Projection methods to solve SDP

Franz Rendl

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

Alpen-Adria-Universität Klagenfurt

Austria

Overview

- Augmented Primal-Dual Method
- Boundary Point Method

Semidefinite Programs

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

Some notation and assumptions:

X, Z symmetric $n \times n$ 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 $A^T(y) = \sum y_i A_i$.

We assume that both the primal and the dual problem have strictly feasible points ($X, Z \succ 0$), so that strong duality holds, and optima are attained.

Optimality conditions

Under these conditions, (X, y, Z) is optimal if and only if the following conditions hold:

$$A(X) = b,\ X \succeq 0, \quad \text{primal feasibility}$$

$$A^T(y) - Z = C,\ Z \succeq 0, \quad \text{dual feasibility}$$

$$\langle X, Z \rangle = 0 \ \text{complementarity.}$$

Last condition is equivalent to $\langle C, X \rangle = b^T y$.

It could also be replaced by the matrix equation

$$ZX = 0.$$

Other solution approaches

- Spectral Bundle method, see Helmberg, Rendl: SIOPT (2000): works on dual problem as eigenvalue optimization problem.
- Low-Rank factorization, see Burer, Monteiro: Math Prog (2003): express $X = LL^T$ and work with L. Leads to nonlinear optimization techniques.
- Iterative solvers for augmented system, see Toh: SIOPT (2004): use iterative methods to solve Newton system.
- Iterative solvers and modified barrier approach, see Kocvara, Stingl: Math Prog (2007): strong computational results using the package PENNSDP.
- and many other methods: sorry for not mentioning them all

Other solution approaches

- Spectral Bundle method
- Low-Rank factorization
- Iterative solvers for augmented system, Toh (2004)
- Iterative solvers and modified barrier approach, Kocvara, Stingl (2007)

Methods based on projection

- boundary point approach: (Povh, R., Wiegele: Computing 2006)
- regularization methods: Malick, Povh, R., Wiegele, 2009
- augmented primal-dual approach: (Jarre, R.: SIOPT 2009)

Comparing IP and projection methods

constraint	IP	BPM	APD
$A(X) = b$	yes	***	yes
$X \succeq 0$	yes	yes	***
$A^T(y) - C = Z$	yes	***	yes
$Z \succeq 0$	yes	yes	***
$\langle Z, X \rangle = 0$	—	—	yes
$ZX = 0$	***	yes	—

IP: Interior-point approach
BPM: boundary point method
APD: augmented primal-dual method
***: means that once this condition is satisfied, the method stops.

Augmented Primal-Dual Method

(This is joint work with Florian Jarre.)

$$FP := \{X : \ A(X) = b\} \text{ primal linear space,}$$

$$FD := \{(y, Z) : Z = C + A^T(y)\} \text{ dual linear space}$$

$$OPT := \{(X, y, Z); \langle C, X \rangle = b^T y\} \text{ optimality hyperplane.}$$

From Linear Algebra:

$$\Pi_{FP}(X) = X - A^T \left((AA^T)^{-1}(A(X) - b) \right),$$

$$\Pi_{FD}(Z) = C + A^T \left((AA^T)^{-1}(A(Z - C)) \right)$$

are the projections of (X, Z) onto FP and FD.

Augmented Primal-Dual Method (2)

Note that both projections essentially need one solve with matrix AA^T. (Needs to be factored only once.)
Projection onto OPT is trivial.
Let $K = FP \cap FD \cap OPT$. Given (X, y, Z), the projection $\Pi_K(X, y, Z)$ onto K requires two solves.

This suggests the following iteration:

Start: Select $(X, y, Z) \in K$
Iteration: while not optimal

- $X^+ = \Pi_{SDP}(X), \;\; Z^+ = \Pi_{SDP}(Z)$.
- $(X, y, Z) \leftarrow \Pi_K(X^+, y, Z^+)$

The projection $\Pi_{SDP}(X)$ of X onto SDP can be computed through an eigenvalue decomposition of X.

Augmented Primal-Dual Method (3)

This approach converges, but possibly very slowly.
The computational effort is two solves (order m) and two factorizations (order n).

