SDPNAL+: A MATLAB software package for large-scale SDPs with a user-friendly interface

Defeng Sun

Department of Applied Mathematics, The Hong Kong Polytechnic University

SIAM ALA Conference 2018

Joint work with: Kim-Chuan Toh, National University of Singapore
Yancheng Yuan (NUS)
Xinyuan Zhao (Beijing U Tech.)
Past contributor: Liuqin Yang

Outline

- SDP and SDP+ (variable is positive semidefinite and bounded)
- Some examples of SDP+
- User-friendly interface
- Phase I: An inexact symmetric Gauss-Seidel (sGS) ADMM for SDP+
- An sGS decomposition theorem for convex composite QP
- Phase II: An augmented Lagrangian method (ALM) for SDP+
- A semismooth Newton-CG (SNCG) method for solving ALM sub-problems
- SDPNAL+: practical implementation of the 2 phase method
- Numerical experiments

SDP and SDP+ problems

$\mathbb{S}^n_+$ = cone of positive semidefinite matrices. Write $X \succeq 0$ if $X \in \mathbb{S}^n_+$.

$$\text{(SDP)} \quad \min \left\{ \langle C,\, X \rangle \mid \mathcal{A}(X) = b,\; X \in \mathbb{S}^n_+ \right\}$$

where $C \in \mathbb{S}^n$, $b \in \mathbb{R}^m$ are given data; $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is a linear map.

$$\text{(SDP+)} \quad \min \left\{ \langle C,\, X \rangle \mid \mathcal{A}(X) = b,\; X \in \mathbb{S}^n_+,\; X \in \mathcal{N} \right\}$$

where $\mathcal{N} = \{X \in \mathbb{S}^n \mid L \leq X \leq U\}$ and L, U are given bounds (entries allow to take $-\infty$, ∞ respectively).

Important case: $\mathcal{N} = \{X \in \mathbb{S}^n \mid X \geq 0\}$, i.e., DNN (doubly nonnegative) SDP.

(SDP) is solvable by powerful interior-point methods if n and m are not too large, say, $n \leq 2000$, $m \leq 10,000$.

m large $\Rightarrow$ $m \times m$ dense "Hessian" matrix cannot be stored explicitly. For $m = 10^5$, needs 100GB RAM memory!

Current research interests focus on $n \leq 5000$ but $m \gg 10,000$.

SDPNAL was developed around 2008/09 for (SDP).

In 2012/13, it was extended to SDPNAL+ for (SDP+) directly without introducing extra equality constraints $X = Y$ to convert $X \in \mathbb{S}^n_+ \cap \mathcal{N}$ to $X \in \mathbb{S}^n_+$ and $Y \in \mathcal{N}$.

Now our solver SDPNAL+ can solve general SDP problems:

$$
\begin{array}{lll}
\text{(genSDP)} & \min & \sum_{i=1}^{N} \langle C_i,\, X_i \rangle \\
& \text{s.t.} & \sum_{i=1}^{N} \mathcal{A}_i(X_i) = b \quad \text{(equalities)} \\
& & l \le \sum_{i=1}^{N} \mathcal{B}_i(X_i) \le u \quad \text{(inequalities)} \\
& & X_i \in \mathbb{K}_i \ \text{(cone)}, \quad X_i \in \mathcal{N}_i \ \text{(bounds)}, \quad i = 1:N
\end{array}
$$

where $\mathbb{K}_i$ is either a PSD cone or nonnegative orthant. Currently extending $\mathbb{K}_i$ to other cones such as SOCP.

Large scale SDP and SDP+: a brief history

- Parallel IPM [Benson, Borchers, Fujisawa, ... 03-present]
- First-order gradient methods on NLP formulation (low accuracy) [Burer-Monteiro 03]
- Inexact IPM [Kojima, Toh 04]
- Gen. Lag. method on barrier-penalized dual [Kocvara-Stingl 03]
- ALM on primal SDP from relaxation of lift-and-project scheme [Burer-Vandenbussche 06]
- Boundary-point method: BCD-ALM on dual [Rendl et al. 06] Reg. methods for SDP $\equiv$ ADMM on dual [Malick-Povh-Rendl 09]
- SDPNAL: ADMM+SNCG-ALM on dual [Zhao-Sun-Toh 10]
- SDPAD: ADMM on dual [Wen et al. 10] (used SDPNAL template)
- 2EBD: hybrid proximal extra-gradient method on primal [Monteiro et al. 13] (used SDPNAL template)
- ADMM+: convergent sGS-ADMM on SDP+ [Sun-Toh-Yang 15]
- SDPNAL+: SNCG-ALM on SDP+ [Yang-Sun-Toh 15]

In nearest correlation matrix problem, given data matrix $U \in \mathbb{S}^n$, we want to solve

$$\text{(NCM)} \quad \min_{X} \left\{ \frac{1}{2} \| H \circ (X - U) \|_1 \;\middle|\; \mathrm{Diag}(X) = \mathbf{1},\; X \succeq 0 \right\}$$

▸ NCM

where $H \in \mathbb{S}^n$ has nonnegative entries and "$\circ$" is the Hardamard product.

