Solution methods for SDP arising from combinatorial optimization problems

Franz Rendl

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

Alpen-Adria-Universität Klagenfurt

Austria

Algorithms for SDP

- Interior-Point Methods
(Existence of) Central-Path (CP), Newton's Method to follow CP, practical limitations and existing software

- Interior-Points and Cutting Planes
How to deal with (too many) combinatorial cutting planes, Lagrangian Dual, Bundle methods combined with SDP ?

- Projection Methods
How to deal with (too many) primal equations if matrix dimension n is not too big ?

- Spectral Bundle Method
How to deal with (too many) primal constraints if n gets bigger?

Strong duality (from before)

Strong duality (primal=dual and optima are attained) holds if we assume that both the primal and the dual problem have strictly feasible points ($X, Z \succ 0$).
Then it follows from the general Karush-Kuhn-Tucker theory that (X, y, Z) is optimal if and only if

$$A(X) = b, \; X \succeq 0, \;\; A^T(y) - Z = C, \; Z \succeq 0, \;\; ZX = 0.$$

Now too many equations as ZX need not be symmetric.

Central Path

For this part we assume:

(A) $\quad \exists$ primal and dual feasible points $X, Z \succ 0$.

Consider, for $\mu > 0$ the system:

$$(CP) \quad A(X) = b,\ Z = A^T y - C,\ ZX = \mu I$$

over $X, Z \succeq 0$.

Central Path

For this part we assume:

$$(A) \quad \exists \text{ primal and dual feasible points } X, Z \succ 0.$$

Consider, for $\mu > 0$ the system:

$$(CP) \quad A(X) = b, \; Z = A^T y - C, \; ZX = \mu I$$

over $X, Z \succeq 0$.
Fundamental Theorem for Interior-Point methods (see SDP Handbook, Chapter 10):

$$(CP) \text{ has unique solution } \forall \mu > 0 \iff (A) \text{ holds.}$$

By inverse function theorem, this solution $(X(\mu)), y(\mu), Z(\mu))$ forms smooth curve, called Central Path.

Central Path Equations (2)

The system defining (CP) is overdetermined. Several ways to fix this:
Replace $ZX - \mu I = 0$ by

1. $Z - \mu X^{-1} = 0$
2. $X - \mu Z^{-1} = 0$
3. $ZX + XZ - 2\mu I = 0$
4. $P(.)P^{-1} + (P(.)P^{-1})^T$ Monteiro-Zhang family

These lead to different linearizations.
Path following methods: Follow the central path by finding points (close to it) for a decreasing sequence of μ.

Interior-Point Methods to solve SDP (1)

Primal-Dual Path-following Methods:

maintain $X, Z \succeq 0$ and try to reach feasibility and optimality.
Use Newton's method applied to perturbed problem
$ZX = \mu I$ or variant from before, and iterate for $\mu \rightarrow 0$.
At start of iteration: $(X \succ 0, y, Z \succ 0)$
Linearized system (CP) 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

Basic SDP Relaxation of Max-Cut

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

Basic SDP Relaxation of Max-Cut

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

Seconds on a PC (Pentium 4, 2.1 Ghz). Implementation in MATLAB, 30 lines of source code

Example: Lovasz Theta Function

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

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

The number of constraints depends on the edge set E.

n	100	200	300	400
$\|E\|$ time	487 1	2047 30	4531 309	7949 1583
$\|E\|$ time	1240 7	5099 371		
$\|E\|$ time	2531 34	10026 2735		

Further details: Dukanovic, Rendl: Math. Prog. 109 (2007)

Timings: Random SDP

Each A_i is nonzero only on randomly chosen 4×4 submatrix, main diagonal is 0.
SEDUMI seconds with default setting.

n	m	secs.
100	1000	11
100	2000	159
200	2000	151
200	5000	2607
300	5000	2395

No attempt with larger m. Memory (!!) and time (!!)

For more results, see Mittelmann's site:
`http://plato.asu.edu/ftp/sdplib.html`

What if m is too large?

We consider

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

where $b \in \mathbb{R}^m$ and m is too large, for instance $m > 10,000$.

Two ideas:

- Suppose we can split the constraints into two parts so that including only one part makes SDP easy $\rightarrow$ work on partial Lagrangian dual
- Use projection methods

Partial Lagrangian

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

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: $(F = \{X : A(X) = a,\ X \succeq 0\})$:

$$f(y) := \max_{X\in F} L(X, y) = b^T y + \max_{x\in F}\langle C - B^T(y), X\rangle$$

Properties of $f(y)$

Recall: $f(y) = b^T y + \max_{x \in F} \langle C - B^T(y), X \rangle$
f is convex (max of linear functions)
weak duality: $z^* \leq f(y) \; \forall y$ (holds by construction)
strong duality: $z^* = \min_y f(y)$ (holds under Slater condition)

Properties of $f(y)$

Recall: $f(y) = b^T y + \max_{x \in F} \langle C - B^T(y), X \rangle$
f is convex (max of linear functions)
weak duality: $z^* \leq f(y) \; \forall y$ (holds by construction)
strong duality: $z^* = \min_y f(y)$ (holds under Slater condition)

Basic assumption: We can compute $f(y)$ easily, yielding also maximizer X^* and $g^* := b - B(X^*)$.
$f(y) = b^T y + \langle C - B^T(y), X^* \rangle = y^T g^* + \langle C, X^* \rangle$, so

$$f(v) \geq v^T g^* + \langle C, X^* \rangle,$$

therefore, using $\langle C, X^* \rangle = f(y) - y^T g^*$ we get

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

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

Minimize $f(y)$ using convex optimization

$$z^* = \min_y f(y) \leq f(y) \ \ \forall y.$$

Any y provides upper bound on z^* by weak duality, and we can try to find best upper bound by minimizing the convex but nonsmooth function $f(y)$.

