Large Scale Semidefinite Programming in ConicBundle

Christoph Helmberg
based on joint work with M. Overton and F. Rendl
TU Chemnitz

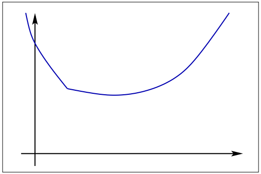
- The Bundle Method and the Aggregate
- SDP, Eigenvalue Optimisation, and the Spectral Bundle Method
- Second Order Approaches to Eigenvalue Optimisation
- Adaptation to the Spectral Bundle Method
- Numerical Examples
- Conclusions

Workshop "Semidefinite Programming: Theory and Applications"
Edinburgh, October 19, 2018

The Bundle Method for Nonsmooth Convex Optimization

$$\min \; f(y) \quad \text{s.t.} \quad y \in \mathbb{R}^m$$

with $f : \mathbb{R}^m \to \mathbb{R}$ convex (nonsmooth)



The Bundle Method for Nonsmooth Convex Optimization

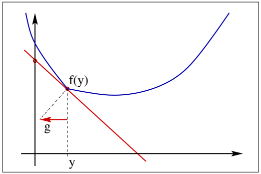
$$\min \ f(y) \quad \text{s.t.} \quad y \in \mathbb{R}^m$$

with $f : \mathbb{R}^m \to \mathbb{R}$ convex (nonsmooth)

f is specified by a **first order oracle**:

given $\bar{y} \in \mathbb{R}^m$ it returns

- $f(\bar{y}) \in \mathbb{R}$ function value
- $g(\bar{y}) \in \mathbb{R}^m$ some subgradient (not nec. unique)



satisfying $f(y) \geq f(\bar{y}) + \langle g(\bar{y}), y - \bar{y} \rangle \quad \forall y \in \mathbb{R}^m$ (subg. ineq.)

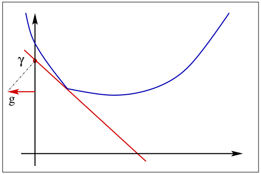
The Bundle Method for Nonsmooth Convex Optimization

$$\min \; f(y) \quad \text{s.t.} \quad y \in \mathbb{R}^m$$

with $f : \mathbb{R}^m \to \mathbb{R}$ convex (nonsmooth)

f is specified by a **first order oracle**:
given $\bar{y} \in \mathbb{R}^m$ it returns

- $f(\bar{y}) \in \mathbb{R}$ function value
- $g(\bar{y}) \in \mathbb{R}^m$ some subgradient (not nec. unique)



satisfying $f(y) \geq f(\bar{y}) + \langle g(\bar{y}), y - \bar{y} \rangle \quad \forall y \in \mathbb{R}^m$ (subg. ineq.)

Each $\omega = (\gamma, g)$, $\gamma = f(\bar{y}) - \langle g, \bar{y} \rangle$ generates a **linear minorant** of f

$$f_\omega(y) := \gamma + \langle g, y \rangle \quad \leq f(y) \qquad \forall y \in \mathbb{R}^m$$

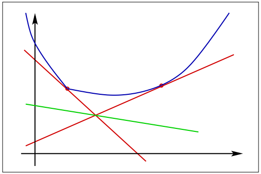
The Bundle Method for Nonsmooth Convex Optimization

$$\min \; f(y) \quad \text{s.t.} \quad y \in \mathbb{R}^m$$

with $f\colon \mathbb{R}^m \to \mathbb{R}$ convex (nonsmooth)

f is specified by a **first order oracle**:
given $\bar{y} \in \mathbb{R}^m$ it returns

- $f(\bar{y}) \in \mathbb{R}$ function value
- $g(\bar{y}) \in \mathbb{R}^m$ some subgradient (not nec. unique)



satisfying $\quad f(y) \geq f(\bar{y}) + \langle g(\bar{y}), y - \bar{y} \rangle \quad \forall y \in \mathbb{R}^m \quad$ (subg. ineq.)

Each $\omega = (\gamma, g)$, $\gamma = f(\bar{y}) - \langle g, \bar{y} \rangle$ generates a **linear minorant** of f

$$f_\omega(y) := \gamma + \langle g, y \rangle \quad \leq f(y) \qquad \forall y \in \mathbb{R}^m$$

The collected minorants form the **bundle**, from this we select a model

$$\widehat{\mathcal{W}} \subseteq \operatorname{conv}\{(\gamma, g)\colon g = g(\bar{y}^i), \gamma = f(\bar{y}^i) - \left\langle g, \bar{y}^i \right\rangle, i = 1, \ldots, k\},$$

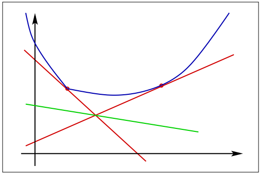
The Bundle Method for Nonsmooth Convex Optimization

$$\min \; f(y) \quad \text{s.t.} \quad y \in \mathbb{R}^m$$

with $f : \mathbb{R}^m \to \mathbb{R}$ convex (nonsmooth)

f is specified by a **first order oracle**:
given $\bar{y} \in \mathbb{R}^m$ it returns

- $f(\bar{y}) \in \mathbb{R}$ function value
- $g(\bar{y}) \in \mathbb{R}^m$ some subgradient (not nec. unique)



satisfying $f(y) \geq f(\bar{y}) + \langle g(\bar{y}), y - \bar{y} \rangle \quad \forall y \in \mathbb{R}^m$ (subg. ineq.)

Each $\omega = (\gamma, g)$, $\gamma = f(\bar{y}) - \langle g, \bar{y} \rangle$ generates a **linear minorant** of f

$$f_\omega(y) := \gamma + \langle g, y \rangle \quad \leq f(y) \qquad \forall y \in \mathbb{R}^m$$

The collected minorants form the **bundle**, from this we select a model

$$\widehat{\mathcal{W}} \subseteq \text{conv}\{(\gamma, g) \colon g = g(\bar{y}^i), \gamma = f(\bar{y}^i) - \left\langle g, \bar{y}^i \right\rangle, i = 1, \ldots, k\},$$

Any closed proper convex function is the sup over its linear minorants,

$$f(y) = \sup_{(\gamma,g) \in \mathcal{W}} \gamma + \langle g, y \rangle, \qquad \text{choose compact } \widehat{\mathcal{W}} \subseteq \mathcal{W}.$$

The Bundle Method for Nonsmooth Convex Optimization

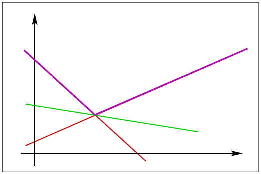
$$\min \ f(y) \quad \text{s.t.} \quad y \in \mathbb{R}^m$$

with $f\colon \mathbb{R}^m \to \mathbb{R}$ convex (nonsmooth)

f is specified by a **first order oracle**:

given $\bar{y} \in \mathbb{R}^m$ it returns

- $f(\bar{y}) \in \mathbb{R}$ function value
- $g(\bar{y}) \in \mathbb{R}^m$ some subgradient (not nec. unique)



satisfying $f(y) \geq f(\bar{y}) + \langle g(\bar{y}), y - \bar{y}\rangle \quad \forall y \in \mathbb{R}^m$ (subg. ineq.)