In clustering, given data vectors $\{p_i\}_{i=1}^n$, the goal is to cluster them into k clusters. A possible model [Peng-Wei 07] is:

$$\min \left\{ \langle D, X \rangle \mid \langle I, X \rangle = k,\; X\mathbf{1} = \mathbf{1},\; X \in \mathbb{S}^n_+,\; X \geq 0 \right\}$$

▸ Clustering

where $D_{ij} = \|p_i - p_j\|^2$.
Note: D can also be other affinity matrix.

Maximum stable set problem a graph $G = (V, \mathcal{E})$

A stable set S is subset of V such that no vertices in S are adjacent.
Maximum stable set problem: find S with maximum cardinality. Let

$$x_i = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases} \qquad \Rightarrow \quad |S| = \sum_{i=1}^{n} x_i.$$

A common formulation of the max-stable-set problem:

$$\alpha(G) := \max\left\{|S| = \tfrac{1}{|S|}\sum_{ij} x_i x_j \mid x_i x_j = 0 \; \forall\, (i,j) \in \mathcal{E}, x \in \{0,1\}^n\right\}$$

$$\Downarrow \quad X := xx^T/|S|$$

$$\max\left\{\langle E, X\rangle \mid X_{ij} = 0 \; \forall\, (i,j) \in \mathcal{E}, \;\; \langle I, X\rangle = 1\right\}$$

SDP relaxation: $X = xx^T/|S| \Rightarrow X \succeq 0$, get

$$\theta(G) \;:=\; \max\left\{\langle E, X\rangle : X_{ij} = 0 \; \forall\, (i,j) \in \mathcal{E}, \;\; \langle I, X\rangle = 1, \; X \succeq 0\right\}$$

$\theta_+(G) \;:=\; n(n+1)/2$ additional constraints $X \geq 0$ ▸ theta

Assign n facilities to n locations [Koopmans and Beckmann (1957)]

$A = (a_{ij})$ where a_{ij}= flow from facility i to facility j

$B = (b_{kl})$ where b_{kl}= distance from location k to location l

cost of assignment $\pi = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{\pi(i)\pi(j)}$

$$\min_P \left\{ \langle B \otimes A,\, \mathrm{vec}(P)\mathrm{vec}(P)^T \rangle \mid P \text{ is } n \times n \text{ permutation matrix} \right\}$$

SDP+ relaxation [Povh and Rendl, 09]:
relax $\mathrm{vec}(P)\mathrm{vec}(P)^T$ to the $n^2 \times n^2$ variable $X \in \mathbb{S}_+^{n^2}$ and $X \geq 0$

$$\text{(QAP)} \ \min \left\{ \langle B \otimes A,\, X \rangle \mid \mathcal{A}(X) - b = 0,\ X \in \mathbb{S}_+^{n^2},\ X \geq 0 \right\}$$

where the linear constraints (with $m = 3n(n+1)/2$) encode the condition $P^T P = I_n$, $P \geq 0$.

A basic user-friendly interface

Consider the NCM problem. ▸ NCM

```
n = 100;
G = randn(n,n);
G = 0.5*(G + G');

model = ccp_model('NCM');
   X = var_sdp(n,n);
   model.add_variable(X);
   model.minimize(l1_norm(X-G));
   model.add_affine_constraint(map_diag(X)==ones(n,1));
model.solve;
```

Consider the $\theta+$ problem of a graph with adjacency matrix G.

