The Spectral Bundle Method with Second Order Information

Franz Rendl
Institut für Mathematik
Universität Klagenfurt
A - 9020 Klagenfurt, Austria

franz.rendl@uni-klu.ac.at

Workshop Chemnitz, November 2004

Overview

- Eigenvalue Optimization and SDP
- Bundle Methods
- Second Order Models for Eigenvalue Optimization
- A practical version with Second order information

Eigenvalue optimization and SDP

Data:
$C, A_i, i = 1, \ldots, m$ symmetric matrices of order n
$b \in \mathbb{R}^m,\ a \in \mathbb{R}$

$$\min_y a\lambda_{\max}(C - A^T(y)) + b^T y$$

If $b = 0$, pure eigenvalue optimization.

Cullum et al (1975), Overton (1988), Oustry (2000)

Constant Trace SDP

$$(P) \quad \max\langle C, X\rangle \quad \text{such that } A(X) = b,\ X \succeq 0.$$

$$(D) \quad \min b^T y \quad \text{such that } A^T(y) - C = Z \succeq 0.$$

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

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

Constant Trace SDP (2)

The constant trace property implies:

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

Constant trace SDP are equivalent to

$$(E) \qquad \min_y a\lambda_{\max}(C - A^T(y)) + b^T y$$

see for instance Helmberg, R. (2000)

Optimality conditions for (E)

f has subdifferential $\partial f(y)$ at y given by

$$\partial f(y) = \{b - A(aW) : \\ \langle W, C - A^T(y) \rangle = \lambda_{\max}(C - A^T(y)),\ \mathrm{tr}(W) = 1,\ W \succeq 0\}.$$

P, U provide optimality certificate for minimizer y iff

$$P^T P = I_k,\ P^T(C - A^T(y))P = \lambda I_k,\ \lambda I \succeq C - A^T(y), \\ A(aPUP^T) = b,\ U \succeq 0,\ \mathrm{tr}(U) = 1.$$

Overview

- Eigenvalue Optimization and SDP
- Bundle Methods
- Second Order Models for Eigenvalue Optimization
- A practical version with Second order information

Bundle Methods to solve (E)

For given P we define the following auxiliary function

$$L_P(y, U) := a\langle C - A^T(y), PUP^T\rangle + b^T y$$

Here P takes the role of the Bundle of 'good eigenvectors'.

The key step is to solve

$$\min_y \max_{U \succeq 0,\ \mathrm{tr} U = 1} L_P(y, U) + \frac{t}{2}\|y - \hat{y}\|^2 =$$

$$\max L_P(y, U) + \frac{t}{2}\|y - \hat{y}\|^2 \text{ such that}$$

$$U \succeq 0,\ trU = 1,\ y = \hat{y} + \frac{1}{t}[aA(PUP^T) - b].$$

Generic Bundle Algorithm to minimize f

Data: $C, A_1, \ldots, A_m, b, a$

Input: $y \in I\!R^m$,   **Output:** $\hat{y} \in I\!R^m$

Start: evaluate f at y (to get $f(y)$ and eigenvector v)

Initialization: $\hat{y} = y,\ \hat{f} = f,\ P = v,$ select $t > 0$

while some stopping condition is not satisfied

(a) Solve

$$\min_{y} \max_{U \succeq 0,\ \mathrm{tr} U = 1} L_P(y, U) + \frac{t}{2}\|y - \hat{y}\|^2 \text{ giving } U, y$$

(b) Evaluate f at new point y (returning $f(y)$ and eigenvector v)

(c) Update P and $\hat{y}$ (serious or null step)

(d) Check the stopping condition

see e.g. Lemarechal, Kiwiel, Overton, Zowe, etc (1970-1990)

Bundle methods (2)

The difference between the standard bundle and spectral bundle lies in the definition of U.

(a) U diagonal leads to standard bundle
need to solve convex quadratic in k variables

(b) U general symmetric gives spectral bundle
need to solve quadratic SDP in matrix variable of order k

see Helmberg Habilitation thesis, and Helmberg, R. SIOPT (2000)

Overview

- Eigenvalue Optimization and SDP
- Bundle Methods
- Second Order Models for Eigenvalue Optimization
- A practical version with Second order information

Second order models

f is nonsmooth in general, but f is smooth on the submanifold, where the largest eigenvalue has constant multiplicity
(under some additional technical assumptions, which are not important)

Basic observation:
If y and $y + d$ have same multiplicity for $\lambda_{\max}$ we get second order expansion as follows:

We assume P is eigenspace of $\lambda_{\max}$:

$$P^T(C - A^T(y))P = \lambda I_k$$

Solve

$$\min \|b - aA(PUP^T)\|^2 \text{ such that } U \succeq 0,\ tr(U) = 1$$

Second order models (2)

U provides 'shortest' subgradient g at y:

$$g = b - aA(PUP^T)$$

(solve quadratic SDP of size k)
We also need full factorization of $C - A^T(y)$ given by P, Q
Use U to form Hessian **H(U)**,

$$H(U) := 2A[(PUP^T) \otimes (Q\tilde{\Lambda}^{-1}Q^T)]A^T$$

Computational effort to find H is nontrivial

see for instance Overton, Womersley 1993

Second order models (3)

Now use H to form prox term, before we had:

$$\min_{y} \max_{U\succeq 0,\ \mathbf{tr}U=1} L_P(y,U) + \frac{t}{2}\|y-\hat{y}\|^2$$

Now

$$\min_{y} \max_{U\succeq 0,\ \mathbf{tr}U=1} L_P(y,U) + \frac{1}{2}(y-\hat{y})^T H (y-\hat{y}) =$$

$$\max L_P(y,U) + \frac{1}{2}(y-\hat{y})^T H (y-\hat{y}) \text{ such that}$$

$$U \succeq 0,\ trU = 1,\ y = \hat{y} + H^{-1}[aA(PUP^T) - b].$$

Note that H^{-1} is used explicitly, therefore impractical.

Second order models (3)

Summary:

- Compute spectral decomposition of $C - A^T(\hat{y})$, (suppose $\lambda_{\max}$ has multiplicity k).
- Solve convex quadratic SDP in U of order k:
- Compute $H(U)$ using spectral decomposition
- Using $H(U)^{-1}$, solve another quadratic SDP in V or order k
- Compute new trial point

$$y = \hat{y} + H^{-1}[aA(PVP^T) - b]$$

Overview

- Eigenvalue Optimization and SDP
- Bundle Methods
- Second Order Models for Eigenvalue Optimization
- A practical version with Second order information

A Practical Variant

Avoid working with all of $H(U)$.

To avoid using inverse explicitly, take only diagonal of $H(U)$:

Still need factorization, but working with $G := \text{Diag}(H(U))$ simplifies the rest.

Amounts to diagonal scaling of update.

Most expensive steps per iteration, in addition to spectral bundle:

- full factorization
- Compute diagonal of H

Preliminary computational experiments are encouraging