Solving large Semidefinite Programs - Part 1 and 2

Franz Rendl

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

Alpen-Adria-Universität Klagenfurt

Austria

Overview

- Limits of Interior Point methods
- Lagrangian Relaxations and the Bundle idea
- Constant Trace SDP and the Spectral Bundle
- Augmented Lagrangian and Boundary Points

Interior-Point Methods to solve SDP (1)

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

Primal-Dual Path-following Methods:
At start of iteration: $(X \succ 0, y, Z \succ 0)$
Linearized system to be solved for $(\Delta X, \Delta y, \Delta Z)$:

$$A(\Delta X) = r_P := b - A(X) \quad \text{primal residue}$$

$$A^T(\Delta y) - \Delta Z = r_D := Z + C - A^T(y) \quad \text{dual residue}$$

$$Z\Delta X + \Delta Z X = \mu I - ZX \quad \text{path residue}$$

The last equation can be reformulated in many ways, which all are derived from the complementarity condition $ZX = 0$

Interior-Point Methods to solve SDP (2)

Direct approach with partial elimination:
Using the second and third equation to eliminate ΔX and ΔZ, and substituting into the first gives

$$\Delta Z = A^T(\Delta y) - r_D, \;\; \Delta X = \mu Z^{-1} - X - Z^{-1}\Delta Z X,$$

and the final system to be solved:

$$A(Z^{-1}A^T(\Delta y)X) = \mu A(Z^{-1}) - b + A(Z^{-1}r_D X)$$

Note that

$$A(Z^{-1}A^T(\Delta y)X) = M\Delta y,$$

but the $m \times m$ matrix M may be expensive to form.

Computational effort

- explicitely determine Z^{-1} $O(n^3)$
- several matrix multiplications $O(n^3)$
- final system of order m to compute Δy $O(m^3)$
- forming the final system matrix $O(mn^3 + m^2n^2)$
- line search to determine

$X^+ := X + t\Delta X, Z^+ := Z + t\Delta Z$ is at least $O(n^3)$

Effort to form system matrix depends on structure of $A(.)$
Limitations: $n \approx 1000,\ m \approx 5000$. See Mittelmann's site:

`http://plato.asu.edu/ftp/sdplib.html`

Example 1: Basic Max-Cut Relaxation

We solve $\max\langle L, X\rangle : \quad diag(X) = e, \quad X \succeq 0$.
Matrices of order n, and n simple equations $x_{ii} = 1$

n	seconds
200	2
400	7
600	16
800	35
1000	80
1500	260
2000	500

Computing times to solve the SDP on a PC (Pentium 4, 2.1 Ghz). Implementation in MATLAB, 30 lines of source code

Example 2: Lovasz Theta Function

Given a graph $G = (V, E)$ with $|V| = n, \ |E| = m$.

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

The number of constraints depends on the edge set $|E|$. If m is small, then this SDP can be solved efficiently

n	100	200	300	400
$\|E\|$	490	2050	4530	8000
time	2	52	470	2240
$\|E\|$	1240	5100	11250	20000
time	11	560	***	***

Times in seconds for computing $\theta(G)$ on random graphs with different densities(p = 0.1 and p=0.25).

Overview

- Limits of Interior Point methods
- Lagrangian Relaxations and the Bundle idea
- Constant Trace SDP and the Spectral Bundle
- Augmented Lagrangian and Boundary Points

What if m too large?

Now we consider

$$z^* := \max\langle C, X\rangle \text{ such that } A(X) = a,\ B(X) = b,\ X \succeq 0.$$

The idea: Optimizing over $A(X) = a$ without $B(X) = b$ is 'easy', but inclusion of $B(X) = b$ makes SDP difficult.
(Could also have inequalities $B(X) \leq b$.)
Partial Lagrangian Dual (y dual to b):

$$L(X, y) := \langle C, X\rangle + y^T(b - B(X))$$

Dual functional:

$$f(y) := \max_{A(X)=a,\ X\succeq 0} L(X, y).$$

Minimize $f(y)$ using Bundle Method

weak duality: $z^* \leq f(y) \ \forall y$
strong duality: $z^* = \min_y f(y)$.
Note: $f(y) = b^T y + \max_{x \in F} \langle C - B^T(y), X \rangle$
with $F := \{X : A(X) = a, \ X \succeq 0\}$.
(Matrix structure of X is not important.)
Basic assumption: We can compute $f(y)$ easily, yielding also maximizer X^* and $g^* := b - B(X^*)$.
$f(y) = b^T y + \langle C - B^T(y), X^* \rangle = y^T g^* + \langle C, X^* \rangle$,
$f(v) \geq v^T g^* + \langle C, X^* \rangle$, therefore

$$f(v) \geq f(y) + \langle g^*, v - y \rangle$$

(This means g^* is subgradient of f at y.)