▸ theta

```
n = 200;
G = triu(sprand(n,n,0.5),1);
[IE,JE] = find(G);
n = length(G);

model = ccp_model('theta');
    X = var_sdp(n,n);
    model.add_variable(X);
    model.maximize(sum(X));
    model.add_affine_constraint(trace(X) == 1);
    model.add_affine_constraint(X(IE,JE) == 0);
    model.add_affine_constraint(X >= 0);
model.solve;
```

$$
\begin{aligned}
\min \quad & \operatorname{trace}(X^{(1)}) + \operatorname{trace}(X^{(2)}) + \operatorname{sum}(X^{(3)}) \\
\text{s.t.} \quad & -X^{(1)}_{12} + 2X^{(2)}_{33} + 2X^{(3)}_{2} = 4 \quad \text{(equalities)} \\
& 2X^{(1)}_{23} + X^{(2)}_{42} - X^{(3)}_{4} = 3 \\
& 2 \le -X^{(1)}_{12} - 2X^{(2)}_{33} + 2X^{(3)}_{2} \le 7 \quad \text{(inequalities)} \\
& X^{(1)} \in \mathbb{S}^{6}_{+},\ X^{(2)} \in \mathbb{R}^{5\times 5},\ X^{(3)} \in \mathbb{R}^{7}_{+} \quad \text{(cones)} \\
& 0 \le X^{(1)} \le 10E_6, \quad 0 \le X^{(2)} \le 8E_5 \quad \text{(bounds)}
\end{aligned}
$$

```
n1 = 6; n2 = 5; n3 = 7;
M = ccp_model('Example_simple');
X1=var_sdp(n1,n1);  X2=var_nn(n2,n2); X3=var_nn(n3);
M.add_variable(X1,X2,X3);
M.minimize(trace(X1) + trace(X2) + sum(X3));
M.add_affine_constraint(-X1(1,2)+2*X2(3,3)+2*X3(2)==4);
M.add_affine_constraint(2*X1(2,3)+X2(4,2)-X3(4) == 3);
M.add_affine_constraint(2<=-X1(1,2)-2*X2(3,3)+2*X3(2)<=7);
M.add_affine_constraint(0 <= X1 <= 10);
M.add_affine_constraint(X2 <= 8);
M.solve;
```

For simplicity, consider only $\mathcal{N} = \{X \in \mathbb{S}^n \mid X \geq 0\}$.
Dual of SDP+ and its augmented Lagrangian function are given by:

$$\text{(D)} \quad \min\{-\langle b,\, y\rangle + \delta_{\mathbb{S}^n_+}(S) + \delta_{\mathcal{N}}(Z) \mid \mathcal{A}^* y + S + Z = C\}$$

(a linearly constrained convex problem with 3 blocks of variables);

$$\begin{aligned}\mathcal{L}_\sigma(y, S, Z; X) = {} & -\langle b,\, y\rangle + \ \delta_{\mathbb{S}^n_+}(S) + \delta_{\mathcal{N}}(Z) \\ & + \langle \mathcal{A}^* y + S + Z - C, X\rangle + \frac{\sigma}{2}\|\mathcal{A}^* y + S + Z - C\|^2\end{aligned}$$

(quadratic in (y, S, Z) + nonsmooth terms in S, Z)

KKT conditions:

$$\mathcal{R}_{\mathrm{KKT}}(y, S, Z; X) := \begin{pmatrix} AX - b \\ S - \Pi_{\mathbb{S}^n_+}(S - X) \\ Z - \Pi_{\mathcal{N}}(Z - X) \\ \mathcal{A}^* y + S + Z - C \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Input $(y_0, S_0, Z_0; X_0)$. For $k = 0, 1, \ldots$, let $\widehat{C}^k = C - \sigma^{-1} X^k$

(1a) $y^{k+1} = \text{argmin}_{y \in \mathbb{R}^m} \mathcal{L}_\sigma(y, S^k, Z^k; X^k)$

(1b) $S^{k+1} = \text{argmin}_{S \in \mathbb{S}^n_+} \mathcal{L}_\sigma(y^{k+1}, S, Z^k; X^k) = \Pi_{\mathbb{S}^n_+}(\widehat{C}^k - \mathcal{A}^* y^{k+1} - Z^k)$

(2) $Z^{k+1} = \text{argmin}_{Z \in \mathcal{N}} \mathcal{L}_\sigma(y^{k+1}, S^{k+1}, Z; X^k) = \Pi_{\mathcal{N}}(\widehat{C}^k - \mathcal{A}^* y^{k+1} - S^{k+1})$

(3) $X^{k+1} = X^k + \tau\sigma(\mathcal{A}^* y^{k+1} + S^{k+1} + Z^{k+1} - C)$, where $\tau \in (0, \frac{1+\sqrt{5}}{2})$ is the step-length.

Direct extension of 2-block ADMM is not guaranteed to converge [Chen-He-Ye-Yuan, v155, MP 2016]

But sGS-ADMM is guaranteed to converge!

