overview

algorithms for solving semidefinite programs

- interior point methods
- spectral bundle methods
- bundle method
- projection methods

semidefinite programs: primal and dual

$$(\textbf{SDP})\begin{cases} \min & \langle C, X\rangle \\ \text{s.t.} & \mathcal{A}(X) = b \\ & X \succeq 0 \end{cases}$$

$$\min_{X\succeq 0}\max_{y\in\mathbb{R}^m} \langle C, X\rangle + \langle b - \mathcal{A}(X), y\rangle \geq \max_{y\in\mathbb{R}^m}\min_{X\succeq 0} \langle b, y\rangle + \langle X, C - \mathcal{A}^\top(y)\rangle$$

$$(\textbf{DSDP})\begin{cases} \max & b^\top y \\ \text{s.t.} & \mathcal{A}^\top(y) + Z = C \\ & y \in \mathbb{R}^m, Z \succeq 0 \end{cases}$$

algorithms for SDP

SDP for max-cut and ϑ-number

- max-cut: sizes of interest n around 100, strengthened relaxation leads to more than 10 000 constraints.

algorithms for SDP

SDP for max-cut and ϑ-number

- max-cut: sizes of interest n around 100, strengthened relaxation leads to more than 10 000 constraints.
- ϑ-number: sizes of interest $n \geq 500$, results in 100 000 constraints.

algorithms for SDP

SDP for max-cut and ϑ-number

- max-cut: sizes of interest n around 100, strengthened relaxation leads to more than 10 000 constraints.
- ϑ-number: sizes of interest $n \geq 500$, results in 100 000 constraints.

$\longrightarrow$ need other algorithmic machinery than interior point methods.

spectral bundle method

assume m and n are large
$\longrightarrow$ avoid Cholesky factorization, matrix multiplication,...

idea: get rid of $Z \succeq 0$ by using eigenvalue arguments.

spectral bundle method

$\mathcal{A}$ has constant trace property if I is in the range of $\mathcal{A}^\top$, i.e., $\exists\mu$ such that $\mathcal{A}^\top(\mu) = I$.

spectral bundle method

$\mathcal{A}$ has constant trace property if I is in the range of $\mathcal{A}^\top$, i.e., $\exists \mu$ such that $\mathcal{A}^\top(\mu) = I$.

The constant trace property implies:

$$\mathcal{A}(X) = b, \mathcal{A}^\top(\mu) = I$$

spectral bundle method

$\mathcal{A}$ has constant trace property if I is in the range of $\mathcal{A}^\top$, i.e., $\exists \mu$ such that $\mathcal{A}^\top(\mu) = I$.

The constant trace property implies:

$$\mathcal{A}(X) = b, \mathcal{A}^\top(\mu) = I$$
$$\implies \operatorname{trace}(X) = \langle I, X \rangle = \langle \mathcal{A}^\top(\mu), X \rangle = \langle \mu, \mathcal{A}(X) \rangle = \mu^\top b =: a$$

spectral bundle method

$\mathcal{A}$ has constant trace property if I is in the range of $\mathcal{A}^\top$, i.e., $\exists\mu$ such that $\mathcal{A}^\top(\mu) = I$.

The constant trace property implies:

$$\mathcal{A}(X) = b, \mathcal{A}^\top(\mu) = I$$
$$\implies \text{trace}(X) = \langle I, X \rangle = \langle \mathcal{A}^\top(\mu), X \rangle = \langle \mu, \mathcal{A}(X) \rangle = \mu^\top b =: a$$

constant trace property holds for many SDP derived from combinatorial optimization problems.

spectral bundle method

Reformulate dual as follows:

$$\min\{b^\top y \colon \mathcal{A}^\top(y) - C = Z \succeq 0\}$$

spectral bundle method

Reformulate dual as follows:

$$\min\{b^\top y \colon \mathcal{A}^\top(y) - C = Z \succeq 0\}$$

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

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

spectral bundle method

Reformulate dual as follows:

$$\min\{b^\top y \colon \mathcal{A}^\top(y) - C = Z \succeq 0\}$$

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

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

X^*,Z^* optimal

$\implies X^* Z^* = 0$, hence Z^* is singular and $\lambda_{min}(Z^*) = 0$.

spectral bundle method

Reformulate dual as follows:

$$\min\{b^\top y \colon \mathcal{A}^\top(y) - C = Z \succeq 0\}$$

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

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

X^*,Z^* optimal

$\Longrightarrow X^* Z^* = 0$, hence Z^* is singular and $\lambda_{min}(Z^*) = 0$.

$\longrightarrow$ used to compute dual variable λ explicitely

spectral bundle method

$$\lambda_{min}(Z^*) = 0 \iff \lambda_{\max}(-Z^*) = 0$$

spectral bundle method

$$\begin{aligned}\lambda_{min}(Z^*) = 0 \quad &\iff \quad \lambda_{\max}(-Z^*) = 0 \\ &\iff \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*) - \lambda^* I) = 0\end{aligned}$$

spectral bundle method

$$
\begin{aligned}
\lambda_{min}(Z^*) = 0 \quad &\Longleftrightarrow \quad \lambda_{\max}(-Z^*) = 0 \\
&\Longleftrightarrow \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*) - \lambda^* I) = 0 \\
&\Longleftrightarrow \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*)) - \lambda^* = 0
\end{aligned}
$$

spectral bundle method

$$
\begin{aligned}
\lambda_{min}(Z^*) = 0 &\iff \lambda_{\max}(-Z^*) = 0 \\
&\iff \lambda_{\max}(C - \mathcal{A}^\top(y^*) - \lambda^* I) = 0 \\
&\iff \lambda_{\max}(C - \mathcal{A}^\top(y^*)) - \lambda^* = 0 \\
&\iff \lambda^* = \lambda_{max}(C - \mathcal{A}^\top(y^*))
\end{aligned}
$$

spectral bundle method

$$
\begin{aligned}
\lambda_{min}(Z^*) = 0 \quad &\iff \quad \lambda_{\max}(-Z^*) = 0 \\
&\iff \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*) - \lambda^* I) = 0 \\
&\iff \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*)) - \lambda^* = 0 \\
&\iff \quad \lambda^* = \lambda_{max}(C - \mathcal{A}^\top(y^*))
\end{aligned}
$$

rewrite the dual:

$$\min\{b^\top y + \lambda_{\max}(C - \mathcal{A}^\top(y))\colon y \in \mathbb{R}^m\}$$

