Properties of a Cutting Plane Method for Semidefinite Programming

John E. Mitchell
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY 12180 USA
mitchj@rpi.edu
http://www.rpi.edu/~mitchj

Joint work with **Kartik Krishnan**.

ISMP 2003
Copenhagen
19 August 2003

Abstract

A semidefinite programming problem is a nonsmooth optimization problem, so it can be solved using a cutting plane approach. In this talk, we analyze properties of such an algorithm. We discuss characteristics of good polyhedral representations the semidefinite program. We show that the complexity of an interior point cutting plane approach based on a semi-infinite formulation of the semidefinite program has complexity comparable with that of a direct interior point solver. We show that cutting planes can always be found efficiently that support the feasible region. Further, we characterize the cutting planes that give high dimensional tangent planes, and show how such cutting planes can be found efficiently.

Contents

1 Semidefinite programming

Standard form:

$$\begin{array}{lrcl} \min & C \bullet X & & \\ \text{subject to} & \mathcal{A}(X) & = & b \qquad (SDP) \\ & X & \succeq & 0, \end{array}$$

with dual

$$\begin{array}{lrcl} \max & b^T y & & \\ \text{subject to} & \mathcal{A}^T y + S & = & C \qquad (SDD) \\ & S & \succeq & 0 \end{array}$$

- C, X and S are square symmetric matrices.
- Require X and S be **positive semidefinite** (psd).
- Inner products are the **Frobenius inner product**, so $C \bullet X = \text{trace}(CX)$.
- $\mathcal{A}X = \begin{bmatrix} A_1 \bullet X \\ \vdots \\ A_m \bullet X \end{bmatrix}$, $\mathcal{A}^T y = \Sigma_{i=1}^m y_j A_j$, where each A_j is a symmetric matrix.
- Denote

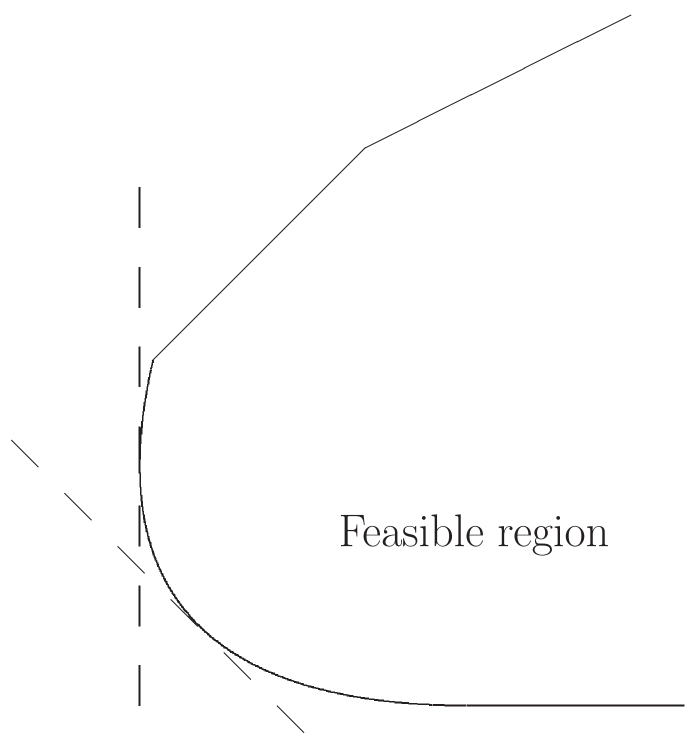
$$Y := \{y \in I\!R^m : C - \mathcal{A}^T y \succeq 0\}.$$

Methods for solving SDPs

- **Primal-dual interior point approaches** are limited in the size of problems they can solve.
- Alternatives include:
  - The **bundle method** of Helmberg, Rendl, and Kiwiel.
  - The **nonlinear programming approach** of Burer, Monteiro, and Zhang.

2 A semi-infinite linear program

- (SDP) and (SDD) are **convex programming** problems.
- Only nonlinearity is the positive semidefiniteness (PSD) requirement.
- **Replace PSD constraint by linear constraints.**



Dual formulation:

$$\begin{array}{lrcll} \max & b^T y & & & (LDD) \\ \text{subject to} & d^T(C - \mathcal{A}^T y)d & \geq & 0 \;\; \forall \text{d with } ||d||^2 = 1. & \end{array}$$

- If (SDP) has m constraints, then (LDD) has m variables.
- Need to be **selective** with the vectors d included in the constraints. For example, use a **cutting plane approach** to select vectors d.
- Typically, we will need some vectors d that are **dense**, leading to a dense linear programming problem.
- Have far fewer variables than in a primal formulation, so get smaller linear programs (although they are still dense).