Use simple subgradient or more refined bundle methods to get approximate minimizer.

See Lemarechal, Kiwiel 1970s, Zowe, Shor, Nesterov 1980' for the general approach.

See Fischer, Gruber, R. and Sotirov Math Prog 105 (2006) for applications to Max-Cut

Computations: SDP + triangles

Big graphs (from Helmberg). Use the bundle method to deal with triangle inequalities. The number of function evaluations of f is 50 for $n = 800$, and 30 for $n = 2000$.

name	n	cut	initial gap (%)	final (%)	minutes
G6	800	2172	22.29	18.15	43.11
G11	800	564	11.56	1.54	60.20
G14	800	3054	4.51	2.84	59.68
G18	800	985	18.38	7.96	69.19
G22	2000	13293	6.34	5.66	278.06
G27	2000	3293	25.77	22.94	406.66
G39	2000	2373	21.27	12.63	533.36

Projection methods (for large m)

- Boundary Point method (based on augmented Lagrangian), see Povh, R., Wiegele, Computing (2007)
- Augmented Primal-Dual Method (based on alternate projections), see Jarre, R. SIOPT (to appear)

$$(D) \quad \min_y b^T y \text{ s.t. } A^T(y) - C = Z \succeq 0.$$

Lagrangian:

$$\max_X \min_{y, Z \succeq 0} b^T y + \langle X, Z + C - A^T(y) \rangle.$$

Augmented Lagrangian

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 minimization:
X and σ are fixed.

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

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

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

Note that dependence on Z looks like projection problem, but with additional variables y.
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:

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

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

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

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

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

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

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

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:

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

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

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 : $\;\; X = -\sigma W(y)_-$

Note: Outer stopping condition is dual feasibility.

See Povh, R., Wiegele, Computing (2006), and Malick, Povh, R., Wiegele (working paper, Klagenfurt 2008)

Theta Number revisited

Comparing Boundary point method (bpm) (Povh, R., Wiegele (2006)) to Kocvara and Stingl's KS iterative SDP solver (2006). Timings of KS from their paper, their machine is at least twice as fast as ours, 5 digits accuracy. Random graphs from the Kim Toh collection.

graph	n	$\|E\|$	KS (secs)	bpm
theta82	400	23871	695	87
theta83	400	39861	852	70
theta102	500	37466	1231	143
theta103	500	62515	1960	110
theta104	500	87244	2105	124
theta123	600	90019	2819	205
theta162	800	127599	6004	570

Theta: big DIMACS graphs

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

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

Comparing IP and projection methods

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

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

Augmented Primal-Dual Method

(This is joint work with Florian Jarre.)

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

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

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

From Linear Algebra:

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

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

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

Augmented Primal-Dual Method (2)

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

This suggests the following iteration:

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

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

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

Augmented Primal-Dual Method (3)

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

An improvement: Consider

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

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

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

We solve SDP by minimizing ϕ over K.

Augmented Primal-Dual Method (4)

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

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

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

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

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

Augmented Primal-Dual Method (5)

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

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

The optimal value is 11951.726.

Random SDP

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

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

Large-Scale SDP

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

What if n is much larger?

Spectral Bundle Method

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

Constant trace SDP

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

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

The constant trace property implies:

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

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

Constant trace property holds for many combinatorially derived SDP!

Reformulating Constant Trace SDP

Reformulate dual as follows:

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

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

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

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

Dual SDP as eigenvalue optimization

Compute dual variable λ explicitely:

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

Dual equivalent to

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

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

Spectral Bundle Method (1)

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

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

Define

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

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

Idea 1: Minorant for $f(y)$

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

Consider W of the form $W = PVP^T$ with new $k \times k$ matrix variable V.

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

Spectral Bundle Method (2)

Idea 2: Proximal point approach

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

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

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

Quadratic Subproblem (1)

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

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

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

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

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

Quadratic Subproblem (2)

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

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

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

Large-Scale Max-Cut SDP

We consider again

$$\max\langle L, X\rangle \text{ s.t. } \operatorname{diag}(X) = e,\ X \succeq 0.$$

Now $n \geq 10,000.$

- We compute upper bound on SDP relaxation for Max-Cut using the spectral bundle method, and also apply the Goemans-Williamson hyperplane rounding technique to generate cuts.
- Sparse graphs with n up to 50,000.
- The graphs are generated as the union of k matchings.

Large Max-Cut instances

n	k	upper bnd	cut	time (secs)
20,000	10	143.3	131.3	330
20,000	20	261.9	244.8	536
20,000	50	598.1	571.1	1255
30,000	10	214.9	197.2	753
30,000	20	393.3	367.4	990
30,000	50	897.9	857.3	2330
40,000	10	286.9	262.7	1180
40,000	20	524.6	489.8	1650
50,000	10	358.9	328.5	1800

About half the time is used to generate cuts, 50 iterations of the spectral bundle method, values scaled (10^{-3}).

Last Slide

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