Each $\omega = (\gamma, g)$, $\gamma = f(\bar{y}) - \langle g, \bar{y}\rangle$ generates a **linear minorant** of f

$$f_\omega(y) := \gamma + \langle g, y\rangle \quad \leq f(y) \qquad \forall y \in \mathbb{R}^m$$

The collected minorants form the **bundle**, from this we select a model

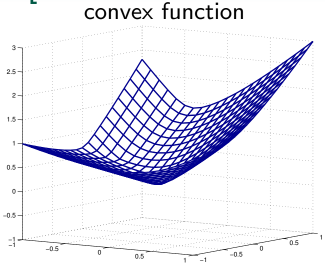
$$\widehat{\mathcal{W}} \subseteq \operatorname{conv}\{(\gamma, g)\colon g = g(\bar{y}^i), \gamma = f(\bar{y}^i) - \left\langle g, \bar{y}^i \right\rangle, i = 1, \ldots, k\},$$

Maximizing over all $\omega \in \widehat{\mathcal{W}}$ gives a **cutting model** minorizing f,

$$f_{\widehat{\mathcal{W}}}(y) := \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) \quad \leq f(y) \qquad \forall y \in \mathbb{R}^m$$

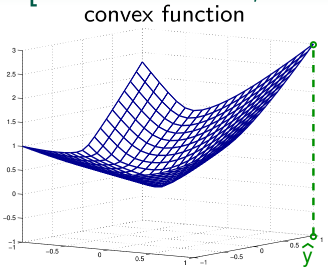
Proximal Bundle Method [Lemaréchal78,Kiwiel90]

Input: a convex function
given by a first order oracle



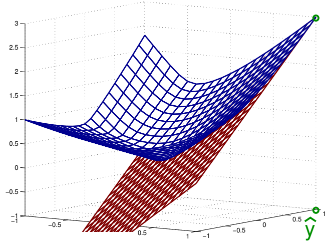
Proximal Bundle Method [Lemaréchal78,Kiwiel90]

Input: a convex function
given by a first order oracle



Proximal Bundle Method [Lemaréchal78,Kiwiel90]

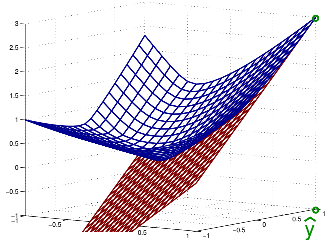
cutting plane model with $g \in \partial f(\hat{y})$



Input: a convex function
given by a first order oracle

Proximal Bundle Method [Lemaréchal78,Kiwiel90]

cutting plane model with $g \in \partial f(\hat{y})$



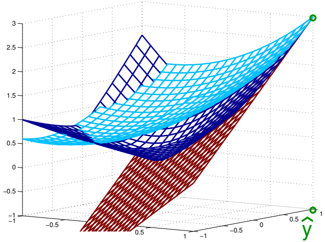
Input: a convex function
given by a first order oracle

1. Find a candidate by solving

$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y)$$

Proximal Bundle Method [Lemaréchal78,Kiwiel90]

solve augmented model $\rightarrow \bar{y}$



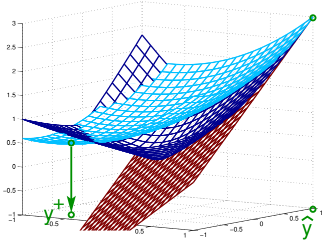
Input: a convex function
given by a first order oracle

1. Find a candidate by solving the quadratic model

$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

Proximal Bundle Method [Lemaréchal78,Kiwiel90]

solve augmented model $\rightarrow \bar{y}$



Input: a convex function
given by a first order oracle

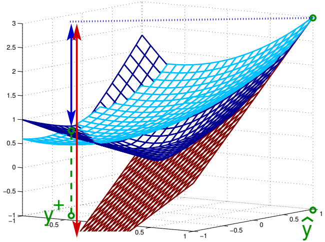
1. Find a candidate by solving the quadratic model

$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_{\omega}(y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

2. Evaluate the function and determine a subgradient (oracle)

Proximal Bundle Method [Lemaréchal78,Kiwiel90]

solve augmented model $\rightarrow \bar{y}$



Input: a convex function
given by a first order oracle

1. Find a candidate by solving the quadratic model

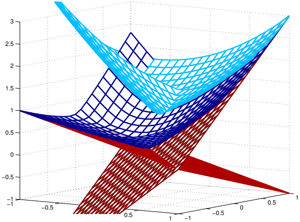
$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

2. Evaluate the function and determine a subgradient (oracle)
3. Decide on
    - null step
    - descent step

Proximal Bundle Method [Lemaréchal78,Kiwiel90]

improve cutting model in $\bar{y}$

Input: a convex function
given by a first order oracle



1. Find a candidate by solving the quadratic model

$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

2. Evaluate the function and determine a subgradient (oracle)
3. Decide on
   - null step
   - descent step
4. Update model to contain at least *aggregate* and new minorant and iterate

The Aggregate and Convergence

Given weight $\mu > 0$, the quadratic subproblem is a saddle point problem

$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

The Aggregate and Convergence

Given weight $\mu > 0$, the quadratic subproblem is a saddle point problem

$$\min_y \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2 = \min_y \max_{\substack{\xi_\omega \geq 0 \\ \sum \xi_\omega = 1}} \sum_{(\gamma, g) \in \widehat{\mathcal{W}}} \xi_\omega(\gamma + g^\top y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

The Aggregate and Convergence

Given weight $\mu > 0$, the quadratic subproblem is a saddle point problem

$$\min_y \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2 = \max_{\substack{\xi_\omega \geq 0 \\ \sum \xi_\omega = 1}} \min_y \sum_{(\gamma, g) \in \widehat{\mathcal{W}}} \xi_\omega(\gamma + g^\top y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

The Aggregate and Convergence

Given weight $\mu > 0$, the quadratic subproblem is a saddle point problem

$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2 = \max_{\substack{\xi_\omega \geq 0 \\ \sum \xi_\omega = 1}} \min_{y} \sum_{(\gamma, g) \in \widehat{\mathcal{W}}} \xi_\omega(\gamma + g^\top y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

Determining the saddle point $(\bar{y}, \bar{\omega})$ over $\mathbb{R}^n \times \operatorname{conv} \widehat{\mathcal{W}}$ yields

- $\bar{\omega} = (\bar{\gamma}, \bar{g})$, the **aggregate** (the "best" minorant in $\operatorname{conv} \widehat{\mathcal{W}}$),
- $\bar{y} = \hat{y} - \frac{1}{\mu}\bar{g}$, the next **candidate** for evaluation.

The Aggregate and Convergence

Given weight $\mu > 0$, the quadratic subproblem is a saddle point problem

$$\min_y \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2 = \max_{\substack{\xi_\omega \geq 0 \\ \sum \xi_\omega = 1}} \min_y \sum_{(\gamma,g)\in\widehat{\mathcal{W}}} \xi_\omega(\gamma + g^\top y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

Determining the saddle point $(\bar{y}, \bar{\omega})$ over $\mathbb{R}^n \times \operatorname{conv}\widehat{\mathcal{W}}$ yields

- $\bar{\omega} = (\bar{\gamma}, \bar{g})$, the **aggregate** (the "best" minorant in $\operatorname{conv}\widehat{\mathcal{W}}$),
- $\bar{y} = \hat{y} - \frac{1}{\mu}\bar{g}$, the next **candidate** for evaluation.

The progress $f(\hat{y}) - f(\bar{y})$ is compared to the **predicted decrease**

$$f(\hat{y}) - f_{\bar{\omega}}(\bar{y}) = f(\hat{y}) - \bar{\gamma} - \langle \hat{y}, \bar{g}\rangle + \frac{1}{\mu}\|\bar{g}\|^2 \geq 0,$$

The Aggregate and Convergence

Given weight $\mu > 0$, the quadratic subproblem is a saddle point problem

$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2 = \max_{\substack{\xi_\omega \geq 0 \\ \sum \xi_\omega = 1}} \min_{y} \sum_{(\gamma, g) \in \widehat{\mathcal{W}}} \xi_\omega(\gamma + g^\top y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

Determining the saddle point $(\bar{y}, \bar{\omega})$ over $\mathbb{R}^n \times \operatorname{conv} \widehat{\mathcal{W}}$ yields

- $\bar{\omega} = (\bar{\gamma}, \bar{g})$, the **aggregate** (the "best" minorant in $\operatorname{conv} \widehat{\mathcal{W}}$),
- $\bar{y} = \hat{y} - \frac{1}{\mu}\bar{g}$, the next **candidate** for evaluation.