An improvement: Consider

$$\phi(X,Z) := dist(X,SDP)^2 + dist(Z,SDP)^2.$$

Here $dist(X,SDP)$ denotes the distance of the matrix X from the cone of semidefinite matrices. The (convex) function ϕ is differentiable with Lipschitz-continuous gradient:

$$\nabla\phi(X,Z) = (X,Z) - \Pi_K(\Pi_{SDP}(X,Z))$$

We solve SDP by minimizing ϕ over K.

Augmented Primal-Dual Method (4)

Practical implementation currently under investigation.
The function ϕ could be modified by

$$\phi(X, Z) + \|XZ\|_F^2$$

Apply some sort of conjugate gradient approach (Polak-Ribiere) to minimize this function. Computational work:

- Projection onto K done by solving a system with matrix AA^T.
  - Evaluating ϕ involves spectral decomposition of X, Z.

This approach is feasible if n not too large ($n \leq 1000$), and if linear system with AA^T can be solved.

Augmented Primal-Dual Method (5)

Recall: (X, y, Z) is optimal once $X, Z \succeq 0$.
A typical run: $n = 400, \ m = 10000$.

iter	secs	$\langle C, X\rangle$	$\lambda_{\min}(X)$	$\lambda_{\min}(Z)$
1	9.7	11953.300	-0.00209	-0.00727
10	55.8	11942.955	-0.00036	-0.00055
20	103.8	11948.394	-0.00013	-0.00015
30	150.7	11950.799	-0.00007	-0.00005
40	196.7	11951.676	-0.00005	-0.00002
50	242.6	11951.781	-0.00004	-0.00001

The optimal value is 11951.726.

Random SDP

n	m	opt	apd	$\lambda_{\min}$
400	40000	-114933.8	-114931.1	-0.0002
500	50000	-47361.2	-47353.4	-0.0003
600	60000	489181.8	489194.5	-0.0004
700	70000	-364458.8	-364476.1	-0.0004
800	80000	-112872.6	-112817.4	-0.0011
1000	100000	191886.2	191954.5	-0.0012

50 iterations of APD.
Largest instance takes about 45 minutes.
$\lambda_{\min}$ is most negative eigenvalue of X and Z.

Boundary Point method

Augmented Lagrangian for (D)
$\min\{b^T y : A^T(y) - C = Z \succeq 0\}$.
$X \ldots$ Lagrange Multiplier for dual equations
$\sigma > 0$ penalty parameter

$$L_\sigma(y, Z, X) = b^T y + \langle X, Z + C - A^T(y)\rangle + \frac{\sigma}{2}\|Z + C - A^T(y)\|^2$$

Generic Method:
repeat until convergence
(a) Keep X fixed: solve $\min_{y, Z \succeq 0} \; L_\sigma(y, Z, X)$ to get $y, Z \succeq 0$
(b) update X: $X \leftarrow X + \sigma(Z + C - A^T(y))$
(c) update σ
Original version: Powell, Hestenes (1969)
σ carefully selected gives linear convergence

Inner Subproblem

Inner minimization:
X and σ are fixed.

$$W(y) := A^T(y) - C - \frac{1}{\sigma}X$$

$$L_\sigma = b^T y + \langle X, Z + C - A^T(y)\rangle + \frac{\sigma}{2}\|Z + C - A^T(y)\|^2 =$$

$$= b^T y + \frac{\sigma}{2}\|Z - W(y)\|^2 + const = f(y, Z) + const.$$

Note that dependence on Z looks like projection problem, but with additional variables y.
Altogether this is convex quadratic SDP!

Optimality conditions (1)

Introduce Lagrange multiplier $V \succeq 0$ for $Z \succeq 0$:

$$L(y, Z, V) = f(y, Z) - \langle V, Z \rangle$$

Recall:

$$f(y, Z) = b^T y + \frac{\sigma}{2}\|Z - W(y)\|^2, \quad W(y) = A^T(y) - C - \frac{1}{\sigma}X.$$

$$\nabla_y L = 0 \text{ gives } \sigma A A^T(y) = \sigma A(Z + C) + A(X) - b,$$

$$\nabla_Z L = 0 \text{ gives } V = \sigma(Z - W(y)),$$

$$V \succeq 0, \quad Z = \succeq 0, \quad VZ = 0.$$

Since Slater constraint qualification holds, these are necessary and sufficient for optimality.