$\longrightarrow$ non-smooth unconstrained convex problem in y.

spectral bundle method

$$
\begin{aligned}
\lambda_{min}(Z^*) = 0 \quad &\iff \quad \lambda_{\max}(-Z^*) = 0 \\
&\iff \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*) - \lambda^* I) = 0 \\
&\iff \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*)) - \lambda^* = 0 \\
&\iff \quad \lambda^* = \lambda_{max}(C - \mathcal{A}^\top(y^*))
\end{aligned}
$$

rewrite the dual:

$$\min\{b^\top y + \lambda_{\max}(C - \mathcal{A}^\top(y)) \colon y \in \mathbb{R}^m\}$$

$\longrightarrow$ non-smooth unconstrained convex problem in y.

note: evaluating $f(y) = b^\top y + \lambda_{\max}(C - \mathcal{A}^\top(y))$ amounts in computing largest eigenvalue of $C - \mathcal{A}^\top(y)$

spectral bundle method

$$\begin{aligned}
\lambda_{min}(Z^*) = 0 \quad &\iff \quad \lambda_{\max}(-Z^*) = 0 \\
&\iff \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*) - \lambda^* I) = 0 \\
&\iff \quad \lambda_{\max}(C - \mathcal{A}^\top(y^*)) - \lambda^* = 0 \\
&\iff \quad \lambda^* = \lambda_{max}(C - \mathcal{A}^\top(y^*))
\end{aligned}$$

rewrite the dual:

$$\min\{b^\top y + \lambda_{\max}(C - \mathcal{A}^\top(y)) \colon y \in \mathbb{R}^m\}$$

$\longrightarrow$ non-smooth unconstrained convex problem in y.

note: evaluating $f(y) = b^\top y + \lambda_{\max}(C - \mathcal{A}^\top(y))$ amounts in computing largest eigenvalue of $C - \mathcal{A}^\top(y)$
$\longrightarrow$ can be done by iterative methods even for very large (sparse) matrices.

spectral bundle method

$$\lambda_{\max}(X) = \max\{\langle X, W \rangle \colon \operatorname{trace}(W) = 1, W \succeq 0\}$$

spectral bundle method

$$\lambda_{\max}(X) = \max\{\langle X, W\rangle \colon \operatorname{trace}(W) = 1, W \succeq 0\}$$

Define

$$\mathcal{L}(W, y) := \langle C - \mathcal{A}^\top(y), W\rangle + b^\top y$$

hence,

$$f(y) = \max\{\mathcal{L}(W, y) \colon \operatorname{trace}(W) = 1, W \succeq 0\}$$

spectral bundle method

$$\lambda_{\max}(X) = \max\{\langle X, W\rangle \colon \operatorname{trace}(W) = 1, W \succeq 0\}$$

Define

$$\mathcal{L}(W, y) := \langle C - \mathcal{A}^\top(y), W\rangle + b^\top y$$

hence,

$$f(y) = \max\{\mathcal{L}(W, y) \colon \operatorname{trace}(W) = 1, W \succeq 0\}$$

two ingredients:

- minorant for $f(y)$

spectral bundle method

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

Define

$$\mathcal{L}(W, y) := \langle C - \mathcal{A}^\top(y), W\rangle + b^\top y$$

hence,

$$f(y) = \max\{\mathcal{L}(W, y) : \text{trace}(W) = 1, W \succeq 0\}$$

two ingredients:

- minorant for $f(y)$
- proximal point approach

spectral bundle method

Idea 1: Minorant for $f(y)$

spectral bundle method

Idea 1: Minorant for $f(y)$

Fix some $m \times k$ matrix P. $k \geq 1$ can be chosen arbitrarily
[choice of P will be explained later]

spectral bundle method

Idea 1: Minorant for $f(y)$

Fix some $m \times k$ matrix P. $k \geq 1$ can be chosen arbitrarily
[choice of P will be explained later]

replace feasible region $\{W \colon \operatorname{trace}(W) = 1, W \succeq 0\}$ by computational more practical subset

$$\mathcal{W} = \{W \colon W = PVP^\top, \operatorname{trace}(V) = 1, V \succeq 0\}$$

with new $k \times k$ matrix variable V.

spectral bundle method

Idea 1: Minorant for $f(y)$

Fix some $m \times k$ matrix P. $k \geq 1$ can be chosen arbitrarily
[choice of P will be explained later]

replace feasible region $\{W : \text{trace}(W) = 1, W \succeq 0\}$ by computational more practical subset

$$\mathcal{W} = \{W : W = PVP^\top, \text{trace}(V) = 1, V \succeq 0\}$$

with new $k \times k$ matrix variable V.

$$\hat{f}(y) := \max\{\mathcal{L}(W, y) : W = PVP^\top, \text{trace}(V) = 1, V \succeq 0\}$$

spectral bundle method

Idea 1: Minorant for $f(y)$

Fix some $m \times k$ matrix P. $k \geq 1$ can be chosen arbitrarily
[choice of P will be explained later]

replace feasible region $\{W : \text{trace}(W) = 1, W \succeq 0\}$ by computational more practical subset

$$\mathcal{W} = \{W : W = PVP^\top, \text{trace}(V) = 1, V \succeq 0\}$$

with new $k \times k$ matrix variable V.

$$\hat{f}(y) := \max\{\mathcal{L}(W, y) : W = PVP^\top, \text{trace}(V) = 1, V \succeq 0\} \leq f(y)$$

spectral bundle method

Idea 2: Proximal point approach

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

spectral bundle method

Idea 2: Proximal point approach

The function $\hat{f}$ (depending on P) will be a good approximation to $f(y)$ only in some neighbourhood of the current iterate $\hat{y}$.
$\longrightarrow$ penalize displacement by adding $\|y - \hat{y}\|^2$.

spectral bundle method

Idea 2: Proximal point approach

The function $\hat{f}$ (depending on P) will be a good approximation to $f(y)$ only in some neighbourhood of the current iterate $\hat{y}$.
$\longrightarrow$ penalize displacement by adding $\|y - \hat{y}\|^2$.