The progress $f(\hat{y}) - f(\bar{y})$ is compared to the **predicted decrease**

$$f(\hat{y}) - f_{\bar{\omega}}(\bar{y}) = f(\hat{y}) - \bar{\gamma} - \langle \hat{y}, \bar{g} \rangle + \frac{1}{\mu}\|\bar{g}\|^2 \geq 0,$$

This decides on **descent step** ($\hat{y} \leftarrow \bar{y}$) or **null step** ($\hat{y} \leftarrow \hat{y}$, new ω).

The Aggregate and Convergence

Given weight $\mu > 0$, the quadratic subproblem is a saddle point problem

$$\min_y \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2 = \max_{\substack{\xi_\omega \geq 0 \\ \sum \xi_\omega = 1}} \min_y \sum_{(\gamma, g) \in \widehat{\mathcal{W}}} \xi_\omega(\gamma + g^\top y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

Determining the saddle point $(\bar{y}, \bar{\omega})$ over $\mathbb{R}^n \times \operatorname{conv} \widehat{\mathcal{W}}$ yields

- $\bar{\omega} = (\bar{\gamma}, \bar{g})$, the **aggregate** (the "best" minorant in $\operatorname{conv} \widehat{\mathcal{W}}$),
- $\bar{y} = \hat{y} - \frac{1}{\mu}\bar{g}$, the next **candidate** for evaluation.

The progress $f(\hat{y}) - f(\bar{y})$ is compared to the **predicted decrease**

$$f(\hat{y}) - f_{\bar{\omega}}(\bar{y}) = f(\hat{y}) - \bar{\gamma} - \langle \hat{y}, \bar{g} \rangle + \frac{1}{\mu}\|\bar{g}\|^2 \geq 0,$$

This decides on **descent step** $(\hat{y} \leftarrow \bar{y})$ or **null step** $(\hat{y} \leftarrow \hat{y}$, new $\omega)$.

Theorem (e.g. [BoGiLeSa2003])

Let $\hat{y}^k$ denote the center of iteration k, then $f(\hat{y}^k) \to \inf f$.
If, in addition, $\hat{y}^{k_0} = \hat{y}^k$ for $k \geq k_0$ (finitely many descent steps) then $\hat{y}^{k_0}$ minimizes f and $(f(\hat{y}^k) - f_{\bar{\omega}^k}(\bar{y}^k))_{k > k_0} \downarrow 0$.

The Aggregate and Convergence

Given weight $\mu > 0$, the quadratic subproblem is a saddle point problem

$$\min_{y} \max_{\omega \in \widehat{\mathcal{W}}} f_\omega(y) + \frac{\mu}{2}\|y - \hat{y}\|^2 = \max_{\substack{\xi_\omega \geq 0 \\ \sum \xi_\omega = 1}} \min_{y} \sum_{(\gamma, g) \in \widehat{\mathcal{W}}} \xi_\omega(\gamma + g^\top y) + \frac{\mu}{2}\|y - \hat{y}\|^2$$

Determining the saddle point $(\bar{y}, \bar{\omega})$ over $\mathbb{R}^n \times \operatorname{conv} \widehat{\mathcal{W}}$ yields

- $\bar{\omega} = (\bar{\gamma}, \bar{g})$, the **aggregate** (the "best" minorant in $\operatorname{conv} \widehat{\mathcal{W}}$),
- $\bar{y} = \hat{y} - \frac{1}{\mu}\bar{g}$, the next **candidate** for evaluation.

The progress $f(\hat{y}) - f(\bar{y})$ is compared to the **predicted decrease**

$$f(\hat{y}) - f_{\bar{\omega}}(\bar{y}) = f(\hat{y}) - \bar{\gamma} - \langle \hat{y}, \bar{g} \rangle + \frac{1}{\mu}\|\bar{g}\|^2 \geq 0,$$

This decides on **descent step** $(\hat{y} \leftarrow \bar{y})$ or **null step** $(\hat{y} \leftarrow \hat{y}$, new $\omega)$.

Theorem (e.g. [BoGiLeSa2003])

Let $\hat{y}^k$ denote the center of iteration k, then $f(\hat{y}^k) \to \inf f$.
If, in addition, $\hat{y}^{k_0} = \hat{y}^k$ for $k \geq k_0$ (finitely many descent steps) then $\hat{y}^{k_0}$ minimizes f and $(f(\hat{y}^k) - f_{\bar{\omega}^k}(\bar{y}^k))_{k > k_0} \downarrow 0$.

f bounded below $\Rightarrow \|\bar{g}^k\| \xrightarrow{K} 0$

The bundle framework offers a lot of flexibility and can be extended in many directions:

- add scaling/"second order" information via the proximal term
- allow constraints on y
- Lagrangian relaxation/decomposition or sums of convex functions
- generate good primal approximations in Lagrangian relaxation
- solve the dual to primal cutting plane approaches
- use specialized cutting models (quadratic subproblem solvable?)
- asynchronous parallel approaches

For me it offers the potential for

"A general tool like the simplex method for LP"

$\rightarrow$ ConicBundle, contains much but not yet all of this ...

Here: choose model and proximal term $+\frac{1}{2}\|y-\hat{y}\|_H^2$
for the maximum eigenvalue function/semidefinite prog.

LP $\leftrightarrow$ SDP

LP	SDP
$\max \; \langle c, x\rangle$ s.t. $Ax = b$, $x \geq 0$	$\max \; \langle C, X\rangle$ s.t. $\mathcal{A}X = b$, $X \succeq 0$
$x \in \mathbb{R}^n_+$ nonneg. orthant (polyhedral)	$X \in S^n_+$ pos. semidef. matrices (non-polyhedral)
$\langle c, x\rangle = \sum_i c_i x_i$	$\langle C, X\rangle = \sum_{i,j} C_{ij} X_{ij}$
$Ax = \begin{pmatrix} \langle a_1, x\rangle \\ \vdots \\ \langle a_m, x\rangle \end{pmatrix}$	$\mathcal{A}X = \begin{pmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{pmatrix}$
$A^T y = \sum_i a_i y_i$	$\mathcal{A}^T y = \sum_i A_i y_i$
$\min \; \langle b, y\rangle$ s.t. $A^T y - z = c$, $z \geq 0$	$\min \; \langle b, y\rangle$ s.t. $\mathcal{A}^T y - Z = C$, $Z \succeq 0$

Example

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \langle I, X \rangle = 1 \\ & X \succeq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & y \\ \text{s.t.} \quad & Z = yI - C \succeq 0 \end{aligned}$$

Example

$$\begin{array}{rl} \max & \langle C, X \rangle \\ \text{s.t.} & \langle I, X \rangle = 1 \\ & X \succeq 0 \end{array} \qquad \begin{array}{rl} \min & y \\ \text{s.t.} & Z = yI - C \succeq 0 \end{array} \qquad [\rightarrow y_* = \lambda_{\max}(C)]$$

Example

$$\begin{array}{rl} \max & \langle C, X \rangle \\ \text{s.t.} & \langle I, X \rangle = 1 \\ & X \succeq 0 \end{array} \qquad \begin{array}{rl} \min & y \\ \text{s.t.} & Z = yI - C \succeq 0 \end{array} \qquad [\to y_* = \lambda_{\max}(C)]$$

$$\mathcal{W} := \{X \succeq 0 : \langle I, X \rangle = 1\} \;=\; \operatorname{conv}\left\{vv^T : \left\langle I, vv^T \right\rangle = v^T v = 1\right\}$$

$$\max_{X \in \mathcal{W}} \langle C, X \rangle \;=\; \max_{\|v\|^2 = 1} \left\langle C, vv^T \right\rangle \;=\; \max_{\|v\|=1} \; v^T C v \;=\; \lambda_{\max}(C)$$