Optimality conditions (2)

Note also: For y fixed we get Z by projection: $Z = W(y)_+$.
From matrix analysis:

$$W = W_+ + W_-, \quad W_+ \succeq 0, \quad -W_- \succeq 0, \quad \langle W_+, W_- \rangle = 0.$$

We have: (y, Z, V) is optimal if and only if:

$$AA^T(y) = \frac{1}{\sigma}(A(X) - b) + A(Z + C),$$

$$Z = W(y)_+, \quad V = \sigma(Z - W(y)) = -\sigma W(y)_-.$$

Solve linear system (of order m) to get y.
Compute eigenvalue decomposition of $W(y)$ (order n).
Note that AA^T does not change during iterations.

Boundary Point Method

Start: $\sigma > 0, X \succeq 0, Z \succeq 0$
repeat until $\|Z - A^T(y) + C\| \leq \epsilon$:

- repeat until $\|A(V) - b\| \leq \sigma\epsilon$ (X, σ fixed):
  - Solve for y: $AA^T(y) = rhs$
  - Compute $Z = W(y)_+, \quad V = -\sigma W(y)_-$
- Update $X : \quad X = -\sigma W(y)_-$

Inner stopping condition is primal feasibility.

Outer stopping condition is dual feasibility.

See: Povh, R, Wiegele (Computing, 2006)

Theta: big DIMACS graphs

graph	n	m	ϑ	ω
keller5	776	74.710	31.00	27
keller6	3361	1026.582	63.00	$\geq$59
san1000	1000	249.000	15.00	15
san400-07.3	400	23.940	22.00	22
brock400-1	400	20.077	39.70	27
brock800-1	800	112.095	42.22	23
p-hat500-1	500	93.181	13.07	9
p-hat1000-3	1000	127.754	84.80	$\geq$68
p-hat1500-3	1500	227.006	115.44	$\geq$94

see Malick, Povh, R., Wiegele (2008): The theta number for the bigger instances has not been computed before.

Random SDP

n	m	secs	iter	secs chol(AA')
300	5000	43	168	1
300	10000	158	229	56
400	10000	130	211	8
400	20000	868	204	593
500	10000	144	136	1
500	20000	431	205	140
600	10000	184	96	1
600	20000	345	155	23
600	30000	975	152	550
800	40000	1298	155	345

relative accuracy of 10^{-5}, coded in MATLAB.

Conclusions and References

- Both methods need more theoretical convergence analysis.
- Speed-up possible making use of limited-memory BFGS type methods.
- The spectral decomposition limits the matrix size n.
- Practical convergence may vary greatly depending on data.

3 papers:

Povh, R., Wiegele: Boundary point method (Computing 2006)
Malick, Povh, R., Wiegele: (SIOPT 2009)
Jarre, R.:, Augmented primal-dual method, (SIOPT 2008)

Large-Scale SDP

Projection methods like the boundary point method assume that a full spectral decomposition is computationally feasible.
This limits n to $n \leq 2000$ but m could be arbitrary.

What if n is much larger?

Spectral Bundle Method

What if m and n is large?
In addition to before, we now assume that working with symmetric matrices X of order n is too expensive (no Cholesky, no matrix multiplication!)
One possibility: Get rid of $Z \succeq 0$ by using eigenvalue arguments.

Constant trace SDP

A has constant trace property if I is in the range of A^T, equivalently

$$\exists \eta \text{ such that } A^T(\eta) = I$$

The constant trace property implies:

$$A(X) = b, \; A^T(\eta) = I \text{ then}$$

$$\text{tr}(X) = \langle I, X \rangle = \langle \eta, A(X) \rangle = \eta^T b =: a$$

Constant trace property holds for many combinatorially derived SDP!