Instead of minimizing $f(y)$ we

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

spectral bundle method

Idea 2: Proximal point approach

The function $\hat{f}$ (depending on P) will be a good approximation to $f(y)$ only in some neighbourhood of the current iterate $\hat{y}$.
$\longrightarrow$ penalize displacement by adding $\|y - \hat{y}\|^2$.

Instead of minimizing $f(y)$ we

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

This is a strictly convex function, if $u > 0$ is fixed.

substituting the definition of $\hat{f}$...

spectral bundle method

$$\min_{y \in \mathbb{R}^m} \hat{f}(y) + \frac{u}{2}\|y - \hat{y}\|^2 =$$

spectral bundle method

$$\min_{y\in\mathbb{R}^m} \hat{f}(y) + \frac{u}{2}\|y - \hat{y}\|^2 =$$

$$= \quad \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \mathcal{L}(W, y) + \frac{u}{2}\|y - \hat{y}\|^2 =$$

spectral bundle method

$$
\begin{aligned}
&\min_{y\in\mathbb{R}^m} \hat{f}(y) + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \quad \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \mathcal{L}(W,y) + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \quad \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \langle C - \mathcal{A}^\top(y), W\rangle + b^\top y + \frac{u}{2}\|y-\hat{y}\|^2 =
\end{aligned}
$$

spectral bundle method

$$
\begin{aligned}
&\min_{y\in\mathbb{R}^m} \hat{f}(y) + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \quad \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \mathcal{L}(W,y) + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \quad \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \langle C - \mathcal{A}^\top(y), W\rangle + b^\top y + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \quad \max_{W\in\mathcal{W}} \{\langle C - \mathcal{A}^\top(y), W\rangle + b^\top y + \frac{u}{2}\|y-\hat{y}\|^2 : \\
&\qquad\qquad y = \hat{y} + \frac{1}{u}(\mathcal{A}(W) - b)\} =
\end{aligned}
$$

spectral bundle method

$$
\begin{aligned}
&\min_{y\in\mathbb{R}^m} \hat{f}(y) + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \mathcal{L}(W,y) + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \langle C - \mathcal{A}^\top(y), W\rangle + b^\top y + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \max_{W\in\mathcal{W}} \{\langle C - \mathcal{A}^\top(y), W\rangle + b^\top y + \frac{u}{2}\|y-\hat{y}\|^2 : \\
&\qquad y = \hat{y} + \frac{1}{u}(\mathcal{A}(W) - b)\} = \\
&= \max_{W\in\mathcal{W}} \{\langle C - \mathcal{A}^\top(\hat{y}), W\rangle + b^\top \hat{y} - \frac{1}{2u}\|\mathcal{A}(W) - b\|^2
\end{aligned}
$$

spectral bundle method

$$
\begin{aligned}
&\min_{y\in\mathbb{R}^m} \hat{f}(y) + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \mathcal{L}(W,y) + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \min_{y\in\mathbb{R}^m} \max_{W\in\mathcal{W}} \langle C-\mathcal{A}^\top(y), W\rangle + b^\top y + \frac{u}{2}\|y-\hat{y}\|^2 = \\
&= \max_{W\in\mathcal{W}} \{\langle C-\mathcal{A}^\top(y), W\rangle + b^\top y + \frac{u}{2}\|y-\hat{y}\|^2 : \\
&\qquad y = \hat{y} + \frac{1}{u}(\mathcal{A}(W)-b)\} = \\
&= \max_{W\in\mathcal{W}} \{\langle C-\mathcal{A}^\top(\hat{y}), W\rangle + b^\top \hat{y} - \frac{1}{2u}\|\mathcal{A}(W)-b\|^2
\end{aligned}
$$

quadratic SDP in the $k \times k$ matrix variable V, since $W = PVP^\top$. k is user defined and can be small, in particular, it is independent of n.

spectral bundle method

- solve maximization problem (using interior point methods) to obtain V and thus $W = PVP^\top$
- $y_{new} = \hat{y} + \frac{1}{u}(\mathcal{A}(W) - b)$

spectral bundle method

- solve maximization problem (using interior point methods) to obtain V and thus $W = PVP^\top$
- $y_{new} = \hat{y} + \frac{1}{u}(\mathcal{A}(W) - b)$

What is P?

spectral bundle method

- solve maximization problem (using interior point methods) to obtain V and thus $W = PVP^\top$
- $y_{new} = \hat{y} + \frac{1}{u}(\mathcal{A}(W) - b)$

What is P?

Having point y_{new}, evaluation $f(y_{new})$ (sparse eigenvalue computation) produces also an eigenvector v to $\lambda_{\max}$.

spectral bundle method

- solve maximization problem (using interior point methods) to obtain V and thus $W = PVP^\top$
- $y_{new} = \hat{y} + \frac{1}{u}(\mathcal{A}(W) - b)$

What is P?

Having point y_{new}, evaluation $f(y_{new})$ (sparse eigenvalue computation) produces also an eigenvector v to $\lambda_{\max}$.