Example

$$\begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \langle I, X\rangle = 1 \\ & X \succeq 0 \end{array} \qquad \begin{array}{rl} \min & y \\ \text{s.t.} & Z = yI - C \succeq 0 \end{array} \qquad [\to y_* = \lambda_{\max}(C)]$$

$$\mathcal{W} := \{X \succeq 0 : \langle I, X\rangle = 1\} \;=\; \operatorname{conv}\left\{vv^T : \left\langle I, vv^T\right\rangle = v^T v = 1\right\}$$

$$\max_{X \in \mathcal{W}} \langle C, X\rangle \;=\; \max_{\|v\|^2=1} \left\langle C, vv^T\right\rangle \;=\; \max_{\|v\|=1} v^T C v \;=\; \lambda_{\max}(C)$$

---

set of primal optimal solutions:

$$\begin{aligned} & \operatorname{conv}\left\{vv^T : \left\langle I, vv^T\right\rangle = 1, v^T C v = \lambda_{\max}(C)\right\} \qquad [v = Pu] \\ = \; & \operatorname{conv}\left\{Puu^T P^T : \left\langle I, uu^T\right\rangle = 1\right\} \\ = \; & \left\{PUP^T : \langle I, U\rangle = 1, U \succeq 0\right\} \end{aligned}$$

columns of P form an orthonormal basis of the eigenspace of $\lambda_{\max}(C)$.

Spectral Bundle Method [H.,Rendl00]

For constant trace, the dual is an eigenvalue optimization problem

$$\begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \langle I, X\rangle = a \\ & \mathcal{A}X = b \\ & X \succeq 0, \end{array} \qquad \min_{y \in \mathbb{R}^m} \; a\lambda_{\max}(C - \mathcal{A}^T y) + \langle b, y\rangle$$

For bounded trace, the dual is

$$\begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \langle I, X\rangle \leq a \\ & \mathcal{A}X = b \\ & X \succeq 0, \end{array} \qquad \min_{y \in \mathbb{R}^m} \; \max\{0, a\lambda_{\max}(C - \mathcal{A}^T y)\} + \langle b, y\rangle$$

In the following we consider constant trace with $a = 1$,
and solve the eigenvalue problem by a specialized bundle approach.

The matrix $C - \sum_i A_i y_i$ inherits the structure of cost matrix and constraints [$\rightarrow$ $\lambda_{\max}$ by iterative methods like Lanczos]

A semidefinite model for $f(y) := \lambda_{\max}(C - \mathcal{A}^T y) + b^T y$

With $\mathcal{W} = \{W \succeq 0 : \operatorname{tr} W = 1\}$

$$f(y) = \max_{W \in \mathcal{W}} \langle W, C - \mathcal{A}^T y \rangle + b^T y$$

evaluate by computing $\lambda_{\max}(C - \mathcal{A}^T y)$, [Lanczos]
any eigenvector v to $\lambda_{\max}$, $\|v\| = 1$, yields a subgradient via $vv^T \in \mathcal{W}$

A semidefinite model for $f(y) := \lambda_{\max}(C - \mathcal{A}^T y) + b^T y$

With $\mathcal{W} = \{W \succeq 0 : \operatorname{tr} W = 1\}$

$$f(y) = \max_{W \in \mathcal{W}} \langle W, C - \mathcal{A}^T y \rangle + b^T y$$

evaluate by computing $\lambda_{\max}(C - \mathcal{A}^T y)$, [Lanczos]
any eigenvector v to $\lambda_{\max}$, $\|v\| = 1$, yields a subgradient via $vv^T \in \mathcal{W}$

For any subset $\widehat{\mathcal{W}}_k \subseteq \mathcal{W}$ one obtains a cutting model

$$\boxed{f_{\widehat{\mathcal{W}}_k}(y) = \max_{W \in \widehat{\mathcal{W}}_k} \langle W, C - \mathcal{A}^T y \rangle + b^T y} \qquad \le f(y) \quad \forall y \in \mathbb{R}^m$$

A semidefinite model for $f(y) := \lambda_{\max}(C - \mathcal{A}^T y) + b^T y$

With $\mathcal{W} = \{W \succeq 0 : \operatorname{tr} W = 1\}$

$$f(y) = \max_{W \in \mathcal{W}} \langle W, C - \mathcal{A}^T y \rangle + b^T y$$

evaluate by computing $\lambda_{\max}(C - \mathcal{A}^T y)$, [Lanczos]
any eigenvector v to $\lambda_{\max}$, $\|v\| = 1$, yields a subgradient via $vv^T \in \mathcal{W}$

For any subset $\widehat{\mathcal{W}}_k \subseteq \mathcal{W}$ one obtains a cutting model

$$\boxed{f_{\widehat{\mathcal{W}}_k}(y) = \max_{W \in \widehat{\mathcal{W}}_k} \langle W, C - \mathcal{A}^T y \rangle + b^T y} \qquad \leq f(y) \quad \forall y \in \mathbb{R}^m$$

We use

$$\boxed{\widehat{\mathcal{W}}_k = \left\{ P_k U P_k^T + \alpha \overline{X}_k : \operatorname{tr} U + \alpha = 1, U \succeq 0, \alpha \geq 0 \right\}} \qquad \subseteq \mathcal{W}$$

with parameters $P_k \in \mathbb{R}^{n \times r}$, $P_k^T P_k = I_r$, and an "aggregate" $\overline{X}_k \in \mathcal{W}$.

A semidefinite model for $f(y) := \lambda_{\max}(C - \mathcal{A}^T y) + b^T y$

With $\mathcal{W} = \{W \succeq 0 : \operatorname{tr} W = 1\}$

$$f(y) = \max_{W \in \mathcal{W}} \langle W, C - \mathcal{A}^T y \rangle + b^T y$$

evaluate by computing $\lambda_{\max}(C - \mathcal{A}^T y)$, [Lanczos]
any eigenvector v to $\lambda_{\max}$, $\|v\| = 1$, yields a subgradient via $vv^T \in \mathcal{W}$

For any subset $\widehat{\mathcal{W}}_k \subseteq \mathcal{W}$ one obtains a cutting model

$$\boxed{f_{\widehat{\mathcal{W}}_k}(y) = \max_{W \in \widehat{\mathcal{W}}_k} \langle W, C - \mathcal{A}^T y \rangle + b^T y} \quad \le f(y) \quad \forall y \in \mathbb{R}^m$$

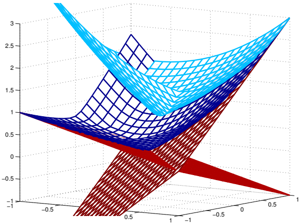
We use

$$\boxed{\widehat{\mathcal{W}}_k = \left\{ P_k U P_k^T + \alpha \overline{X}_k : \operatorname{tr} U + \alpha = 1, U \succeq 0, \alpha \ge 0 \right\}} \quad \subseteq \mathcal{W}$$

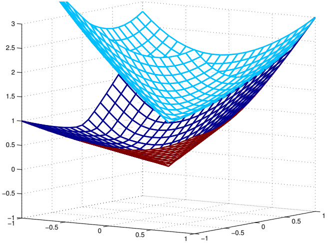
with parameters $P_k \in \mathbb{R}^{n \times r}$, $P_k^T P_k = I_r$, and an "aggregate" $\overline{X}_k \in \mathcal{W}$.
Convergence: $P = v$ and $\overline{X}$ or no $\overline{X}$ and big r with $\binom{r+1}{2} \le m$.