3 Cutting plane algorithm for (SDP)

- Choose an **initial set** of constraints for (LDR).
- **Solve** (LDR) and (LPR) approximately using an interior point method.
- Get **trial point** $\bar{y}$.
- If $C - \mathcal{A}^T \bar{y}$ is not psd, find vectors d and corresponding **violated constraints** $d^T(C - \mathcal{A}^T y)d \geq 0$.
- Modify (LDR) and (LPR), and **repeat**.

Finding violated constraints

- We have a vector $\bar{y}$ from (LDR).
- Calculate $\bar{S} = C - \mathcal{A}^T\bar{y}$.
- Solve the **quadratic subproblem**

$$\begin{array}{lrcl} \min_d & d^T\bar{S}d & & \\ \text{subject to} & \|d\| & \leq & 1 \end{array}$$

- Subproblem only needs to be solved approximately: just need to find a solution with negative objective function value to cut off the current solution $\bar{y}$.
- Can use different norms:
  - **2-norm**: Optimal solution is the eigenvector with most negative eigenvalue. Can use eigenvalue schemes such as Lanczos to find eigenbases for several negative eigenvalues simultaneously.
  - **∞-norm**: Solutions d may contain a number of ± 1 components. Useful if we are solving a combinatorial optimization problem such as Maxcut.

4 Dimension of added face

Assume:

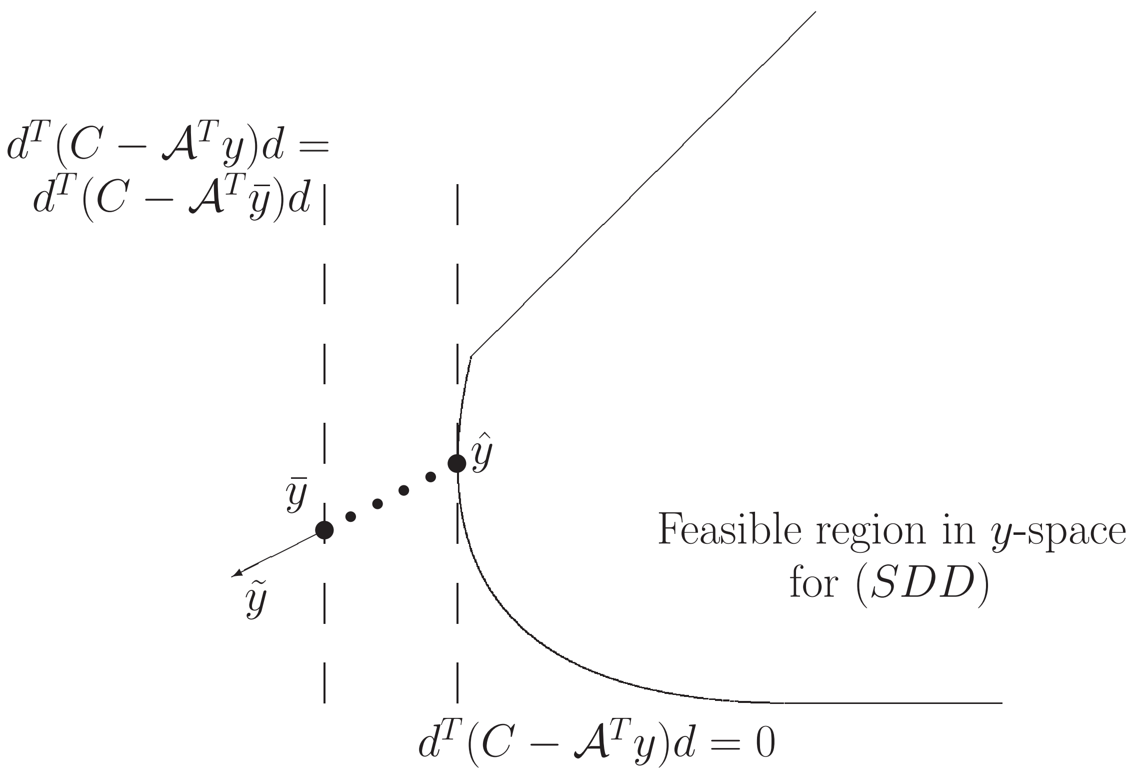
1. The constraints $\mathcal{A}(X) = b$ imply $\textbf{trace}(X) = a$ for some fixed a.
2. Given $\bar{y}$ and $\bar{S} = C - \mathcal{A}^T\bar{y}$, the vector d is an eigenvector of $\bar{S}$ corresponding to the **most negative eigenvalue** $\lambda_{\min}$ of $\bar{S}$. Let r be the multiplicity of this eigenvalue.