Input $(y_0, S_0, Z_0; X_0)$. For $k = 0, 1, \ldots,$ let $\widehat{C}^k = C - \sigma^{-1} X^k$

(1a) $\widehat{\mathbf{y}}^{k+1} \approx \text{argmin}_{y \in \mathbb{R}^m} \mathcal{L}_\sigma(y, S^k, Z^k; X^k)$

(1b) $S^{k+1} = \text{argmin}_{S \in \mathbb{S}^n_+} \mathcal{L}_\sigma(\widehat{\mathbf{y}}^{k+1}, S, Z^k; X^k) = \Pi_{\mathbb{S}^n_+}(\widehat{C}^k - \mathcal{A}^* \widehat{\mathbf{y}}^{k+1} - Z^k)$

(1c) $\boxed{y^{k+1} \approx \text{argmin}_{y \in \mathbb{R}^m} \mathcal{L}_\sigma(y, S^{k+1}, Z^k; X^k)}$

(2) $Z^{k+1} = \text{argmin}_{Z \in \mathcal{N}} \mathcal{L}_\sigma(y^{k+1}, S^{k+1}, Z; X^k) = \Pi_{\mathcal{N}}(\widehat{C}^k - \mathcal{A}^* y^{k+1} - S^{k+1})$

(3) $X^{k+1} = X^k + \tau\sigma(\mathcal{A}^* y^{k+1} + S^{k+1} + Z^{k+1} - C)$

In Step 1, the AL function $\mathcal{L}_\sigma$ for the block (y, S) has the form:

$$\mathcal{L}_\sigma(y, S) \equiv \delta_{\mathbb{S}^n_+}(S) + \frac{\sigma}{2}\|\mathcal{A}^* y + S + Z^k + \widehat{C}^k\|^2 - \langle b, y \rangle$$

(QP in (y, S) + nonsmooth term in S)

(1a)–(1c) is equivalent to minimizing $\mathcal{L}_\sigma(y, S)$ + sGS proximal term.
The steps are based on an sGS decomposition theorem.

Theorem Suppose that the KKT conditions of (SDP+) has a solution. Let $\{(y^k, S^k, Z^k, X^k)\}$ be the sequence generated by the inexact sGS-ADMM. Then $\{X^k\}$ converges to an optimal solution of (SDP+) and $\{(y^k, S^k, Z^k)\}$ converges to an optimal solution of its dual.

[1] D.F. Sun, K.C. Toh and L.Q. Yang, A convergent 3-block semi-proximal ADMM for conic programming with 4-type constraints, v25, SIOPT 2015.

[2] X.D. Li, D.F. Sun, K.C. Toh, A Schur complement based semiproximal ADMM for convex ..., v155, MP 2016. Schur-complement-ADMM

[3] X.D. Li, D.F. Sun, K.C. Toh, QSDPNAL: A two-phase augmented Lagrangian method for convex quadratic SDP, MPC 2018. Section 2: sGS decomposition theorem, Schur-complement-ADMM = sGS-ADMM

[4] L. Chen, D.F. Sun, K.C. Toh, An efficient inexact symmetric Gauss-Seidel based majorized ADMM for ..., v161, MP 2017. inexact sGS-ADMM

[5] X.D. Li, D.F. Sun, K.C. Toh, A block sGS decomposition theorem for convex composite quadratic programming and its applications, MP 2018. sGS-ADMM = Schur-complement-ADMM, sSOR-extension

Theorem [Han-Sun-Zhang, MOR 2018: exact version]
Let $\Omega_{\rm KKT} \neq \emptyset$ be the KKT solution set. Suppose that an error bound condition holds for $\mathcal{R}_{\rm KKT}$ at an optimal solution $u^* = (y^*, S^*, Z^*, X^*)$ that $u^k = (y^k, S^k, Z^k, X^k)$ converges to, i.e, $\exists$ $\eta, r > 0$ s.t.

$$\mathrm{dist}(u, \Omega_{\rm KKT}) \;\; \leq \;\; \eta \|\mathcal{R}_{\rm KKT}(u)\| \quad \forall\, u \in B_r(u^*).$$

Then $\exists$ $\mu \in (0,1)$ depending on η s.t.

$$\mathrm{dist}(u^{k+1}, \Omega_{\rm KKT}) \;\; \leq \;\; \mu\, \mathrm{dist}(u^k, \Omega_{\rm KKT}) \quad \forall\, k \text{ sufficiently large.}$$

Inexact version can be established via the analysis in [Chen-Sun-Toh, MP 2017] and [Han-Sun-Zhang, MOR 2018].