Spectral Bundle Model

cutting plane and augmented model



semidefinite model



Solving the augmented model $\min f_{\widehat{\mathcal{W}}}(y) + \frac{\mu}{2}\|y - \hat{y}\|^2$

$$\min_{y} \max_{W \in \widehat{\mathcal{W}}} \left\langle C - \mathcal{A}^T y, W \right\rangle + \langle b, y \rangle + \frac{\mu}{2}\|y - \hat{y}\|^2$$

$$= \max_{W \in \widehat{\mathcal{W}}} \min_{y} \langle C, W \rangle + \langle b - \mathcal{A}W, y \rangle + \frac{\mu}{2}\|y - \hat{y}\|^2$$

Solving the augmented model $\min f_{\widehat{\mathcal{W}}}(y) + \frac{\mu}{2}\|y - \hat{y}\|^2$

$$\min_{y} \max_{W \in \widehat{\mathcal{W}}} \left\langle C - \mathcal{A}^T y, W \right\rangle + \langle b, y \rangle + \frac{\mu}{2}\|y - \hat{y}\|^2$$

$$= \max_{W \in \widehat{\mathcal{W}}} \min_{y} \langle C, W \rangle + \langle b - \mathcal{A}W, y \rangle + \frac{\mu}{2}\|y - \hat{y}\|^2$$

Solve unconstrained quadratic inner optimization over y explicitly:

$$\boxed{y_+(W) = \hat{y} - \frac{1}{\mu}(b - \mathcal{A}W)}$$

[μ "step size/trust region control"]

Solving the augmented model $\min f_{\widehat{\mathcal{W}}}(y) + \frac{\mu}{2}\|y-\hat{y}\|^2$

$$\min_{y} \max_{W\in\widehat{\mathcal{W}}} \left\langle C-\mathcal{A}^T y, W\right\rangle + \langle b,y\rangle + \frac{\mu}{2}\|y-\hat{y}\|^2$$

$$= \max_{W\in\widehat{\mathcal{W}}} \min_{y} \langle C,W\rangle + \langle b-\mathcal{A}W, y\rangle + \frac{\mu}{2}\|y-\hat{y}\|^2$$

Solve unconstrained quadratic inner optimization over y explicitly:

$$\boxed{y_+(W) = \hat{y} - \tfrac{1}{\mu}(b-\mathcal{A}W)} \qquad [\mu \text{ “step size/trust region control”}]$$

Substitute for y to obtain a quadratic semidefinite problem in W,

$$(\text{QSP}) \quad \begin{array}{ll} \min & \frac{1}{2\mu}\|b-\mathcal{A}W\|^2 - \left\langle W, C-\mathcal{A}^T\hat{y}\right\rangle - \langle b,\hat{y}\rangle \\ \text{s.t.} & W = PUP^T + \alpha\overline{X} \\ & \operatorname{tr} U + \alpha = 1 \\ & U \succeq 0, \alpha \geq 0. \end{array}$$

small if r is small ($U \in S^r_+$) $\rightarrow$ interior point system matrix $\binom{r+1}{2}+1$ [!]

$\rightarrow$ “best (eps)subgradient” $W_+ = PU_+P^T + \alpha_+\overline{X}$

$\rightarrow$ new candidate $y_+ = y_+(W_+)$.

Second Order Approaches

[Overton8*, OvertonWomersley95, Oustry200*]

Local quadratic convergence for correct multiplicity t in the optimum y^*,

$$C - \mathcal{A}^T y^* = [Q_1^* Q_2^*] \begin{bmatrix} \Lambda_1^* & 0 \\ 0 & \Lambda_2^* \end{bmatrix} [Q_1^* Q_2^*]^T$$

$$\lambda_1^* = \cdots = \lambda_t^* > \lambda_{t+1}^* > \cdots > \lambda_n^*$$

Second Order Approaches

[Overton8*, OvertonWomersley95, Oustry200*]
Local quadratic convergence for correct multiplicity t in the optimum y^*,

$$C - \mathcal{A}^T y^* = [Q_1^* Q_2^*] \begin{bmatrix} \Lambda_1^* & 0 \\ 0 & \Lambda_2^* \end{bmatrix} [Q_1^* Q_2^*]^T$$

$$\lambda_1^* = \cdots = \lambda_t^* > \lambda_{t+1}^* > \cdots > \lambda_n^*$$

1. Guess t_k, compute Q_1^k, Q_2^k and an interior subgradient U_k by

$$\min \|b - \mathcal{A} Q_1 U Q_1^T\|^2 \text{ s.t. } \operatorname{tr} U = 1, \; U \succeq 0$$

Second Order Approaches

[Overton8*, OvertonWomersley95, Oustry200*]

Local quadratic convergence for correct multiplicity t in the optimum y^*,

$$C - \mathcal{A}^T y^* = [Q_1^* Q_2^*] \begin{bmatrix} \Lambda_1^* & 0 \\ 0 & \Lambda_2^* \end{bmatrix} [Q_1^* Q_2^*]^T$$

$$\lambda_1^* = \cdots = \lambda_t^* > \lambda_{t+1}^* > \cdots > \lambda_n^*$$

1. Guess t_k, compute Q_1^k, Q_2^k and an interior subgradient U_k by

$$\min \|b - \mathcal{A} Q_1 U Q_1^T\|^2 \text{ s.t. } \operatorname{tr} U = 1,\ U \succeq 0$$

2. Compute the Newton candidate by solving

$$\begin{array}{ll} \min & \frac{1}{2}\|y - \hat{y}_k\|_{H_k}^2 + \langle b, y\rangle + \delta \\ \text{s.t.} & \delta I = Q_1^T (C - \mathcal{A}^T y) Q_1 \end{array}$$

where

$$H_k = 2\mathcal{A}\left((Q_1 U_k Q_1^T) \otimes (Q_2[\lambda_1^k I - \Lambda_2^k]^{-1} Q_2^T)\right)\mathcal{A}^T \qquad \text{[regularity } \succ 0]$$

Adaptation of Step 2 for Spectral Bundle

Step 2 $\begin{array}{ll} \min & \frac{1}{2}\|y-\hat{y}\|_H^2 + \langle b, y\rangle + \delta \\ \text{s.t.} & \delta I = Q_1^T(C - \mathcal{A}^T y)Q_1 \end{array}$ is relaxed to

$$\begin{array}{ll} \min & \frac{1}{2}\|y-\hat{y}\|_H^2 + \langle b, y\rangle + \delta \\ \text{s.t.} & \delta I \succeq Q_1^T(C - \mathcal{A}^T y)Q_1, \end{array} \Rightarrow \delta = \lambda_{\max}(Q_1^T(C - \mathcal{A}^T y)Q_1).$$

Adaptation of Step 2 for Spectral Bundle

Step 2 $\quad \begin{array}{ll} \min & \frac{1}{2}\|y-\hat{y}\|_H^2 + \langle b, y\rangle + \delta \\ \text{s.t.} & \delta I = Q_1^T(C - \mathcal{A}^T y)Q_1 \end{array} \quad$ is relaxed to

$$\begin{array}{ll} \min & \frac{1}{2}\|y-\hat{y}\|_H^2 + \langle b, y\rangle + \delta \\ \text{s.t.} & \delta I \succeq Q_1^T(C - \mathcal{A}^T y)Q_1, \end{array} \quad \Rightarrow \quad \delta = \lambda_{\max}(Q_1^T(C - \mathcal{A}^T y)Q_1).$$

With $\widehat{\mathcal{W}} := \{Q_1 U Q_1^T : \operatorname{tr} U = 1, U \succeq 0\}$ the problem reads

$$\min_y \max_{W \in \widehat{\mathcal{W}}} \left\langle W, C - \mathcal{A}^T y \right\rangle + b^T y + \frac{1}{2}\|y-\hat{y}\|_H^2$$