Then:

1. **There exists $\hat{y}$ feasible in (SDD) satisfying the constraint $d^T(C - \mathcal{A}^T y)d \geq 0$ at equality.**
2. **If $r = 1$ then the hyperplane $d^T(C - \mathcal{A}^T y)d = 0$ induces a tangent space of dimension $m - 1$.**
3. **For general r, a vector d can be found that defines a hyperplane $d^T(C - \mathcal{A}^T y)d = 0$ which induces a tangent space of dimension at least $m - r$, under certain weak assumptions.**

Constraint satisfied at equality:

- Since $\text{trace}(X) = a$, there exists $\tilde{y}$ with $\mathcal{A}^T\tilde{y} = I$.
- Let $\hat{y} := \bar{y} + \lambda_{\min}\tilde{y}$.
- Then $\hat{S} := C - \mathcal{A}^T\hat{y}$ is psd, with nullity r, and $d^T\hat{S}d = 0$.

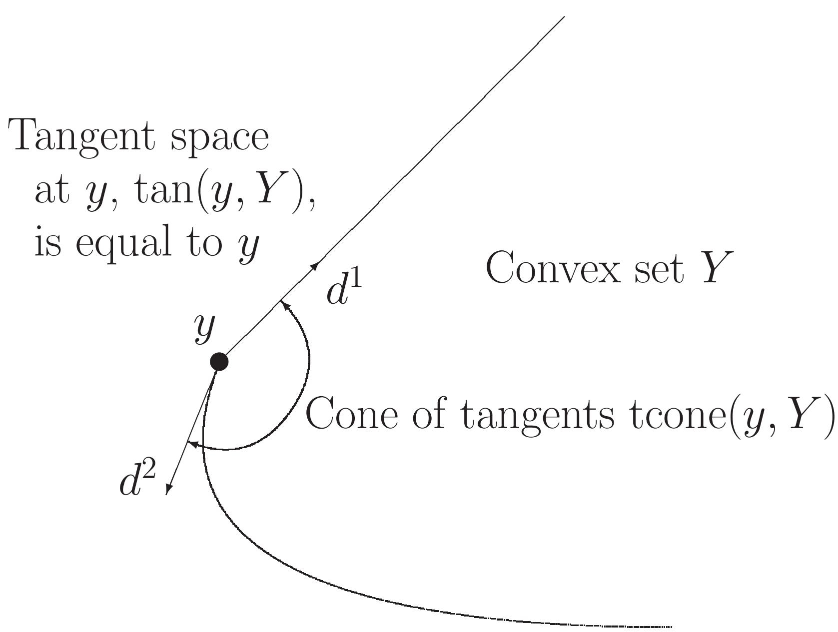


Tangent space and cone of tangents

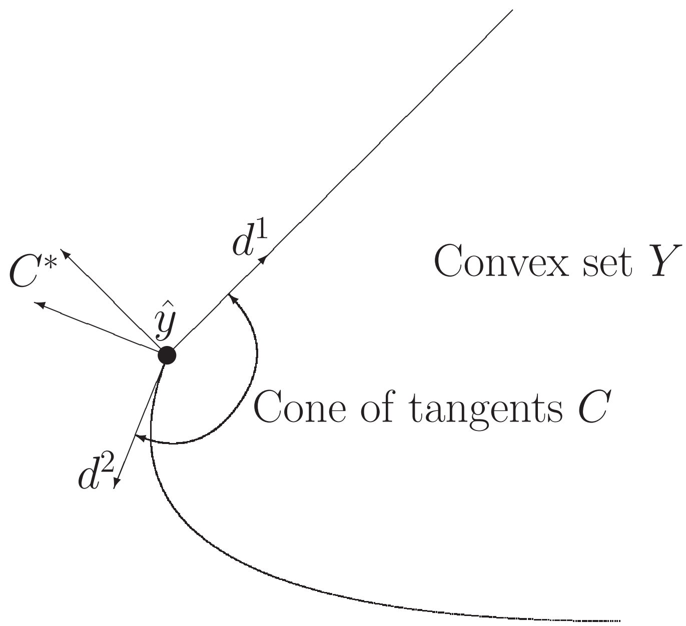
Take $Y \subseteq I\!R^2$ to be the vectors y that make the following matrix S positive semidefinite:

$$S = \begin{bmatrix} y_1 & y_2 & 0 \\ y_2 & y_1 - 3 & 0 \\ 0 & 0 & 2 - y_2 \end{bmatrix}.$$

Let $y = [4 \ \ 2]^T \in Y$. The corresponding matrix S has rank one and nullity two.



Dual cone



$$d_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad d_2 = \begin{bmatrix} -4 \\ -5 \end{bmatrix}$$

Extreme rays of dual cone C^* give the constraints

$$\begin{aligned} y_2 &\leq 2 \\ 5y_1 - 4y_2 &\geq 12 \end{aligned}$$

Using an eigendecomposition

- Define **eigendecomposition** of $\hat{S}$:

$$\hat{S} = [P \;\; Q] \begin{bmatrix} 0 & 0 \\ 0 & \Gamma \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$$

- Note: we do not assume strict complementarity.
- (Alizadeh et al.)

  Tangent space of cone of psd matrices at $\hat{S}$:

  $$\{[P \;\; Q] \begin{bmatrix} 0 & V \\ V^T & U \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix} : U \in \mathcal{S}^{m-r}, V \in I\!R^{r \times (m-r)}\}$$

  After a step of length $O(\epsilon)$ in the tangent space, still within $O(\epsilon^2)$ of the cone of psd matrices.

What about Y?

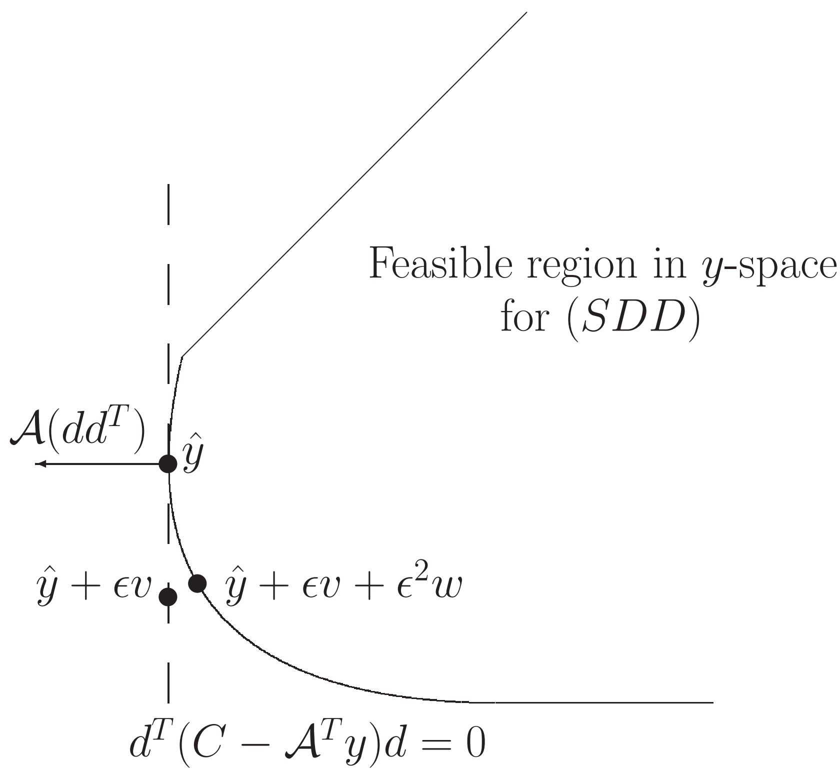
- Not every psd matrix S corresponds to a choice of y.
- At a point $\hat{y} + v$, the slack matrix is

$$\begin{aligned} S &= C - \mathcal{A}^T \hat{y} - \mathcal{A}^T v \\ &= [P\ Q] \left(\begin{bmatrix} 0 & 0 \\ 0 & \Gamma \end{bmatrix} - \begin{bmatrix} P^T \\ Q^T \end{bmatrix} \mathcal{A}^T v [P\ Q] \right) \begin{bmatrix} P^T \\ Q^T \end{bmatrix} \end{aligned}$$