Detour: block symmetric Gauss-Seidel (sGS) decomposition

Consider a convex composite QP with 3 blocks:

$$\min\left\{p(x_1)+h(x)\mid x=(x_1;\,x_2,x_3)\in\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}\times\mathbb{R}^{n_3}\right\}$$

Convex quadratic function $h(x):=\frac{1}{2}\langle x,\,\mathcal{H}x\rangle-\langle b,\,x\rangle$

Closed proper convex fun. $p:\mathbb{R}^{n_1}\to(-\infty,+\infty]$, e.g. $p(x_1)=\|x_1\|_\infty$

Write $\mathcal{H}=\mathcal{U}^*+\mathcal{D}+\mathcal{U}$, $\mathcal{D}$ diagonal blocks, $\mathcal{U}$ strict upper triangular part. Assume $\mathcal{D}$ invertible.

Define $\mathrm{sGS}(\mathcal{H}):=\mathcal{U}\mathcal{D}^{-1}\mathcal{U}^*$ (symmetric Gauss-Seidel decomp)

Given $\bar{x}$, define

$$x^+:=\text{argmin}_x\left\{p(x_1)+h(x)+\frac{1}{2}\|x-\bar{x}\|^2_{\mathrm{sGS}(\mathcal{H})}\right\}$$

Next theorem: can compute x^+ using one sGS cycle!

If $p(x_1)$ is absent, we get the classical block sGS iteration.

Theorem [Li-Sun-Toh 2015]

It holds that $\mathcal{H} + \mathrm{sGS}(\mathcal{H}) = (\mathcal{D} + \mathcal{U})\mathcal{D}^{-1}(\mathcal{D} + \mathcal{U}^*) \succ 0.$

Backward GS: $3 \to 2.$ Compute

$$x_3' = \text{argmin } p(\bar{x}_1) + h(\bar{x}_1, \bar{x}_2, x_3) \;=\; \mathcal{H}_{33}^{-1}(b_3 - \mathcal{H}_{13}^*\bar{x}_1 - \mathcal{H}_{23}^*\bar{x}_2)$$

$$x_2' = \text{argmin } p(\bar{x}_1) + h(\bar{x}_1, x_2, x_3') \;=\; \mathcal{H}_{22}^{-1}(b_2 - \mathcal{H}_{12}^*\bar{x}_1 - \mathcal{H}_{23}\bar{x}_3')$$

Forward GS: $1 \to 2 \to 3.$ Compute

$$x_1^+ = \text{argmin } p(x_1) + h(x_1, x_2', x_3') \quad \text{(non-smooth/non-quadratic)}$$

$$x_2^+ = \text{argmin } p(x_1^+) + h(x_1^+, x_2, x_3') = \mathcal{H}_{22}^{-1}(b_2 - \mathcal{H}_{12}^* x_1^+ - \mathcal{H}_{23} x_3')$$

$$x_3^+ = \text{argmin } p(x_1^+) + h(x_1^+, x_2^+, x_3) = \mathcal{H}_{33}^{-1}(b_3 - \mathcal{H}_{13}^* x_1^+ - \mathcal{H}_{23}^* x_2^+)$$

Inexact computation is also allowed! So can use PCG to solve large linear systems.

Theorem [Li-Sun-Toh 2015]
Backward GS: For $i = s, \ldots, 2$, compute

$$x'_i \;=\; \mathcal{H}_{ii}^{-1}\big(b_i + e'_i - \textstyle\sum_{j=1}^{i-1} \mathcal{H}_{ji}^* \bar{x}_j - \sum_{j=i+1}^{s} \mathcal{H}_{ij} x'_j\big).$$

Forward GS: For $i = 2, \ldots, s$,

$$x_1^+ \;=\; \text{argmin}\; p(x_1) + h(x_1, x'_{\geq 2}) - \langle e_1^+,\, x_1 \rangle,$$

$$x_i^+ \;=\; \mathcal{H}_{ii}^{-1}(b_i + e_i^+ - \textstyle\sum_{j=1}^{i-1} \mathcal{H}_{ji}^* x_j^+ - \sum_{j=i+1}^{s} \mathcal{H}_{ij} x'_j)$$

e^+, e' are error vectors. In this case, x^+ is the exact solution to a slightly perturbed proximal problem:

$$x^+ := \text{argmin}_x \left\{ p(x_1) + h(x) + \frac{1}{2}\|x - \bar{x}\|^2_{\mathrm{sGS}(\mathcal{H})} - \langle x,\, \Delta(e', e^+) \rangle \right\}$$

$$\Delta(e', e^+) = e^+ + \mathcal{U}\mathcal{D}^{-1}(e^+ - e').$$

Adding a large proximal term slows the convergence of sGS-ADMM!

With no proximal term added, we consider the ALM for solving dual SDP+.