Minimize f using Bundle Method (2)

Current iterate: $\hat{y}$ with maximizer $\hat{X}$, i.e. $f(\hat{y}) = L(\hat{X}, \hat{y})$. We also assume to have a 'bundle' of other $X_i \in F,\ i = 1, \ldots, k$ with $\hat{X}$ being one of them.
Compute $g_i := b - B(X_i), \phi_i := \langle C, X_i \rangle$.
Using subgradient inequalities for g_i we can minorize f by

$$f(y) \geq l(y) := \max_i \{\phi_i + \langle g_i, y \rangle\}$$

$$= \max_{\lambda \in \Lambda} \phi^T \lambda + \langle G\lambda, y \rangle.$$

Here $\lambda := \{\lambda : \lambda_i \geq 0,\ \sum_i \lambda_i = 1\}$. The key idea:

$$\min_y l(y) + \frac{1}{2t} \|y - \hat{y}\|^2$$

Minimize f using Bundle Method (3)

This is essentially convex quadratic programming in k variables. After exchanging min and max we get:

$$\max_{\lambda \in \Lambda} (\phi + G^T \hat{y})^T \lambda - \frac{t}{2} \|G\lambda\|^2,$$

and new trial point is given by

$$y = \hat{y} - tG\lambda.$$

Note: $G\lambda$ is convex combination of subgradients g_i.
We move in the direction of a subgradient of ’small’ norm!!

See Lemarechal, Kiwiel 1970s, Zowe, Shor, Nesterov 1980’

Computations: Max-Cut + triangles

Big graphs (from Helmberg). The number of bundle iterations is 50 for $n = 800$, and 30 for $n = 2000$.

problem	n	cut	initial bd	gap	fi nal bd	gap
G1	800	11612	12083.2	4.06	12005.4	3.39
G11	800	564	629.2	11.56	572.7	1.54
G14	800	3054	3191.6	4.51	3140.7	2.84
G18	800	985	1166.0	18.38	1063.4	7.96
G22	2000	13293	14135.9	6.34	14045.8	5.66
G27	2000	3293	4141.7	25.77	4048.4	22.94
G39	2000	2373	2877.7	21.27	2672.7	12.63

Overview

- Limits of Interior Point methods
- Lagrangian Relaxations and the Bundle idea
- Constant Trace SDP and the Spectral Bundle
- Augmented Lagrangian and Boundary Points

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 $W = PVP^T$.

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

Can approximately solve SDP with quite large matrices, $n \approx 5000$.

References

See web-site of Christoph Helmberg and his software packa-
ge SBMETHOD

Overview

- Limits of Interior Point methods
- Lagrangian Relaxations and the Bundle idea
- Constant Trace SDP and the Spectral Bundle
- Augmented Lagrangian and Boundary Points

Augmented Lagrangian Method

Augmented Lagrangian applied to (D)
$X \ldots$ 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 mimization: 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.

Alltogether 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).

Coordinatewise Minimization

If Z (and X) is kept constant, y given by unconstrained quadratic minimization:

$$\sigma A A^T y = \sigma A(C + Z) + A(X) - b$$

If y (and X) is kept constant, Z is given by projection onto PSD:

$$\min_{Z \succeq 0} \|Z - W(y)\|^2$$

Solved by eigenvalue decomposition of $W(y)$. Optimal Z given by $Z = W(y)_+$.

see also Burer and Vandenbussche (2004) for a similar approach applied to primal SDP

Why boundary-point method?

Observe that the update on X is given by

$$X \leftarrow X + \sigma(Z + C - A^T(y)) =$$

$$(X + \sigma C - \sigma A^T(y)) + \sigma Z = \sigma(-W(y) + W(y)_+) = -\sigma W(y)_- \succeq 0$$

We have

$$Z = W(y)_+, \; X = -\sigma W(y)_-$$

therefore X and Z are always in PSD and

$$ZX = 0.$$

Maintain complementarity and semidefiniteness. Once we reach primal and dual feasibility, we are optimal.

Inner stopping condition

Inner optimality conditions:

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

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

Equations defining Z and V hold for current y. So error occurs only in first equation.

$A(V) = A(\sigma(Z + C - A^T(y)) + X)$, so

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

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

Inner error is primal infeasibility of V.

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 $Z = W(y)_+, \;\; V = -\sigma W(y)_-$
- Update $X : \quad X = -\sigma W(y)_-$

Note: Outer stopping condition is dual feasibility.

See working paper: Malick, Povh, Rendl, Wiegele (Klagenfurt 2006)

Application to Theta Function

We solve

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

Constraints are of form

$$2x_{ij} = \langle E_{ij}, X\rangle = 0, \text{ with } E_{ij} = e_i e_j^T + e_j e_i^T.$$

But $\langle E_{ij}, E_{kl}\rangle \neq 0$ only if $[ij] = [kl]$, hence AA^T is diagonal.

Currently best computational results by Kim Toh (2003), and M. Kocvara (2005)

Theta Number revisited

Boundary point times in seconds to solve SDP, (Pentium 4, 3 Ghz), 7 digits accuracy

n	$\|E\|$	secs
200	10000	13
300	22500	40
400	40000	120
500	62500	170
600	90100	270
800	160000	670
1000	250000	1360

Interior Point method takes half hour for n=200

Convergence Analysis

Unfortunately, the convergence analysis of the boundary point method is rudimentary. We can show convergence, but no analysis of speed of convergence.
There are instances, where convergence is much slower.
Currently under investigation.

A technical report by Malick (Grenoble), Povh, Rendl and Wiegele (Klagenfurt) is under preparation.