$\longrightarrow$ eigenvector v is added as new column to P,
and P is purged by removing unnecessary other columns.

spectral bundle method

computational effort:

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

spectral bundle method

computational effort:

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

software `SBmethod` a C++ implementation of the spectral bundle method of [Helmberg and Rendl 00; Helmberg and Kiwiel 99] no longer supported, but now there is the ConicBundle callable library instead; available at `https://www-user.tu-chemnitz.de/~helmberg/ConicBundle/`

spectral bundle method

example: consider again the basic max-cut relaxation

$$\max\{\langle L, X\rangle \colon \operatorname{diag}(X) = e, X \succeq 0\}$$

Now $20\,000 \leq n \leq 50\,000$, sparse graphs.

spectral bundle method

n	upper-bnd	cut	time (secs)
20,000	143.3	131.3	330
20,000	261.9	244.8	536
20,000	598.1	571.1	1255
30,000	214.9	197.2	753
30,000	393.3	367.4	990
30,000	897.9	857.3	2330
40,000	286.9	262.7	1180
40,000	524.6	489.8	1650
50,000	358.9	328.5	1800

spectral bundle method

spectral bundle method summarized

- using eigenvalue optimization and classical methods from convex analysis
- general tool for solving SDP having matrices of large dimension
- convergence is slow, once close to optimum

overview

Algorithms for solving semidefinite programs

- interior point methods
- spectral bundle methods
- bundle method
- projection methods

bundle method

We would like to compute

$$z^* = \max\{\langle C, X\rangle \colon \mathcal{A}(X) = a,\ \mathcal{B}(X) = b,\ X \succeq 0\}$$

Optimizing over $\mathcal{A}(X) = a, X \succeq 0$ without $\mathcal{B}(X) = b$ is "easy", but inclusion of $\mathcal{B}(X) = b$ makes SDP difficult.

bundle method

We would like to compute

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

Optimizing over $\mathcal{A}(X) = a, X \succeq 0$ without $\mathcal{B}(X) = b$ is "easy", but inclusion of $\mathcal{B}(X) = b$ makes SDP difficult.
partial Lagrangian dual (y dual to $\mathcal{B}(X) = b$):

$$\mathcal{L}(X; y) = \langle C, X\rangle + y^\top (b - \mathcal{B}(X))$$

dual functional:

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

bundle method

We would like to compute

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

Optimizing over $\mathcal{A}(X) = a, X \succeq 0$ without $\mathcal{B}(X) = b$ is "easy", but inclusion of $\mathcal{B}(X) = b$ makes SDP difficult.
partial Lagrangian dual (y dual to $\mathcal{B}(X) = b$):

$$\mathcal{L}(X; y) = \langle C, X\rangle + y^\top (b - \mathcal{B}(X))$$

dual functional:

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

and thus

$$z^* = \min_{y \in \mathbb{R}^m} f(y)$$

bundle method

$$z^* = \min_{y \in \mathbb{R}^m} f(y)$$

with

$$f(y) \quad = \quad \max_{\mathcal{A}(X)=a, X \succeq 0} \langle C, X \rangle + y^\top (b - \mathcal{B}(X))$$

bundle method

$$z^* = \min_{y \in \mathbb{R}^m} f(y)$$

with

$$\begin{aligned} f(y) &= \max_{\mathcal{A}(X)=a, X \succeq 0} \langle C, X \rangle + y^\top (b - \mathcal{B}(X)) \\ &= b^\top y + \max_{\mathcal{A}(X)=a, X \succeq 0} \langle C - \mathcal{B}^\top(y), X \rangle \end{aligned}$$

bundle method

$$z^* = \min_{y \in \mathbb{R}^m} f(y)$$

with

$$\begin{aligned} f(y) &= \max_{\mathcal{A}(X)=a, X \succeq 0} \langle C, X \rangle + y^\top (b - \mathcal{B}(X)) \\ &= b^\top y + \max_{\mathcal{A}(X)=a, X \succeq 0} \langle C - \mathcal{B}^\top(y), X \rangle \end{aligned}$$

evaluating $f(y)$ amounts in solving an SDP.

bundle method

$$z^* = \min_{y \in \mathbb{R}^m} f(y)$$

with

$$\begin{aligned} f(y) &= \max_{\mathcal{A}(X)=a, X \succeq 0} \langle C, X \rangle + y^\top (b - \mathcal{B}(X)) \\ &= b^\top y + \max_{\mathcal{A}(X)=a, X \succeq 0} \langle C - \mathcal{B}^\top(y), X \rangle \end{aligned}$$

evaluating $f(y)$ amounts in solving an SDP.

basic assumption: we can evaluate $f(y)$ easily, yielding also a maximizer X^* and $g^* = b - \mathcal{B}(X^*)$.

bundle method

using $g^* = b - \mathcal{B}(X^*)$ if X^* is the optimizer for given $\bar{y}$:

$$f(\bar{y}) \quad = \quad b^\top \bar{y} + \langle C - \mathcal{B}^\top(\bar{y}), X^* \rangle =$$

bundle method

using $g^* = b - \mathcal{B}(X^*)$ if X^* is the optimizer for given $\bar{y}$:

$$
\begin{aligned}
f(\bar{y}) &= b^\top \bar{y} + \langle C - \mathcal{B}^\top(\bar{y}), X^* \rangle = \\
&= \langle b - \mathcal{B}^\top(X^*), \bar{y} \rangle + \langle C, X^* \rangle = \langle g^*, \bar{y} \rangle + \langle C, X^* \rangle
\end{aligned}
$$

bundle method

using $g^* = b - \mathcal{B}(X^*)$ if X^* is the optimizer for given $\bar{y}$:

$$\begin{aligned} f(\bar{y}) &= b^\top \bar{y} + \langle C - \mathcal{B}^\top(\bar{y}), X^* \rangle = \\ &= \langle b - \mathcal{B}^\top(X^*), \bar{y} \rangle + \langle C, X^* \rangle = \langle g^*, \bar{y} \rangle + \langle C, X^* \rangle \end{aligned}$$