- Thus, v is in the tangent space if $P^T \mathcal{A}^T v P = 0$. This is a system of linear equations in v.
- Take constraint defined by a vector in the nullspace of $\hat{S}$, so $d = Pu$ for some u. So hyperplane is $d^T \mathcal{A}^T (\hat{y} - y) = 0$.

Nullity of one

If $\hat{S}$ has nullity $m - r = 1$, then P consists of a single column. So **the dimension of the tangent space is** $r = m - 1$. Further, everything in the tangent space also satisfies $d^T \mathcal{A}^T v = 0$.



Larger values for the nullity

- Generalize the idea of the tangent space:

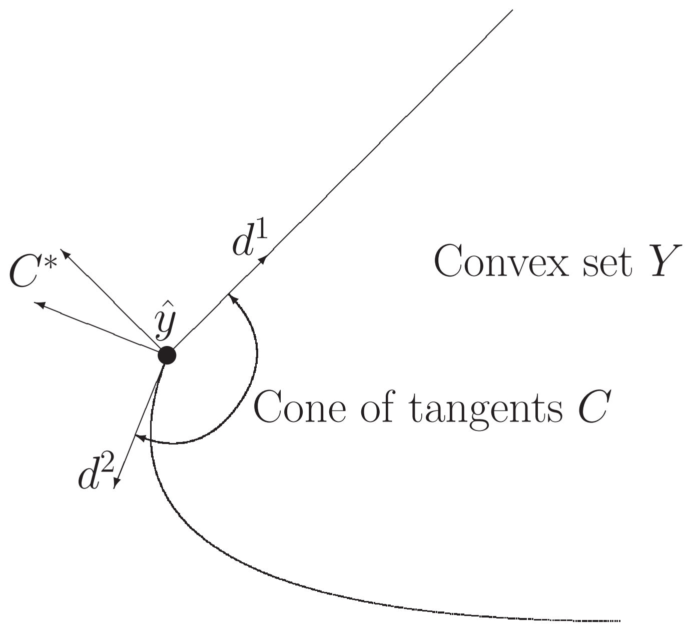
  > Given a supporting hyperplane H for a convex set Y, and given $y \in Y \cap H$, we call the intersection of H with $\mathrm{tcone}(y, Y)$ the **tangent space induced by the hyperplane at the point**.

- The **cone of tangents** C of Y at $\hat{y}$ is the set of vectors d for which the matrix $P^T(\mathcal{A}^T d)P$ is **negative semidefinite**.

- Let C^* be the dual cone to the cone of tangents, taken in the sense $w \in C^*$ if and only if $w^T d \leq 0$ for all $d \in C$.

- Any $w \in C^*$ gives a supporting hyperplane of Y at $\hat{y}$.

Dual cone



$$d_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad d_2 = \begin{bmatrix} -4 \\ -5 \end{bmatrix}$$

Extreme rays of dual cone C^* give the constraints

$$\begin{aligned} y_2 &\leq 2 \\ 5y_1 - 4y_2 &\geq 12 \end{aligned}$$

The form of the of the dual cone

- In general, the set C^* is equal to the closure of the set of vectors of the form

$$\tilde{C}^* := \{\mathcal{A}(PZP^T) : Z \succeq 0\}.$$

- The set $\tilde{C}^*$ may not be closed.

