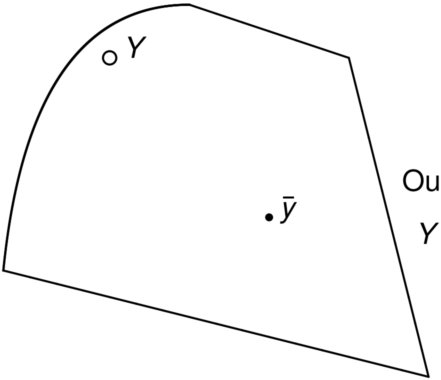
$$
\begin{array}{llcl}
\min_{V,\alpha} & C \bullet (\alpha W + PVP^T) & & \\
\text{s.t.} & \mathcal{A}(\alpha W + PVP^T) & = & 0 \\
 & V & \succeq & 0 \\
 & \alpha & \geq & 0
\end{array}
\qquad
\begin{array}{llcl}
\max_{y,S} & 0 & & \\
\text{s.t.} & \mathcal{A}^*(y) + S & = & C \\
 & P^T SP & \succeq & 0 \\
 & W \bullet S & \geq & 0
\end{array}
$$

Include an extra aggregation matrix W in the problem.
As columns of P are dropped, they are incorporated into W.

Look for analytic center: $P^T SPV = I, \quad \alpha W \bullet S = 1.$

How to restart effectively after modifying P?

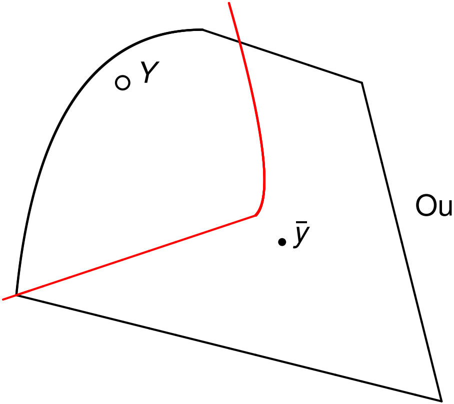
Find approximate analytic center $\bar{y}$



Outer approximation of

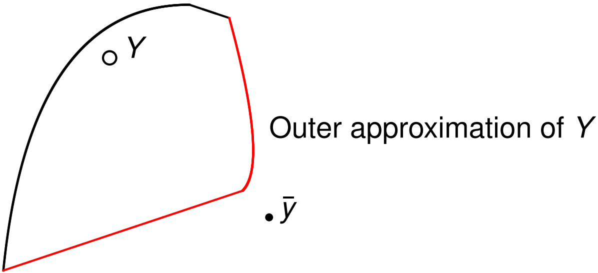
$$Y := \{y : C - \mathcal{A}^*(y) \succeq 0\}$$

Update P using eigenvector(s) of $\bar{S}$ at $\bar{y}$

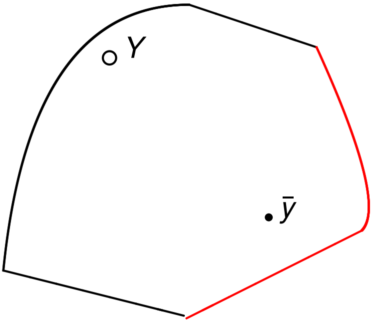


Outer approximation of Y

Update outer approximation



Weaken constraint to make it easier to restart



Updated outer
approximation of Y

Adding a column to P

Have current approximate analytic center $(\bar{V}, \bar{\alpha}, \bar{y}, \bar{S})$ to

$$\begin{array}{llcl} \min_{V,\alpha} & C \bullet (\alpha W + PVP^T) & & \\ \text{s.t.} & \mathcal{A}(\alpha W + PVP^T) & = & 0 \\ & V & \succeq & 0 \\ & \alpha & \geq & 0 \end{array} \qquad \begin{array}{llcl} \max_{y,S} & 0 & & \\ \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ & P^T S P & \succeq & 0 \\ & W \bullet S & \geq & 0 \end{array}$$

$\bar{p}$ is an eigenvector of S with negative eigenvalue.

$P^T P = I$ by assumption.

Normalize: Let $p = \gamma(\bar{p} - PP^T\bar{p})$ with $\gamma = \frac{1}{\sqrt{\bar{p} - PP^T\bar{p}}}$.

Let $\hat{P} = [P \;\; p]$.

Want strictly feasible solutions after P is replaced by $\hat{P}$.

Adding a column to P

Have current approximate analytic center $(\bar{V}, \bar{\alpha}, \bar{y}, \bar{S})$ to

$$\begin{array}{llcl} \min_{V,\alpha} & C \bullet (\alpha W + PVP^T) & & \\ \text{s.t.} & \mathcal{A}(\alpha W + PVP^T) & = & 0 \\ & V & \succeq & 0 \\ & \alpha & \geq & 0 \end{array} \qquad \begin{array}{llcl} \max_{y,S} & 0 & & \\ \text{s.t.} & \mathcal{A}^*(y) + S & = & C \\ & P^T SP & \succeq & 0 \\ & W \bullet S & \geq & 0 \end{array}$$

$\bar{p}$ is an eigenvector of S with negative eigenvalue.

$P^T P = I$ by assumption.

Normalize: Let $p = \gamma(\bar{p} - PP^T\bar{p})$ with $\gamma = \frac{1}{\sqrt{\bar{p} - PP^T\bar{p}}}$.