for any y we have

$$f(y) \geq b^\top y + \langle C - \mathcal{B}^\top(y), X^* \rangle =$$

bundle method

using $g^* = b - \mathcal{B}(X^*)$ if X^* is the optimizer for given $\bar{y}$:

$$\begin{aligned} f(\bar{y}) &= b^\top \bar{y} + \langle C - \mathcal{B}^\top(\bar{y}), X^* \rangle = \\ &= \langle b - \mathcal{B}^\top(X^*), \bar{y} \rangle + \langle C, X^* \rangle = \langle g^*, \bar{y} \rangle + \langle C, X^* \rangle \end{aligned}$$

for any y we have

$$\begin{aligned} f(y) &\geq b^\top y + \langle C - \mathcal{B}^\top(y), X^* \rangle = \\ &= \langle b - \mathcal{B}(X^*), y \rangle + \langle C, X^* \rangle = \langle g^*, y \rangle + \langle C, X^* \rangle \end{aligned}$$

bundle method

using $g^* = b - \mathcal{B}(X^*)$ if X^* is the optimizer for given $\bar{y}$:

$$\begin{aligned} f(\bar{y}) &= b^\top \bar{y} + \langle C - \mathcal{B}^\top(\bar{y}), X^* \rangle = \\ &= \langle b - \mathcal{B}^\top(X^*), \bar{y} \rangle + \langle C, X^* \rangle = \langle g^*, \bar{y} \rangle + \langle C, X^* \rangle \end{aligned}$$

for any y we have

$$\begin{aligned} f(y) &\geq b^\top y + \langle C - \mathcal{B}^\top(y), X^* \rangle = \\ &= \langle b - \mathcal{B}(X^*), y \rangle + \langle C, X^* \rangle = \langle g^*, y \rangle + \langle C, X^* \rangle \end{aligned}$$

combining, we get:

$$f(y) - f(\bar{y}) \geq \langle g^*, y \rangle + \langle C, X^* \rangle - \langle g^*, \bar{y} \rangle - \langle C, X^* \rangle$$

i.e.,

bundle method

using $g^* = b - \mathcal{B}(X^*)$ if X^* is the optimizer for given $\bar{y}$:

$$\begin{aligned} f(\bar{y}) &= b^\top \bar{y} + \langle C - \mathcal{B}^\top(\bar{y}), X^* \rangle = \\ &= \langle b - \mathcal{B}^\top(X^*), \bar{y} \rangle + \langle C, X^* \rangle = \langle g^*, \bar{y} \rangle + \langle C, X^* \rangle \end{aligned}$$

for any y we have

$$\begin{aligned} f(y) &\geq b^\top y + \langle C - \mathcal{B}^\top(y), X^* \rangle = \\ &= \langle b - \mathcal{B}(X^*), y \rangle + \langle C, X^* \rangle = \langle g^*, y \rangle + \langle C, X^* \rangle \end{aligned}$$

combining, we get:

$$f(y) - f(\bar{y}) \geq \langle g^*, y \rangle + \langle C, X^* \rangle - \langle g^*, \bar{y} \rangle - \langle C, X^* \rangle$$

i.e.,

$$f(y) \geq f(\bar{y}) + \langle g^*, y - \bar{y} \rangle \qquad \forall y$$

hence $g* = b - \mathcal{B}(X^*)$ is subgradient.

bundle method

since

$$z^* = \min_{y \in \mathbb{R}^m} f(y)$$

bundle method

since

$$z^* = \min_{y \in \mathbb{R}^m} f(y) \leq f(\tilde{y}) \quad \forall \tilde{y} \in \mathbb{R}^m$$

any $\tilde{y} \in \mathbb{R}^m$ provides upper bound on z^*

bundle method

since

$$z^* = \min_{y \in \mathbb{R}^m} f(y) \leq f(\tilde{y}) \quad \forall \tilde{y} \in \mathbb{R}^m$$

any $\tilde{y} \in \mathbb{R}^m$ provides upper bound on z^*
$\longrightarrow$ try to find tight upper bound (i.e., approximate minimizer of $f(y)$) by using bundle methods.

bundle method

two ingredients:

- work with a “bundle” of X_i's and maximize over $\mathrm{conv}\{X_1, \ldots, X_k\}$ instead of over $\{\mathcal{A}(X) = a, X \succeq 0\}$
- penalize displacement from current iterate, i.e., add penalty term $\frac{1}{2t}\|y - \hat{y}\|^2$

bundle method

two ingredients:

- work with a "bundle" of X_i's and maximize over $\mathrm{conv}\{X_1, \ldots, X_k\}$ instead of over $\{\mathcal{A}(X) = a, X \succeq 0\}$
- penalize displacement from current iterate, i.e., add penalty term $\frac{1}{2t}\|y - \hat{y}\|^2$

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

$$\hat{f}(y) = \max_{X \in \mathrm{conv}\{X_1, \ldots, X_k\}} \mathcal{L}(X; y)$$

bundle method

iterative procedure:

- solve approximately

$$\min_{\gamma \geq 0} \hat{f}(y) + \frac{1}{2t}||y - \hat{y}||^2$$

where

$$\hat{f}(y) = \max\{\mathcal{L}(X; y) \colon X \in \mathrm{conv}\{X_1, \ldots, X_k\}$$

giving y

bundle method

iterative procedure:

- solve approximately

$$\min_{\gamma \geq 0} \hat{f}(y) + \frac{1}{2t}||y - \hat{y}||^2$$

  where

$$\hat{f}(y) = \max\{\mathcal{L}(X; y) \colon X \in \operatorname{conv}\{X_1, \ldots, X_k\}$$

  giving y
- evaluate

$$f(y) = b^\top y + \max\{\langle C - \mathcal{B}^\top(y), X\rangle \colon \mathcal{A}(X) = a,\ X \succeq 0\}$$

bundle method

iterative procedure:

- solve approximately

$$\min_{\gamma \geq 0} \hat{f}(y) + \frac{1}{2t}||y - \hat{y}||^2$$

  where

$$\hat{f}(y) = \max\{\mathcal{L}(X; y) \colon X \in \operatorname{conv}\{X_1, \dots, X_k\}$$

  giving y
- evaluate

$$f(y) = b^\top y + \max\{\langle C - \mathcal{B}^\top(y), X\rangle \colon \mathcal{A}(X) = a,\ X \succeq 0\}$$

computational effort in each iteration:

- solve a convex quadratic program in k variables
- evaluate $f(\hat{y})$ to yield new $\hat{X}$ and a subgradient $\hat{g}$

bundle method

example: solving the max-cut relaxation strengthened by triangle-inequalities:

$$\max\{\langle L, X\rangle : \operatorname{diag}(X) = e, \mathcal{M}(X) \leq -e, X \succeq 0\}$$

bundle method

example: solving the max-cut relaxation strengthened by triangle-inequalities:

$$\max\{\langle L, X\rangle \colon \operatorname{diag}(X) = e, \mathcal{M}(X) \le -e, X \succeq 0\}$$

dualize the triangle-inequalities $\mathcal{M}(X) \le -e$, evaluating $f(y)$ is then solving

$$\max\{\langle L - \mathcal{M}^\top(y), X\rangle \colon \operatorname{diag}(X) = e, X \succeq 0\}$$

bundle method

example: solving the max-cut relaxation strengthened by triangle-inequalities:

$$\max\{\langle L, X\rangle \colon \operatorname{diag}(X) = e, \mathcal{M}(X) \leq -e, X \succeq 0\}$$

dualize the triangle-inequalities $\mathcal{M}(X) \leq -e$, evaluating $f(y)$ is then solving

$$\max\{\langle L - \mathcal{M}^\top(y), X\rangle \colon \operatorname{diag}(X) = e, X \succeq 0\}$$

SDP with matrix dimension n and only n linear constraints.

bundle method

SDP relaxation for max-cut; triangle inequalities are dualized.
50 bundle iterations for $n = 800$, and 30 for $n = 2000$.

graph	n	initial gap (%)	final (%)	time (secs)
G6	800	22.29	18.15	43.11
G11	800	11.56	1.54	60.20
G14	800	4.51	2.84	59.68
G18	800	18.38	7.96	69.19
G22	2000	6.34	5.66	278.06
G27	2000	25.77	22.94	406.66
G39	2000	21.27	12.63	533.36

see [Fischer, Gruber, Rendl, Sotirov, 06]

bundle method

software: ConicBundle C++ library of Ch. Helmberg, available at
`https://www-user.tu-chemnitz.de/~helmberg/ConicBundle/`

bundle method

bundle method summarized

- in combination with interior point methods is a good tool to approximate SDPs with a huge number of constraints
- the number of function evaluations to reach good approximations is surprisingly small
- getting to the “real” optimum is hard

overview

algorithms for solving semidefinite programs

- interior point methods
- spectral bundle method
- bundle methods
- projection methods

augmented Lagrange algorithm

$$\min f(x) \quad \text{such that} \quad x \in \mathcal{X},\ h(x) = 0$$

augmented Lagrange algorithm

$$\min f(x) \quad \text{such that} \quad x \in \mathcal{X},\ h(x) = 0$$

$f : \mathbb{R}^n \mapsto \mathbb{R}$, $h : \mathbb{R}^n \mapsto \mathbb{R}^m$ sufficiently smooth functions, $\mathcal{X} \subseteq \mathbb{R}^n$ nonempty closed convex set of simple structure

augmented Lagrange algorithm

$$\min f(x) \quad \text{such that} \quad x \in \mathcal{X},\ h(x) = 0$$

$f : \mathbb{R}^n \mapsto \mathbb{R}$, $h : \mathbb{R}^n \mapsto \mathbb{R}^m$ sufficiently smooth functions, $\mathcal{X} \subseteq \mathbb{R}^n$ nonempty closed convex set of simple structure

$$\mathcal{L}_\sigma(x, y) := f(x) + y^\top h(x) + \frac{\sigma}{2}\|h(x)\|^2$$

augmented Lagrange algorithm

$$\min f(x) \quad \text{such that} \quad x \in \mathcal{X},\ h(x) = 0$$

$f : \mathbb{R}^n \mapsto \mathbb{R}$, $h : \mathbb{R}^n \mapsto \mathbb{R}^m$ sufficiently smooth functions, $\mathcal{X} \subseteq \mathbb{R}^n$ nonempty closed convex set of simple structure

$$\mathcal{L}_\sigma(x, y) := f(x) + y^\top h(x) + \frac{\sigma}{2}\|h(x)\|^2$$

repeat until convergence
(a) Keep y fixed: solve $\min_x \mathcal{L}_\sigma(x, y)$ to get x
(b) update y: $y \leftarrow y + \sigma h(x)$
(c) update σ