(1) Compute

$$(y^{k+1}, S^{k+1}, Z^{k+1}) \approx \text{argmin}\Big\{\mathcal{L}_k(y, S, Z) := \mathcal{L}_{\sigma_k}(y, S, Z; X^k)\Big\}$$

$$= \text{argmin}\Big\{ -\langle b,\ y\rangle + \frac{\sigma}{2}\|\mathcal{A}^* y + S + Z + \widehat{C}^k\|^2 + \delta_{\mathbb{S}^n_+}(S) + \delta_{\mathcal{N}}(Z)\Big\}$$

(2) Update $X^{k+1} = X^k + \sigma_k(\mathcal{A}^* y^{k+1} + S^{k+1} + Z^{k+1} - C)$;
update $\sigma_{k+1} \uparrow \sigma_\infty \leq \infty$.

Define $X^{k+1} = X^k + \sigma_k R_D(y^{k+1}, S^{k+1}, Z^{k+1})$,

$$e^{k+1} = \left[\begin{array}{c} \mathcal{A}X^{k+1} - b \\ X^{k+1} - \Pi_{\mathbb{S}^n_+}(X^{k+1} - S^{k+1}) \\ X^{k+1} - \Pi_{\mathcal{N}^n}(X^{k+1} - Z^{k+1}) \end{array} \right].$$

In Step 1, we use the following easy-to-check stopping conditions:

$$\text{(A)} \ \|e^{k+1}\| \ \leq \ \frac{\epsilon_k^2}{1 + \|(X, y, S, Z)^{k+1}\|} \min\left\{ \frac{1}{\sigma_k}, \frac{1}{1+\|X^{k+1}-X^k\|} \right\}$$

$$\text{(B)} \ \|e^{k+1}\| \ \leq \ \frac{\eta_k^2 \|X^{k+1} - X^k\|^2}{1 + \|(X, y, S, Z)^{k+1}\|} \min\left\{ \frac{1}{\sigma_k}, \frac{1}{1+\|X^{k+1}-X^k\|} \right\}$$

where $\{\epsilon_k\}$ and $\{\delta_k\}$ are nonnegative summable sequences.

Theorem [Rockafellar 76] Let $\Omega_P \neq \emptyset$ be the primal optimal solution set and Slater's condition holds for primal problem (P). Under stopping condition (A), we have $X^k \to X^*$ and $(y^{k+1}, S^{k+1}, Z^{k+1})$ converges to a dual optimal solution.

Theorem [Cui-Sun-Toh] If in addition, the blue stopping conditions are added, and the essential primal objective function P^{obj} satisfies a quadratic growth condition at X^*, i.e., $\exists$ a neighborhood $\mathcal{U}$ of X^* and $\kappa > 0$ s.t.

$$P^{\text{obj}}(X) \;\geq\; P^{\text{obj}}(X^*) + \kappa^{-1}\text{dist}^2(X, \Omega_P) \quad \forall\, X \in \mathcal{U}$$

Then for k large, we have

$$\begin{aligned} \text{dist}(X^{k+1}, \Omega_P) \;&\leq\; \theta_k \,\text{dist}(X^k, \Omega_P) \\ \text{dual feasibility at } (y^{k+1}, S^{k+1}, Z^{k+1}) \;&\leq\; \tau_k \,\text{dist}(X^k, \Omega_P) \\ \text{dual objective gap at } (y^{k+1}, S^{k+1}, Z^{k+1}) \;&\leq\; \tau'_k \,\text{dist}(X^k, \Omega_P) \end{aligned}$$

where $\theta_k \approx \frac{\kappa}{\sqrt{\kappa^2+\sigma_k^2}}$, $\quad \tau_k \approx \frac{1}{\sigma_k}$, $\quad \tau'_k \approx \frac{\|X^k\|+\|X^{k+1}\|}{2\sigma_k}$

Larger σ_k gives faster convergence, but the inner problem is harder to solve.

For simplicity, assume $\mathcal{N} = \mathbb{S}^n$ and hence the variable Z is absent.

$$\text{argmin}_{y,S}\Big\{\mathcal{L}_\sigma(y,S) \equiv \delta_{\mathbb{S}^n_+}(S) + \frac{\sigma}{2}\|\mathcal{A}^* y + S - \widehat{C}^k\|^2 - \langle b,\, y\rangle\Big\}$$

$$\equiv \text{argmin}_y\Big\{\Phi^k(y) := -\langle b,\, y\rangle + \frac{\sigma}{2}\|\Pi_{\mathbb{S}^n_+}(\mathcal{A}^* y - \widehat{C}^k)\|^2\Big\} \text{ (project out } S\text{)}$$

Optimality condition of unconstrained subproblem in y is:

$$\nabla\Phi^k(y) = -b + \sigma\mathcal{A}\,\Pi_{\mathbb{S}^n_+}(\mathcal{A}^* y - \widehat{C}^k) = 0.$$

Solve for solution y^{k+1} by the semismooth Newton-CG (SNCG) method. Then compute $S^{k+1} = \Pi_{\mathbb{S}^n_+}(\widehat{C}^k - \mathcal{A}^* y^{k+1})$.