Let $\hat{P} = [P \;\; p]$.

Want strictly feasible solutions after P is replaced by $\hat{P}$.

Updating the primal solution

Let the square matrix L satisfy $LL^T = P^T \bar{S} P$. Eg, $L =$ Cholesky factor, or square root, or ...

Use a Dikin ellipsoid to choose a direction leading to a strictly feasible primal solution:

Require $||L^T(\Delta V)L||_F \leq 1$ (Frobenius norm).

Solve direction finding problem:

$$\begin{array}{lll} \max_{\beta,\Delta V} & \beta & \\ \text{s.t.} & \mathcal{A}\left([P \;\; p]\begin{bmatrix} \Delta V & 0 \\ 0 & \beta \end{bmatrix}\begin{bmatrix} P^T \\ p^T \end{bmatrix}\right) & = \;\; 0 \\ & ||L^T(\Delta V)L||_F & \leq \;\; 1 \end{array}$$

(Keep α unchanged.)

Updating the primal solution

Let the square matrix L satisfy $LL^T = P^T \bar{S} P$. Eg, $L =$ Cholesky factor, or square root, or ...

Use a Dikin ellipsoid to choose a direction leading to a strictly feasible primal solution:

Require $||L^T(\Delta V)L||_F \leq 1$ (Frobenius norm).

Solve direction finding problem:

$$
\begin{array}{lccc}
\max_{\beta,\Delta V} & \beta & & \\
\text{s.t.} & \mathcal{A}([P \;\; p]\begin{bmatrix} \Delta V & 0 \\ 0 & \beta \end{bmatrix}\begin{bmatrix} P^T \\ p^T \end{bmatrix}) & = & 0 \\
 & ||L^T(\Delta V)L||_F & \leq & 1
\end{array}
$$

(Keep α unchanged.)

Shift the dual constraint

1. Replace C by $C + \epsilon pp^T$ for some appropriate $\epsilon > 0$.
2. Choose ϵ so that it is easy to find a new approximate analytic center.
3. Gradually reduce ϵ to zero while maintaining an approximate analytic center.

Shift the dual constraint

1. Replace C by $C + \epsilon pp^T$ for some appropriate $\epsilon > 0$.
2. Choose ϵ so that it is easy to find a new approximate analytic center.
3. Gradually reduce ϵ to zero while maintaining an approximate analytic center.

Shift the dual constraint

1. Replace C by $C + \epsilon pp^T$ for some appropriate $\epsilon > 0$.
2. Choose ϵ so that it is easy to find a new approximate analytic center.
3. Gradually reduce ϵ to zero while maintaining an approximate analytic center.

Choosing ϵ (page 1)

With the current $\bar{y}$, get slack $\hat{S} = \bar{S} + \epsilon pp^T$, so we have

$$\begin{aligned} \hat{P}^T \hat{S} \hat{P} &= \begin{bmatrix} P^T \\ p^T \end{bmatrix} (\bar{S} + \epsilon pp^T)[P, p] = \begin{bmatrix} P^T SP & P^T Sp \\ p^T SP & p^T Sp + \epsilon \end{bmatrix} \\ &= \begin{bmatrix} L & 0 \\ a^T & \nu \end{bmatrix} \begin{bmatrix} L^T & a \\ 0 & \nu \end{bmatrix} \end{aligned}$$

where a satisfies

$$La = P^T Sp$$

and

$$\nu^2 = p^T Sp + \epsilon - a^T a.$$

Choosing ϵ (page 2)

Centrality condition for $\hat{P}^T \hat{S} \hat{P}$ and the modified V:

$$I \approx \begin{bmatrix} L^T & a \\ 0 & \nu \end{bmatrix} \begin{bmatrix} V + \Delta V & 0 \\ 0 & \beta \end{bmatrix} \begin{bmatrix} L & 0 \\ a^T & \nu \end{bmatrix}$$
$$= \begin{bmatrix} L^T(V + \Delta V)L & 0 \\ 0 & 0 \end{bmatrix} + \beta \begin{bmatrix} a \\ \nu \end{bmatrix} \begin{bmatrix} a^T & \nu \end{bmatrix}.$$

Scale primal step so that $\beta = \frac{1}{\nu^2}$: change in the centrality violation is no greater than

$$||L^T \Delta VL||_F + 1 + \beta a^T a.$$

This quantity can be made arbitrarily close to 1 by picking ϵ large and hence β and $||\Delta V||_F$ small.

Reducing ϵ

Reduce ϵ so that still in a larger neighborhood of the central path and can return to the central path quickly.

Can possibly exploit work on sensitivity analysis for SDPs (Yildirim et al).

Dropping columns of P

Can regain a new approximate analytic center immediately, without changing (y, S).

Diagonalize PVP^T as $U\Sigma U^T$, where U is of the same dimension as P.

If $P^TSPV = I$ then $U^TSU\Sigma = I$.

Drop the column u_0 of U with smallest eigenvalue σ_0.

Update W to $W + \frac{\sigma_0}{\alpha} u_0 u_0^T$.

Then can keep same α and just truncate Σ.

Outline

Conclusions

- Can approximate an SDP using a smaller SDP.
- Can shift the constraint in order to enable a quick restart.
- Can shrink the size of the SDP by aggregating columns.
- If don't shrink the active set then eventually get the original SDP so get convergence.