Adaptation of Step 2 for Spectral Bundle

Step 2 $\quad \begin{array}{ll} \min & \frac{1}{2}\|y-\hat{y}\|_H^2 + \langle b, y\rangle + \delta \\ \text{s.t.} & \delta I = Q_1^T(C - \mathcal{A}^T y)Q_1 \end{array}$ is relaxed to

$$\begin{array}{ll} \min & \frac{1}{2}\|y-\hat{y}\|_H^2 + \langle b, y\rangle + \delta \\ \text{s.t.} & \delta I \succeq Q_1^T(C - \mathcal{A}^T y)Q_1, \end{array} \quad \Rightarrow \quad \delta = \lambda_{\max}(Q_1^T(C - \mathcal{A}^T y)Q_1).$$

With $\widehat{\mathcal{W}} := \{Q_1 U Q_1^T : \operatorname{tr} U = 1, U \succeq 0\}$ the problem reads

$$\min_y \max_{W \in \widehat{\mathcal{W}}} \left\langle W, C - \mathcal{A}^T y \right\rangle + b^T y + \frac{1}{2}\|y - \hat{y}\|_H^2$$

Dualize, then $\boxed{y_+(W) = \hat{y} - H^{-1}(b - \mathcal{A}W)}$

$$\text{(QSP)} \quad \begin{array}{ll} \min & \frac{1}{2}\|b - \mathcal{A}W\|_{H^{-1}}^2 - \left\langle W, C - \mathcal{A}^T\hat{y}\right\rangle - \langle b, \hat{y}\rangle \\ \text{s.t.} & W = Q_1 U Q_1^T \\ & \operatorname{tr} U = 1 \\ & U \succeq 0. \end{array}$$

Scope of a second order bundle method

If QSP is solved by an interior point method with r columns, each iteration of QSP requires the factorization of a $\binom{r+1}{2}$ matrix.

For m constraints we can expect $r \approx \sqrt{m}$.
$\rightarrow$ Several $O(m^3)$ operations for each solution of QSP.

Scope of a second order bundle method

If QSP is solved by an interior point method with r columns, each iteration of QSP requires the factorization of a $\binom{r+1}{2}$ matrix.

For m constraints we can expect $r \approx \sqrt{m}$.
$\rightarrow$ Several $O(m^3)$ operations for each solution of QSP.

Typically, a full interior point code requires several $O(n^3)$ and one $O(m^3)$ operation per iteration.

$\rightarrow$ Second order SB is unlikely to be attractive for $m \geq n$,
but might be relevant for small $m \leq n$ or if r is small.

$\rightarrow$ Emphasis on large n and rather small m.

Scaling Variants

- **No scaling, bounded bundle (SB)**

Scaling Variants

- **No scaling, bounded bundle (SB)**
- **No scaling, fat bundle (CB)**

Scaling Variants

- **No scaling, bounded bundle (SB)**
- **No scaling, fat bundle (CB)**
- **Modified Newton (CB-fN):** full eigenvalue decomposition, minimum norm subgradient, compute full Newton H $(+\rho I)$

Scaling Variants

- **No scaling, bounded bundle (SB)**
- **No scaling, fat bundle (CB)**
- **Modified Newton (CB-fN):** full eigenvalue decomposition, minimum norm subgradient, compute full Newton H $(+\rho I)$
- **Low-Rank Newton (CB-lrN):** collect approximate subspace to large eigenvalues, compute min. norm subgradient for this, low rank approximation of Newton matrix $(+\rho I)$

Scaling Variants

- **No scaling, bounded bundle (SB)**
- **No scaling, fat bundle (CB)**
- **Modified Newton (CB-fN):** full eigenvalue decomposition, minimum norm subgradient, compute full Newton H $(+\rho I)$
- **Low-Rank Newton (CB-lrN):** collect approximate subspace to large eigenvalues, compute min. norm subgradient for this, low rank approximation of Newton matrix $(+\rho I)$
- **Approximate Low-Rank Newton (CB-alrN):** collect approx. subspace to large eigenvalues, use subgradient induced by W_+ of (QSP), approximate Newton matrix with available low rank information $(+\rho I)$

Scaling Variants

- **No scaling, bounded bundle (SB)**
- **No scaling, fat bundle (CB)**
- **Modified Newton (CB-fN):** full eigenvalue decomposition, minimum norm subgradient, compute full Newton H $(+\rho I)$
- **Low-Rank Newton (CB-lrN):** collect approximate subspace to large eigenvalues, compute min. norm subgradient for this, low rank approximation of Newton matrix $(+\rho I)$
- **Approximate Low-Rank Newton (CB-alrN):** collect approx. subspace to large eigenvalues, use subgradient induced by W_+ of (QSP), approximate Newton matrix with available low rank information $(+\rho I)$
- **Diagonal Low-Rank (CB-diag):** Collect approximate subspace to large eigenvalues, use subgradient W_+ of (QSP) and the diagonal of the approximate Newton matrix $(+\rho I)$

Low Rank Structure

$$H = 2\mathcal{A}\left((Q_1 U Q_1^T) \otimes (Q_2[\lambda_1 I - \Lambda_2]^{-1} Q_2^T)\right)\mathcal{A}^T$$

decompose $U = Q_u \Lambda_u Q_u^T$, set $\bar{Q}_1 = Q_1 Q_u$ and rewrite H as

$$H = 2\mathcal{A}\left((\bar{Q}_1 \otimes Q_2)(\Lambda_u \otimes [\lambda_1 I - \Lambda_2]^{-1})(\bar{Q}_1 \otimes Q_2^T)\right)\mathcal{A}^T$$

Truncate $[\lambda_1 I - \Lambda_2]_{1,\ldots,h}$ and $Q_2 \to Q_h$,

Low Rank Structure

$$H = 2\mathcal{A}\left((Q_1 U Q_1^T) \otimes (Q_2[\lambda_1 I - \Lambda_2]^{-1} Q_2^T)\right)\mathcal{A}^T$$

decompose $U = Q_u \Lambda_u Q_u^T$, set $\bar{Q}_1 = Q_1 Q_u$ and rewrite H as

$$H = 2\mathcal{A}\left((\bar{Q}_1 \otimes Q_2)(\Lambda_u \otimes [\lambda_1 I - \Lambda_2]^{-1})(\bar{Q}_1 \otimes Q_2^T)\right)\mathcal{A}^T$$

Truncate $[\lambda_1 I - \Lambda_2]_{1,\dots,h}$ and $Q_2 \to Q_h$,
compute a QR-decomposition of $\mathcal{A}(\bar{Q}_1 \otimes Q_h) \ \to \ Q_\mathcal{A} R$

$$H_h = 2Q_\mathcal{A} \underbrace{R(\Lambda_u \otimes [\lambda_1 I - \Lambda_2]^{-1}_{1,\dots,h})R^T}_{\to \ \tilde{Q}\Lambda_H \tilde{Q}^T, \ Q_H := Q_\mathcal{A}\tilde{Q}} Q_\mathcal{A}^T$$

truncate $\Lambda_H \to \hat{\Lambda}_H, \hat{Q}_H$

$$\to \quad \hat{H} = \rho I + 2\hat{Q}_H \hat{\Lambda}_H \hat{Q}_H^T$$

for some regularization parameter $\rho > 0$.

Implementation Details

Multiplicity Detection.

Use Tapia indicators based on the development of the eigenvalues of the last two iterates of the (QSP) solver.

Implementation Details

Multiplicity Detection.

Use Tapia indicators based on the development of the eigenvalues of the last two iterates of the (QSP) solver.

Bundle Update.

- maintain approximate subspace $\overline{Q}$ to large eigenvalues
- old P: keep the active subspace of (QSP) and that having a large contribution to $\operatorname{diag}(H)$
- add the (5) top most Ritz vectors of $\overline{Q}$

Implementation Details

Multiplicity Detection.

Use Tapia indicators based on the development of the eigenvalues of the last two iterates of the (QSP) solver.

Bundle Update.