Original version: Powell, Hestenes, 1969

augmented Lagrange algorithm

$$(\textbf{DSDP}) \qquad \min \; b^\top y \quad \text{s.t. } \mathcal{A}^\top(y) - C = Z, \; Z \succeq 0$$

augmented Lagrange algorithm

$$(\textbf{DSDP}) \qquad \min \; b^\top y \quad \text{s.t. } \mathcal{A}^\top(y) - C = Z, \; Z \succeq 0$$

$$\mathcal{L}_\sigma(y, Z; X) = \; b^\top y + \langle X, Z + C - \mathcal{A}^\top(y) \rangle + \frac{\sigma}{2} \| Z + C - \mathcal{A}^\top(y) \|^2$$

projection methods

inner minimization

$$\min_{y, Z \succeq 0} \mathcal{L}_{\sigma_k}(y, Z; X)$$

projection methods

inner minimization

$$\min_{y, Z \succeq 0} \mathcal{L}_{\sigma_k}(y, Z; X)$$

$$\text{Define } \mathcal{W}(y) := \mathcal{A}^\top(y) - C - \frac{1}{\sigma}X$$

projection methods

inner minimization

$$\min_{y,Z\succeq 0} \mathcal{L}_{\sigma_k}(y,Z;X)$$

$$\text{Define } \mathcal{W}(y) := \mathcal{A}^\top(y) - C - \frac{1}{\sigma}X$$

$$\begin{aligned}\mathcal{L}_\sigma(y,Z;X) &= b^\top y + \langle X, Z + C - \mathcal{A}^\top(y)\rangle + \tfrac{\sigma}{2}\|Z + C - \mathcal{A}^\top(y)\|^2 \\ &= b^\top y + \tfrac{1}{\sigma}\|Z - \mathcal{W}(y)\|^2 - \tfrac{1}{2\sigma}\|X\|^2\end{aligned}$$

projection methods

inner minimization

$$\min_{y, Z \succeq 0} \mathcal{L}_{\sigma_k}(y, Z; X)$$

$$\text{Define } \mathcal{W}(y) := \mathcal{A}^\top(y) - C - \frac{1}{\sigma} X$$

$$\begin{array}{rcl} \mathcal{L}_\sigma(y, Z; X) & = & b^\top y + \langle X, Z + C - \mathcal{A}^\top(y) \rangle + \frac{\sigma}{2} \| Z + C - \mathcal{A}^\top(y) \|^2 \\ & = & b^\top y + \frac{1}{\sigma} \| Z - \mathcal{W}(y) \|^2 - \frac{1}{2\sigma} \| X \|^2 \end{array}$$

$$\min_{y, Z \succeq 0} b^\top y + \frac{1}{\sigma} \| Z - \mathcal{W}(y) \|^2$$

projection methods

inner minimization: optimality conditions

$$\mathcal{L}_\sigma(y, Z; V) = b^\top y + \frac{1}{\sigma}\|Z - \mathcal{W}(y)\|^2 - \langle V, Z\rangle$$

projection methods

inner minimization: optimality conditions

$$\mathcal{L}_\sigma(y, Z; V) = b^\top y + \frac{1}{\sigma}\|Z - \mathcal{W}(y)\|^2 - \langle V, Z\rangle$$

$$\nabla_y \mathcal{L} = 0 \Rightarrow \sigma\mathcal{A}(\mathcal{A}^\top(y)) = \sigma\mathcal{A}(Z + C) + (\mathcal{A}(X) - b)$$

projection methods

inner minimization: optimality conditions

$$\mathcal{L}_\sigma(y, Z; V) = b^\top y + \frac{1}{\sigma}\|Z - \mathcal{W}(y)\|^2 - \langle V, Z\rangle$$

$$\nabla_y \mathcal{L} = 0 \Rightarrow \sigma\mathcal{A}(\mathcal{A}^\top(y)) = \sigma\mathcal{A}(Z + C) + (\mathcal{A}(X) - b)$$

$$\nabla_Z \mathcal{L} = 0 \Rightarrow V = \sigma(Z - \mathcal{W}(y))$$

projection methods

inner minimization: optimality conditions

$$\mathcal{L}_\sigma(y, Z; V) = b^\top y + \frac{1}{\sigma}\|Z - \mathcal{W}(y)\|^2 - \langle V, Z\rangle$$

$$\nabla_y \mathcal{L} = 0 \Rightarrow \sigma\mathcal{A}(\mathcal{A}^\top(y)) = \sigma\mathcal{A}(Z + C) + (\mathcal{A}(X) - b)$$

$$\nabla_Z \mathcal{L} = 0 \Rightarrow V = \sigma(Z - \mathcal{W}(y))$$

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

projection methods

solve coordinatewise:
keep Z (and X) constant, y is given by the unconstrained minimization

$$\sigma\mathcal{A}(\mathcal{A}^\top(y)) = \sigma\mathcal{A}(Z + C) + (\mathcal{A}(X) - b)$$

projection methods

solve coordinatewise:
keep Z (and X) constant, y is given by the unconstrained minimization

$$\sigma\mathcal{A}(\mathcal{A}^\top(y)) = \sigma\mathcal{A}(Z + C) + (\mathcal{A}(X) - b)$$

keep y (and X) constant, Z is given by the projection onto the positive semidefinite cone

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

projection methods

solve coordinatewise:
keep Z (and X) constant, y is given by the unconstrained minimization

$$\sigma \mathcal{A}(\mathcal{A}^\top(y)) = \sigma \mathcal{A}(Z + C) + (\mathcal{A}(X) - b)$$

keep y (and X) constant, Z is given by the projection onto the positive semidefinite cone

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

$$Z = \mathcal{W}(y)_+$$

projection methods

boundary point method

Initialization: $k = 0$, select $\sigma_k > 0$, $X_k \succeq 0, Z_k \succeq 0$

projection methods

boundary point method

Initialization: $k = 0$, select $\sigma_k > 0$, $X_k \succeq 0, Z_k \succeq 0$

repeat until convergence

(a) Keep X_k fixed.

repeat until convergence

projection methods

boundary point method

Initialization: $k = 0$, select $\sigma_k > 0$, $X_k \succeq 0, Z_k \succeq 0$

repeat until convergence

(a) Keep X_k fixed.

repeat until convergence

– solve $\mathcal{A}(\mathcal{A}^\top(y)) = rhs$ giving y_k

projection methods

boundary point method

Initialization: $k = 0$, select $\sigma_k > 0$, $X_k \succeq 0, Z_k \succeq 0$

repeat until convergence

(a) Keep X_k fixed.