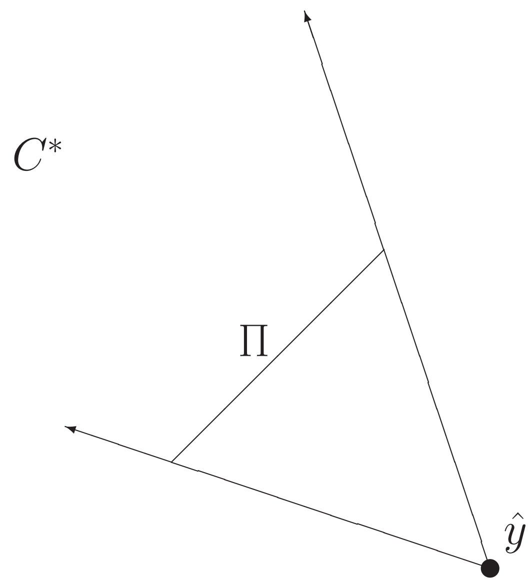
  Example (Pataki):

$$P^T A_1 P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad P^T A_2 P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

  The ray $(0,1)^T$ is not in the cone $\{\mathcal{A}(PZP^T) : Z \succeq 0\}$, but it is in its closure.

- Under our first assumption, the set $\tilde{C}^*$ is closed.

Finding the extreme rays of the dual cone



$$\Pi = \{\mathcal{A}(PZP^T) : \text{trace}(Z) = 1, Z \succeq 0\}$$

- The extreme rays of C^* correspond to the extreme points of the compact set Π.
- Let u be an eigenvector of minimum eigenvalue of the matrix $(P^T \mathcal{A} P)^T g$ for some vector g. The **extreme rays** of C^* are vectors of the form $w = \mathcal{A}(Puu^T P^T)$.
- Such a w gives a valid constraint $w^T y \leq w^T \hat{y}$.
- If the minimum eigenvalue of the matrix $(P^T \mathcal{A} P)^T g$ has multiplicity equal to one then **the tangent space induced by w has dimension at least** $m - r$.

Example

- $C = ee^T + I$
- $A_i = e_i e_i^T, i = 1, \ldots, n$. Thus, $m = n$.
- Assume $n \geq 4$, n is even.
- $S = ee^T + I - \text{diag}(y)$.
- $\hat{y} = e$.
- $\hat{S} = ee^T$, so nullity is $r = n - 1$ and $m - r = 1$.
- Take $d = e_1 - e_2$, in the nullspace of $\hat{S}$.

  Get valid constraint: $y_1 + y_2 \leq 2$.

  Gives **tangent plane of dimension** $n - 2 > m - r$.
- Take $d_i = \begin{cases} 1 & \text{for } i = 1, \ldots, n/2 \\ -1 & \text{for } i = 1 + n/2, \ldots, n \end{cases}$

  Get valid constraint: $e^T y \leq n$.

  Tangent plane has dimension $0 < m - r$**.**
- There is **no tangent plane with dimension** $m - 1$**.**

Regaining a primal solution

- Because the linear programs are smaller, it was more efficient to work with the **dual formulation**.
- We only use a finite number of vectors d, so we are solving a **linear programming relaxation** of the dual:

$$\begin{array}{lrcll} \max_y & b^T y & & & (LDR) \\ \text{subject to} & d_i^T(C - \mathcal{A}^T y)d_i & \geq & 0 & \\ & \text{for vectors } d_i, i = 1, \ldots, m. & & & \end{array}$$

- The optimal value to (LDR) gives an upper bound on the optimal value of (SDP).
- The corresponding constrained version of (SDP) given by **LP duality** can be written:

$$\begin{array}{lrcll} \min_x & C \bullet (\Sigma_{i=1}^m x_i d_i d_i^T) & & & \\ \text{subject to} & \mathcal{A}(\Sigma_{i=1}^m x_i d_i d_i^T) & = & b & (LPR) \\ & x & \geq & 0. & \end{array}$$

- Thus, any feasible solution to (LPR) gives a **feasible solution to (SDP)**, $X = \Sigma_{i=1}^m x_i d_i d_i^T$

A perfect set of constraints

Let X^* solve (SDP) and y^*, S^* solve (SDD).

- X^* has an **eigendecomposition**:

$$X^* = [P \;\; Q] \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$$

- X^* and S^* are **simultaneously diagonalizable**:

$$S^* = [P \;\; Q] \begin{bmatrix} 0 & 0 \\ 0 & \Gamma \end{bmatrix} \begin{bmatrix} P^T \\ Q^T \end{bmatrix}$$

- Let $p^1, \ldots, p^r$ be the columns of P. Then the optimal solutions to

$$\begin{array}{lrcl} \max & b^T y & & \\ \text{subject to} & {p^i}^T (C - \mathcal{A}^T y) p^i & \geq & 0, \;\; i = 1, \ldots, r \end{array}$$

  and its dual give the **optimal solutions** to (SDD) and (SDP): set $x = \lambda$, where $\lambda = \text{diag}(\Lambda)$.