$\nabla\Phi^k(y)$ is not differentiable, but is strongly semismooth [Sun-Sun, 2002]. Thus SNCG is expected to have at least superlinear convergence.

A semismooth Newton-CG method (SNCG) for ALM-subproblem

Solve $\nabla\Phi^k(y) = -b + \sigma\mathcal{A}\Pi_{\mathbb{S}^n_+}(U) = 0,\ U = \mathcal{A}^*y - \widehat{C}^k$.

At the current iteration, y_l, we solve a generalized Newton equation:

$$\mathcal{H}\Delta y \approx \nabla\Phi^k(y_l), \quad \text{where } \mathcal{H}\Delta y = \sigma\mathcal{A}\,\Pi'_{\mathbb{S}^n_+}(U)[\mathcal{A}^*\Delta y] \tag{1}$$

From eigenvalue decomp: $U = QDQ^T$ with $d_1 \geq \cdots \geq d_r \geq 0 > d_{r+1} \geq \cdots \geq d_n$, we choose

$$\Pi'_{\mathbb{S}^n_+}(U)[M] = Q(\Omega \circ (Q^T M Q))Q^T, \tag{2}$$

$\Omega_{ij} = (d_i^+ - d_j^+)/(d_i - d_j)$. Let $\gamma = \{1,\ldots,r\}$, $\bar{\gamma} = \{r+1,\ldots,n\}$,

$$\Omega = \begin{bmatrix} E_{\gamma\gamma} & \Omega_{\gamma\bar{\gamma}} \\ \Omega_{\bar{\gamma}\gamma} & 0 \end{bmatrix}.$$

When problem is primal nondegenerate, cond$(\mathcal{H})$ is bounded:

$$\text{cond}(\mathcal{H}) \leq \sigma\,\Theta(1)\,\text{cond}([\mathcal{A}Q_\gamma \otimes Q_\gamma,\ \mathcal{A}Q_\gamma \otimes Q_{\bar{\gamma}}])^2$$

Exploiting second-order structured sparsity

The structure in Ω allows for efficient computation of matrix-vector multiply for CG in solving (1). Direct evaluation of

$$Y := \Pi'_{\mathbb{S}^n_+}(U)[M] = Q(\Omega \circ (Q^T M Q))Q^T$$

needs 4 matrix-matrix multiplications $= 8n^3$ operations. But with the structure of Ω, can compute Y as follows:

$$Y = H + H^T, \quad H = Q_\gamma \Big[\frac{1}{2}(UQ_\gamma)Q_\gamma^T + (\Omega_{\gamma\bar{\gamma}} \circ (UQ_{\bar{\gamma}}))Q_{\bar{\gamma}}^T \Big]$$

where $U = Q_\gamma M$. The cost is at most $6rn^2$.

If $r \approx n$, then use

$$\begin{aligned} Y &= Q(E \circ (Q^T M Q))Q^T - Q(\overline{\Omega} \circ (Q^T M Q))Q^T \\ &= M - Q(\overline{\Omega} \circ (Q^T M Q))Q^T \end{aligned}$$

where $\overline{\Omega} = E - \Omega$ has a similar structure as Ω but with a large block of 0. The cost is $6(n-r)n^2$.

Let ADMM+ denote the sGS-ADMM.

1. Generate a good starting point to warm-start SNCG-ALM:
   $(y^0, S^0, Z^0, X^0, \sigma_0) \leftarrow$ ADMM+$(\bar{y}^0, \bar{S}^0, \bar{Z}^0, \bar{X}^0, \bar{\sigma}_0)$

2. For $k = 0, 1, \ldots$

   Generate $(y^{k+1}, S^{k+1}, Z^{k+1})$ in ALM-subproblem via SNCG

   Compute X^{k+1} based on $(y^{k+1}, S^{k+1}, Z^{k+1})$, update σ_{k+1}

   If progress of SNCG-ALM is slow,

   Rescale data

   Let $(\bar{y}^k, \bar{S}^k, \bar{Z}^k, \bar{X}^k, \bar{\sigma}_k)$ denote rescaled $(y^k, S^k, Z^k, X^k, \sigma_k)$
   Rescaling is chosen such that $\|\bar{X}^k\| \approx \max\{\|\bar{S}^k\|, \|\bar{Z}^k\|\}$