repeat until convergence

– solve $\mathcal{A}(\mathcal{A}^\top(y)) = rhs$ giving y_k

– compute $Z_k = \mathcal{W}(y_k)_+$

projection methods

boundary point method

Initialization: $k = 0$, select $\sigma_k > 0$, $X_k \succeq 0, Z_k \succeq 0$

repeat until convergence

(a) Keep X_k fixed.

repeat until convergence

– solve $\mathcal{A}(\mathcal{A}^\top(y)) = rhs$ giving y_k

– compute $Z_k = \mathcal{W}(y_k)_+$

(b) Update X: $X_{k+1} = -\sigma\mathcal{W}(y_k)_-$

projection methods

boundary point method

Initialization: $k = 0$, select $\sigma_k > 0$, $X_k \succeq 0, Z_k \succeq 0$

repeat until convergence

(a) Keep X_k fixed.

repeat until convergence

– solve $\mathcal{A}(\mathcal{A}^\top(y)) = rhs$ giving y_k

– compute $Z_k = \mathcal{W}(y_k)_+$

(b) Update X: $X_{k+1} = -\sigma\mathcal{W}(y_k)_-$

(c) Select $\sigma_{k+1} \geq \sigma_k$.

projection methods

boundary point method

Initialization: $k = 0$, select $\sigma_k > 0$, $X_k \succeq 0, Z_k \succeq 0$

repeat until convergence

(a) Keep X_k fixed.

repeat until convergence

– solve $\mathcal{A}(\mathcal{A}^\top(y)) = rhs$ giving y_k

– compute $Z_k = \mathcal{W}(y_k)_+$

(b) Update X: $X_{k+1} = -\sigma \mathcal{W}(y_k)_-$

(c) Select $\sigma_{k+1} \geq \sigma_k$.

(d) Check the stopping condition and increase k

See Malick, Povh, Rendl, W., 2007

projection methods

observation: $X \succeq 0$, $Z \succeq 0$, $ZX = 0$ hold throughout the algorithm; once primal and dual feasibilty reached we are optimal.

projection methods

observation: $X \succeq 0$, $Z \succeq 0$, $ZX = 0$ hold throughout the algorithm; once primal and dual feasibilty reached we are optimal.

computational effort in each iteration:

- solve $\mathcal{A}(\mathcal{A}^\top(y)) = rhs$ giving updated $\tilde{y}$
- compute $\tilde{Z} = (\mathcal{A}^\top(y) - C - \frac{1}{\sigma}X)_+$

projection methods

observation: $X \succeq 0$, $Z \succeq 0$, $ZX = 0$ hold throughout the algorithm; once primal and dual feasibilty reached we are optimal.

computational effort in each iteration:

- solve $\mathcal{A}(\mathcal{A}^\top(y)) = rhs$ giving updated $\tilde{y}$
- compute $\tilde{Z} = (\mathcal{A}^\top(y) - C - \frac{1}{\sigma}X)_+$

example: when computing the ϑ-number: $\mathcal{A}(\mathcal{A}^\top(.))$ is diagonal. Therefore, the main computational effort is the projection on the positive semidefinite cone.

projection methods

example: computing the ϑ-number for random graphs from the Kim Toh collection

graph	n	$\|E\|$	time (secs)
theta82	400	23871	87
theta83	400	39861	70
theta102	500	37466	143
theta103	500	62515	110
theta104	500	87244	124
theta123	600	90019	205
theta162	800	127599	570

projection methods

boundary point method summarized

- works “orthogonal” to interior point methods
- convergence behavior not well understood
- for matrices of moderate size, but can deal with a large number of constraints

conclusions

Semidefinite Programming solvers at the NEOS site:
`http://www.neos-server.org/`

- csdp
- dsdp
- penbmi
- pensdp
- sdpa
- sdplr
- sdpt3
- sedumi

moreover: sdplib [B. Borchers] at
`http://euler.nmt.edu/~brian/sdplib/` and the sdp website
[Ch. Helmberg] at
`http://www-user.tu-chemnitz.de/~helmberg/semidef.html`

conclusions

- solution method to choose depends on sizes and on structure of problems

conclusions

- solution method to choose depends on sizes and on structure of problems
- interior point methods: many implementations available, limit $n \approx 1000$, $m \approx 10\,000$

conclusions

- solution method to choose depends on sizes and on structure of problems
- interior point methods: many implementations available, limit $n \approx 1000$, $m \approx 10\,000$
- spectral bundle method: general tool for matrices of large dimension

conclusions

- solution method to choose depends on sizes and on structure of problems
- interior point methods: many implementations available, limit $n \approx 1000$, $m \approx 10\,000$
- spectral bundle method: general tool for matrices of large dimension
- bundle method: if partial Lagrangian dual is “nice”

conclusions

- solution method to choose depends on sizes and on structure of problems
- interior point methods: many implementations available, limit $n \approx 1000$, $m \approx 10\,000$
- spectral bundle method: general tool for matrices of large dimension
- bundle method: if partial Lagrangian dual is “nice”
- more methods: augmented Lagrangian methods, projection methods, low-rank methods,...

conclusions

- solution method to choose depends on sizes and on structure of problems
- interior point methods: many implementations available, limit $n \approx 1000$, $m \approx 10\,000$
- spectral bundle method: general tool for matrices of large dimension
- bundle method: if partial Lagrangian dual is "nice"
- more methods: augmented Lagrangian methods, projection methods, low-rank methods,. . .
- SDP is standard tool in optimization and sufficiently easy to use(?)

conclusions

- solution method to choose depends on sizes and on structure of problems
- interior point methods: many implementations available, limit $n \approx 1000$, $m \approx 10\,000$
- spectral bundle method: general tool for matrices of large dimension
- bundle method: if partial Lagrangian dual is "nice"
- more methods: augmented Lagrangian methods, projection methods, low-rank methods,...
- SDP is standard tool in optimization and sufficiently easy to use(?)

thank you for your attention!