- maintain approximate subspace $\overline{Q}$ to large eigenvalues
- old P: keep the active subspace of (QSP) and that having a large contribution to diag(H)
- add the (5) top most Ritz vectors of $\overline{Q}$

Update of Q_2 for the Low Rank Representation?

Heuristic: dynamically enlarge $\overline{Q}$ in case of too many null steps

Numerical Experiments

Sparse SDP Random Generator: A_i nonzero submatrices of order p
small instances:

$n \in \{100, 300, 500\}$, $m \in \{100, 500, 1000\}$, $p \in \{3, 5, 7\}$

larger instances:

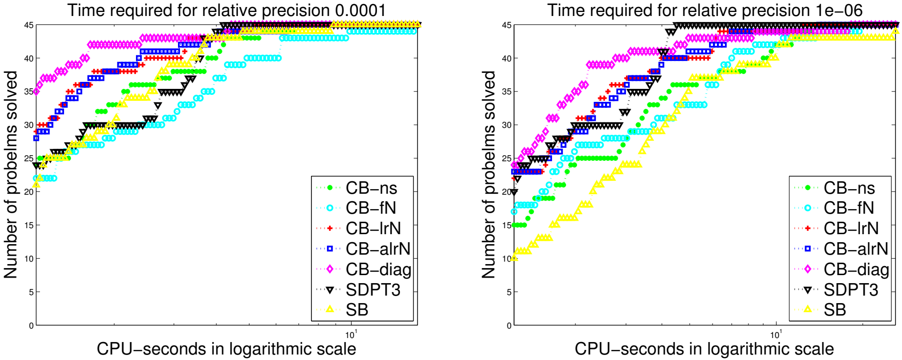
$n \in \{1, \ldots, 6\} \cdot 1000$, $m \in \{1, 3, 5\} \cdot 1000$, $p \in \{3, 4, 5\}$

Intel(R) Core(TM) i7 CPU 920 machines
8 MB cache, 12 GB RAM, openSUSE Linux 11.1 (x86_64)
in single processor mode

ConicBundle: start scaling at 10^{-2}
Termination: 10^{-8} or 10000 evaluations

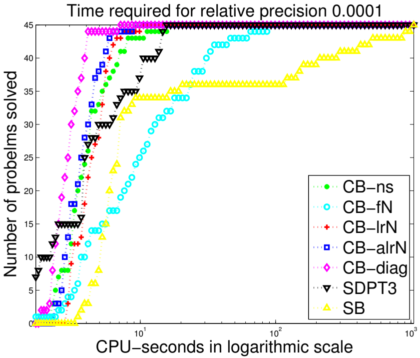
compare to SDPT3 4.0 beta [ToddTohTütüncü]

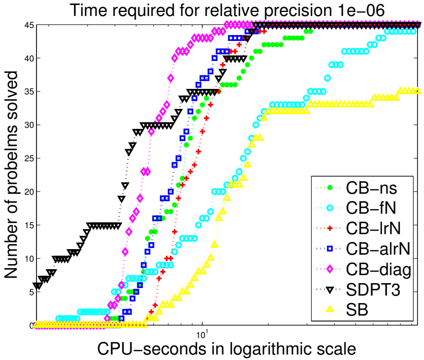
Small Instances: $n \in \{100, 300, 500\}$ and $m = 100$



Five instances per choice of n and constraint support order $\in \{3, 5, 7\}$

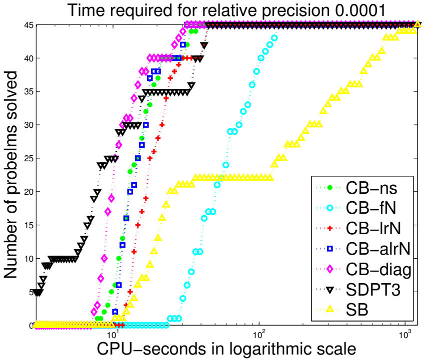
Small Instances: $n \in \{100, 300, 500\}$ and $m = 500$

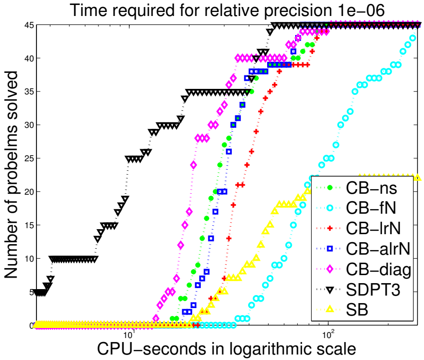




Five instances per choice of n and constraint support order $\in \{3, 5, 7\}$

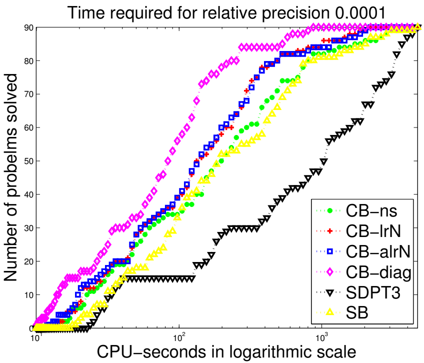
Small Instances: $n \in \{100, 300, 500\}$ and $m = 1000$

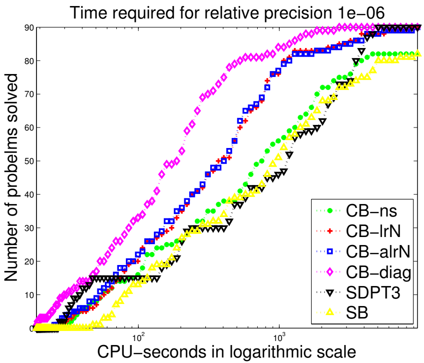




Five instances per choice of n and constraint support order $\in \{3, 5, 7\}$

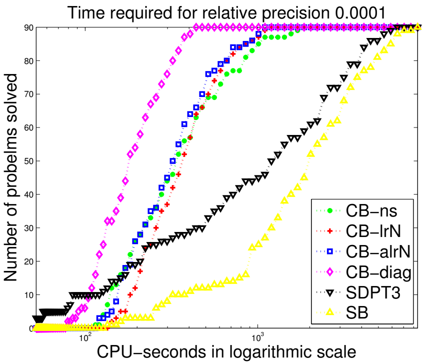
Larger Instances: $n \in \{1, \ldots, 6\} \cdot 1000$ and $m = 1000$

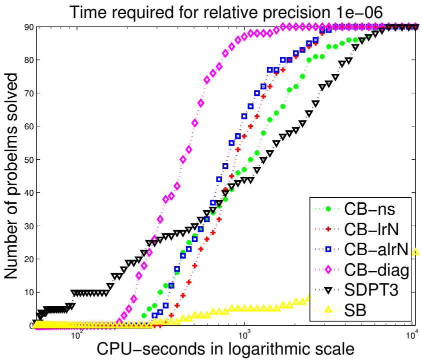




Five instances per choice of n and constraint support order $\in \{3, 4, 5\}$

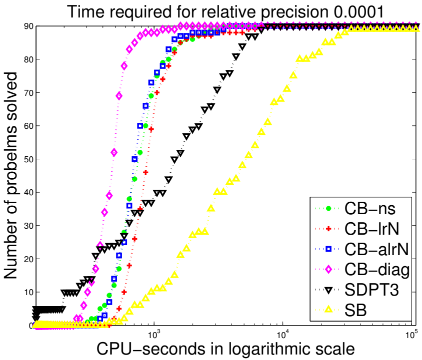
Larger Instances: $n \in \{1, \dots, 6\} \cdot 1000$ and $m = 3000$