   Goto Step 1: Restart with ADMM+$(\bar{y}^k, \bar{S}^k, \bar{Z}^k, \bar{X}^k, \bar{\sigma}_k)$

Robustness of SDPNAL+

$$\eta \equiv \frac{\|\mathcal{R}_{\rm KKT}(y^{k+1}, S^{k+1}, Z^{k+1}, X^{k+1})\|}{1 + \|(y^{k+1}, S^{k+1}, Z^{k+1}, X^{k+1})\|} \leq 10^{-6}.$$

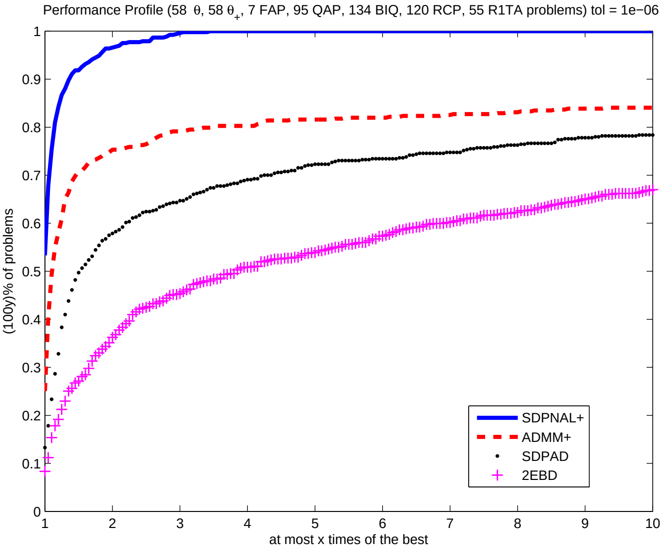
Performance of our SDPNAL+ and ADMM+ versus
SDPAD: the directly extended ADMM implemented in [Wen et al.]
2EBD-HPE [Monteiro et al.]

Numbers of problems which are solved to the accuracy $\eta \leq 10^{-6}$

problem set (No.)	SDPNAL+	ADMM+	SDPAD	2EBD
θ (58)	58	56	53	53
θ_+ (58)	58	58	58	56
FAP (7)	7	7	7	7
QAP (95)	95	39	30	16
BIQ (134)	134	134	134	134
RCP (120)	120	120	114	109
R1TA (55)	55	42	47	18
Total (527)	527	456	443	393

Performance profile on 527 large SDPs

Performance profiles of SDPNAL+, ADMM+, SDPAD and 2EBD



Implemented the algorithms in MATLAB.
Runs perform on PC with (12 cores) Intel Xeon CPU E5-2680 @ 2.50 GHz and 128 GB RAM.

Stop SDPAD and 2EBD after 25000 iterations or 20 hours.

Prob	$m; n$	η SDPAD|2EBD|SDPNAL+	time (hour:minute)
1dc.2048	58368+$\mathcal{N}$; 2048	9.9-7| 9.9-7| 9.9-7	3:56| 2:10| 1:08
fap25	2118+$\mathcal{N}$; 2118	9.9-7| 9.9-7| 9.5-7	3:26| 0:54| 0:43
nug30	1393+$\mathcal{N}$; 900	1.1-5| 1.7-5| 9.6-7	2:10| 1:46| 0:09
tai30a	1393+$\mathcal{N}$; 900	4.6-6| 1.3-5| 9.9-7	2:34| 1:47| 0:10
nsym_rd[40,40,40]	672399; 1600	**1.5-3**| **2.0-3**| 8.6-7	2:48| 4:54| 0:04
nonsym(14,4)	1.16M; 2744	**1.4-2**| **5.2-3**| 1.3-7	7:39| 14:01| 0:20

Results show that it is essential to use **second-order information** and **second-order structured sparsity** to solve hard problems!

Summary and future work

- We have tested SDPNAL+ on about 520 SDPs from θ, θ_+, QAP, binary QP, rank-1 tensor approximation, etc
- When the problems are primal-dual nondegenerate, SDPNAL+ can efficiently solve large SDPs to high accuracy. SDPAD and 2EDB also performed well, though SDPNAL+ is often much more efficient.
- Many of the tested SDPs are degenerate, but SDPNAL+ can still solve them accurately with $\eta < 10^{-6}$. On the other hand, SDPAD and 2EDB were not able to solve many such problems.

Currently under development:

1. sparse SDPNAL+ so as to handle larger matrix variable when the data has conducive sparsity structure
2. a more advanced user-friendly interface

Thank you for your attention!