Reformulating Constant Trace SDP

Reformulate dual as follows:

$$\min\{b^T y : A^T(y) - C = Z \succeq 0\}$$

Adding (redundant) primal constraint tr$(X) = a$ introduces new dual variable, say λ, and dual becomes:

$$\min\{b^T y + a\lambda : A^T(y) - C + \lambda I = Z \succeq 0\}$$

At optimality, Z is singular, hence $\lambda_{\min}(Z) = 0$.
Will be used to compute dual variable λ explicitely.

Dual SDP as eigenvalue optimization

Compute dual variable λ explicitely:

$$\lambda_{\max}(-Z) = \lambda_{\max}(C - A^T(y)) - \lambda = 0, \Rightarrow \lambda = \lambda_{\max}(C - A^T(y))$$

Dual equivalent to

$$\min\{a\ \lambda_{\max}(C - A^T(y)) + b^T y : y \in \Re^m\}$$

This is non-smooth unconstrained convex problem in y.
Minimizing $f(y) = \lambda_{\max}(C - A^T(y)) + b^T y$:
Note: Evaluating $f(y)$ at y amounts to computing largest eigenvalue of $C - A^T(y)$.
Can be done by iterative methods for very large (sparse) matrices.

Spectral Bundle Method (1)

If we have some y, how do we move to a better point?

$$\lambda_{\max}(X) = \max\{\langle X, W\rangle : \ \text{tr}(W) = 1, \ W \succeq 0\}$$

Define

$$L(W, y) := \langle C - A^T(y), W\rangle + b^T y.$$

Then $f(y) = \max\{L(W, y) : \ \text{tr}(W) = 1, \ W \succeq 0\}$.

Idea 1: Minorant for $f(y)$

Fix some $m \times k$ matrix P. $k \geq 1$ can be chosen arbitrarily.
The choice of P will be explained later.
Consider W of the form $W = PVP^T$ with new $k \times k$ matrix variable V.

$$\hat{f}(y) := \max\{L(W, y) : \ W = PVP^T, \ V \succeq 0\} \ \leq \ f(y)$$

Spectral Bundle Method (2)

Idea 2: Proximal point approach

The function $\hat{f}$ depends on P and will be a good approximation to $f(y)$ only in some neighbourhood of the current iterate $\hat{y}$.
Instead of minimizing $f(y)$ we minimize

$$\hat{f}(y) + \frac{u}{2}\|y - \hat{y}\|^2.$$

This is a strictly convex function, if $u > 0$ is fixed.
Substitution of definition of $\hat{y}$ gives the following min-max problem

Quadratic Subproblem (1)

$$\min_{y} \max_{W} L(W,y) + \frac{u}{2}\|y-\hat{y}\|^2 = \dots$$

$$= \max_{W,\ y=\hat{y}+\frac{1}{u}(A(W)-b)} L(W,y) + \frac{u}{2}\|y-\hat{y}\|^2$$

$$= \max_{W} \langle C - A^T(\hat{y}), W\rangle + b^T\hat{y} - \frac{1}{2u}\langle A(W)-b, A(W)-b\rangle.$$

Note that this is a quadratic SDP in the $k \times k$ matrix V, because $W = PVP^T$.

k is user defined and can be small, independent of n!!

Quadratic Subproblem (2)

Once V is computed, we get with $W = PVP^T$ that
$y = \hat{y} + \frac{1}{u}(A(W) - b)$
see: Helmberg, Rendl: SIOPT 10, (2000), 673ff

Update of P:
Having new point y, we evaluate f at y (sparse eigenvalue computation), which produces also an eigenvector v to $\lambda_{\max}$.
The vector v is added as new column to P, and P is purged by removing unnecessary other columns.
Convergence is slow, once close to optimum

- solve quadratic SDP of size k
- compute $\lambda_{\max}$ of matrix of order n

Last Slide

• Interior Point methods are fine and work robustly, but $n \leq 1000$ and $m \leq 10,000$ is a severe limit.
• If n small enough for matrix operations ($n \leq 2,000$), then projection methods allow to go to large m. These algorithms have weaker convergence properties and need some nontrivial parameter tuning.
• Partial Lagrangian duality can always be used to deal with only a part of the constraints explicitely. But we still need to solve some basic SDP and convergence of bundle methods for the Lagrangian dual may be slow.
• Currently, only spectral bundle is suitable as a general tool for very-large scale SDP.