- Pataki: expect $r \approx O(\sqrt{m})$.

5 Nonpolyhedral variants of (LPR)

- (LPR) can be rewritten as

$$\begin{array}{llcll} \min & C \bullet (DMD^T) & & & \\ \text{s.t.} & A_j \bullet (DMD^T) & = & b_j & j = 1, \ldots, m \\ & M & \succeq & 0 & \\ & M \ \textbf{diagonal} & & & \end{array} \tag{1}$$

  where the columns of the $n \times m$ matrix D are the vectors $d_i, i = 1, \ldots, m$.

- The requirement that M be diagonal can be relaxed or omitted.

- **Omitting** the requirement gives back (SDP) if D has full row rank n. This is the basis of the spectral bundle method when M is small.

- Requiring M be **block-diagonal** corresponds to a cutting plane method that adds semidefinite cuts. (Oskoorouchi and Goffin.)

6 Complexity of the algorithm

- Theoretically, could use a **volumetric barrier cutting plane algorithm** to solve (SDD).
- Want to get within ϵ of optimality.
- Requires $O(m \log(\frac{1}{\epsilon}))$ **calls to the quadratic subproblem** and a similar number of Newton steps.
- Each **solution of the subproblem** requires $O(n^3)$ work if the QR algorithm is used to find the solutions to the subproblems. It also requires $O(mn^2)$ work to calculate $\bar{S}$.
- Each **Newton step** in (LDR) requires $O(m^3)$ work.
- Thus, for $m \geq n$, **total number of arithmetic operations** is of the order of $O(m^4 \log(\frac{1}{\epsilon}))$.
- A **primal dual SDP method** requires $O(mn^3 + m^2n^2 + m^3)$ arithmetic operations at each iteration and $O(\sqrt{n} \log(\frac{1}{\epsilon}))$ iterations. So overall complexity is $O(m^2n^{2.5} \log(\frac{1}{\epsilon}))$ for $n \leq m \leq n^2$. (Complexity can be reduced slightly if constraint matrices have special structure.)

7 Conclusions

1. Cutting plane approaches to the solution of semidefinite programs have attractive theoretical complexity.
2. High-dimensional cutting planes can be found efficiently.
3. The cone of tangents can be captured exactly through the use of a nonpolyhedral cutting plane.

References

[1] F. Alizadeh, J.-P. Haeberly, and M. L. Overton. Complementarity and nondegeneracy in semidefinite programming. *Mathematical Programming*, 77(2):111–128, 1997.

[2] S. Burer, R. D. C. Monteiro, and Y. Zhang. Solving a class of semidefinite programs via nonlinear programming. *Mathematical Programming*, 93(1):97–122, 2002.

[3] C. Helmberg. Semidefinite programming for combinatorial optimization. Technical Report ZR-00-34, TU Berlin, Konrad-Zuse-Zentrum, Berlin, October 2000. Habilitationsschrift.

[4] C. Helmberg and K. Kiwiel. A spectral bundle method with bounds. *Mathematical Programming*, 93(2):173–194, 2002.

[5] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10(3):673–696, 2000.

[6] K. Krishnan and J. E. Mitchell. Cutting plane methods for semidefinite programming. Technical report, Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, November 2002.

[7] K. Krishnan and J. E. Mitchell. Properties of a cutting plane method for semidefinite programming. Technical report, Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, May 2003.

[8] K. Krishnan and J. E. Mitchell. Semi-infinite linear programming approaches to semidefinite programming (SDP) problems. In P. Pardalos and H. Wolkowicz, editors, *Novel approaches to hard discrete optimization problems*, volume 37 of *Fields Institute Communications Series*, pages 123–142. AMS, 2003.

[9] M. R. Oskoorouchi and J.-L. Goffin. The analytic center cutting plane method with semidefinite cuts. *SIAM Journal on Optimization*, 13(4):1029–1053, 2003.

[10] G. Pataki. On the closedness of the linear image of a closed convex cone. Part I: necessary and sufficient conditions for nice cones, and core

separation. Technical report, Department of Operations Research, University of North Carolina, Chapel Hill, 2003.