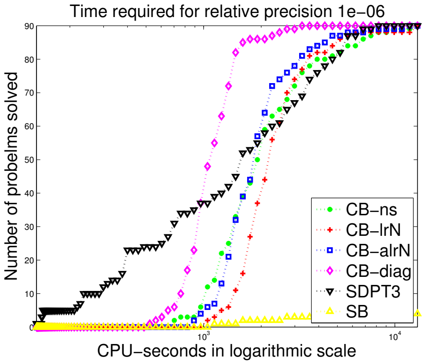




Five instances per choice of n and constraint support order $\in \{3, 4, 5\}$

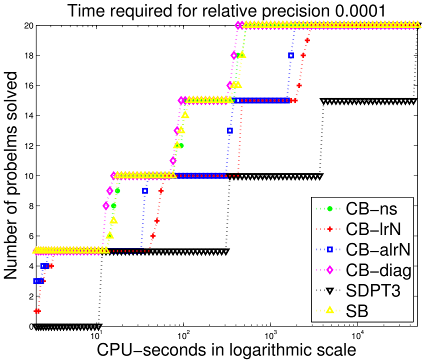
Larger Instances: $n \in \{1, \ldots, 6\} \cdot 1000$ and $m = 5000$

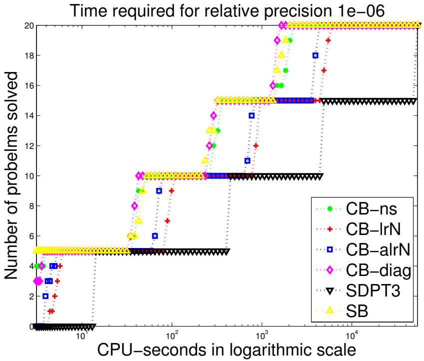




Five instances per choice of n and constraint support order $\in \{3, 4, 5\}$

Max-Cut 3D-Grids: n^3, $n \in \{10, 15, 20, 25\}$





Five instances with random ± 1 edge weights per choice of n

Number of Descent Steps, Small Instances

relative precision 10^{-6}, average and variance over 15 instances

n	m	CB-ns	CB-fN	CB-lrN	CB-alrN	CB-diag	SDPT3	SB
100	100	37 (6.11)	20 (3.44)	33 (6.86)	33 (6.09)	38 (21.3)	11 (0.573)	*43 (10.4)
300	100	43 (5.96)	22 (4.7)	38 (8.5)	39 (9.86)	37 (8.65)	13 (0.49)	53 (10.2)
500	100	58 (12.7)	27 (6.67)	50 (11.1)	51 (11.2)	52 (20)	14 (0.611)	69 (25.1)
100	500	42 (5.44)	27 (3.07)	42 (5.56)	42 (5.3)	50 (15.4)	11 (0.499)	*48 (3.35)
300	500	59 (11.1)	34 (5.04)	56 (10.2)	57 (11.3)	57 (11.6)	13 (0.806)	54 (6.3)
500	500	66 (11.5)	37 (5.23)	62 (12.2)	63 (12.4)	59 (15.9)	14 (0.596)	64 (15.6)
100	1000	51 (7)	32 (3.25)	50 (8.13)	49 (8.26)	60 (17.9)	10 (0.249)	*55 (2.46)
300	1000	59 (6.76)	36 (5.84)	59 (6.81)	59 (6.31)	60 (7.8)	12 (0.442)	*55 (3.26)
500	1000	67 (10.8)	42 (5.44)	67 (11.2)	67 (11.1)	67 (10.5)	13 (0.442)	*58 (3.64)

$*$ not all instances achieved the required precision

Number of Oracle Calls, Small Instances

relative precision 10^{-6}, average and variance over 15 instances

n	m	CB-ns	CB-fN	CB-lrN	CB-alrN	CB-diag	SDPT3	SB
100	100	75 (25.6)	44 (24.9)	49 (15.8)	52 (15.6)	54 (31.3)	11 (0.573)	255 (504)
300	100	155 (60.4)	75 (44.2)	104 (41.4)	110 (49.1)	86 (29.3)	13 (0.49)	279 (171)
500	100	314 (135)	95 (44.6)	195 (102)	199 (108)	163 (132)	14 (0.611)	464 (399)
100	500	83 (18.8)	68 (27.8)	69 (13.7)	68 (12.6)	76 (20.7)	11 (0.499)	*119453 ($1.03\cdot10^5$)
300	500	178 (110)	142 (132)	125 (46.3)	127 (54.4)	107 (32.3)	13 (0.806)	289 (207)
500	500	295 (211)	180 (129)	187 (99.9)	188 (99.8)	143 (75.5)	14 (0.596)	532 (462)
100	1000	117 (35.6)	90 (25.4)	96 (21)	97 (22.6)	113 (34)	10 (0.249)	*213306 ($3.65\cdot10^4$)
300	1000	151 (41.8)	110 (59.8)	123 (23.3)	124 (24.1)	118 (19.9)	12 (0.442)	*25553 ($3.58\cdot10^4$)
500	1000	238 (159)	152 (65.1)	177 (83.8)	178 (86.7)	148 (37)	13 (0.442)	*15803 ($3.12\cdot10^4$)

$*$ not all instances achieved the required precision

Conclusions

Scaling works well and behaves as (or even better than) expected:

- The number of oracle calls is reduced significantly
  Newton $<$ Low Rank $<$ fat Bundle

Conclusions

Scaling works well and behaves as (or even better than) expected:

- The number of oracle calls is reduced significantly
  Newton $<$ Low Rank $<$ fat Bundle
- CB-diag is more stable, accurate, and efficient than SB

Conclusions

Scaling works well and behaves as (or even better than) expected:

- The number of oracle calls is reduced significantly
  Newton $<$ Low Rank $<$ fat Bundle
- CB-diag is more stable, accurate, and efficient than SB
- Newton is attractive for small matrices and many constraints, but interior point methods seem preferable.
  [In the end the QSP system is of size $O(m)$.]

Conclusions

Scaling works well and behaves as (or even better than) expected:

- The number of oracle calls is reduced significantly
  Newton $<$ Low Rank $<$ fat Bundle
- CB-diag is more stable, accurate, and efficient than SB
- Newton is attractive for small matrices and many constraints, but interior point methods seem preferable.
  [In the end the QSP system is of size $O(m)$.]
- Diagonal low rank scaling is attractive for large matrices and few constraints.

Conclusions

Scaling works well and behaves as (or even better than) expected:

- The number of oracle calls is reduced significantly
  Newton $<$ Low Rank $<$ fat Bundle
- CB-diag is more stable, accurate, and efficient than SB
- Newton is attractive for small matrices and many constraints, but interior point methods seem preferable.
  [In the end the QSP system is of size $O(m)$.]
- Diagonal low rank scaling is attractive for large matrices and few constraints.
- Scaling allows a relative precision of 10^{-6} routinely with fast initial convergence.

Conclusions

Scaling works well and behaves as (or even better than) expected:

- The number of oracle calls is reduced significantly
  Newton $<$ Low Rank $<$ fat Bundle
- CB-diag is more stable, accurate, and efficient than SB
- Newton is attractive for small matrices and many constraints, but interior point methods seem preferable.
  [In the end the QSP system is of size $O(m)$.]
- Diagonal low rank scaling is attractive for large matrices and few constraints.
- Scaling allows a relative precision of 10^{-6} routinely with fast initial convergence.
- The cost of solving QSP might be reducible by Toh's approach.

Conclusions

Scaling works well and behaves as (or even better than) expected:

- The number of oracle calls is reduced significantly
  Newton $<$ Low Rank $<$ fat Bundle
- CB-diag is more stable, accurate, and efficient than SB
- Newton is attractive for small matrices and many constraints, but interior point methods seem preferable.
  [In the end the QSP system is of size $O(m)$.]
- Diagonal low rank scaling is attractive for large matrices and few constraints.
- Scaling allows a relative precision of 10^{-6} routinely with fast initial convergence.
- The cost of solving QSP might be reducible by Toh's approach.

$\rightarrow$ Scope of scaled CB: fast low precision results, cutting plane approaches, high precision results with large matrices and few constraints.

Thank you for your attention!