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 subproblems
- SDPNAL+: practical implementation of the 2 phase method
- Numerical experiments
$S^n_+ = \text{cone of positive semidefinite matrices.}$ Write $X \succeq 0$ if $X \in S^n_+$.

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

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

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

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

Important case: $\mathcal{N} = \{ X \in S^n \mid X \geq 0 \}$, 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 $X \in S^n_+ \cap \mathcal{N}$ to $X \in S^n_+$ and $Y \in \mathcal{N}$.

Now our solver SDPNAL+ can solve general SDP problems:

$$
\begin{align*}
\text{(genSDP)} \quad & \min \sum_{i=1}^{N} \langle C_i, X_i \rangle \\
\text{s.t.} \quad & \sum_{i=1}^{N} A_i(X_i) = b \quad \text{(equalities)} \\
\quad & l \leq \sum_{i=1}^{N} B_i(X_i) \leq u \quad \text{(inequalities)} \\
\quad & X_i \in K_i \quad \text{(cone)}, \quad X_i \in \mathcal{N}_i \quad \text{(bounds)}, \quad i = 1 : N
\end{align*}
$$

where $K_i$ is either a PSD cone or nonnegative orthant. Currently extending $K_i$ 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 ≡ 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

\[
\text{(NCM)} \quad \min_X \left\{ \frac{1}{2} \| H \circ (X - U) \|_1 \mid \text{Diag}(X) = 1, \; X \succeq 0 \right\}
\]

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:

\[
\min \left\{ \langle D, X \rangle \mid \langle I, X \rangle = k, \; X1 = 1, \; X \in \mathbb{S}^n_+, \; X \succeq 0 \right\}
\]

where \( D_{ij} = \| p_i - p_j \|_2^2 \).

Note: \( D \) can also be other affinity matrix.
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

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

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

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

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

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

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

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

$$\theta_+(G) := n(n+1)/2 \text{ additional constraints } X \succeq 0$$
Quadratic assignment problem (QAP)

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

\[ A = (a_{ij}) \quad \text{where } a_{ij} = \text{flow from facility } i \text{ to facility } j \]

\[ B = (b_{kl}) \quad \text{where } b_{kl} = \text{distance from location } k \text{ to location } l \]

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

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

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

(QAP) \[
\min \left\{ \langle B \otimes A, X \rangle \mid A(X) - b = 0, \ X \in \mathbb{S}^n_{++}, \ X \geq 0 \right\}
\]

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

\[
\begin{align*}
n &= 100; \\
G &= \text{randn}(n,n); \\
G &= 0.5*(G + G'); \\
\end{align*}
\]

```python
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 $G$. 

\[
\begin{align*}
n &= 200; \\
G &= \text{triu}(\text{sprand}(n,n,0.5),1); \\
[\text{IE}, \text{JE}] &= \text{find}(G); \\
n &= \text{length}(G);
\end{align*}
\]

\[
\begin{align*}
\text{model} &= \text{ccp\_model('theta')} \\
X &= \text{var\_sdp}(n,n); \\
\text{model.add\_variable}(X); \\
\text{model.maximize}(\text{sum}(X)); \\
\text{model.add\_affine\_constraint}(\text{trace}(X) == 1); \\
\text{model.add\_affine\_constraint}(X(\text{IE}, \text{JE}) == 0); \\
\text{model.add\_affine\_constraint}(X >= 0); \\
\text{model.solve};
\end{align*}
\]
\[
\begin{align*}
\text{min} & \quad \text{trace}(X^{(1)}) + \text{trace}(X^{(2)}) + \text{sum}(X^{(3)}) \\
\text{s.t.} & \quad -X^{(1)}_{12} + 2X^{(2)}_{33} + 2X^{(3)}_2 = 4 \quad \text{(equalities)} \\
& \quad 2X^{(1)}_{23} + X^{(2)}_{42} - X^{(3)}_4 = 3 \\
& \quad 2 \leq -X^{(1)}_{12} - 2X^{(2)}_{33} + 2X^{(3)}_2 \leq 7 \quad \text{(inequalities)} \\
& \quad X^{(1)} \in \mathbb{S}^6_+, \ X^{(2)} \in \mathbb{R}^{5\times5}_+, \ X^{(3)} \in \mathbb{R}^7_+ \quad \text{(cones)} \\
& \quad 0 \leq X^{(1)} \leq 10E_6, \ 0 \leq X^{(2)} \leq 8E_5 \quad \text{(bounds)}
\end{align*}
\]

\[n1 = 6; \ n2 = 5; \ n3 = 7;\]
M = ccp_model(’Example_simple’);
X1=var_sdp(n1,n1); \quad X2=var_nn(n2,n2); \quad 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:

\[
\text{(D)} \quad \min \left\{ -\langle b, y \rangle + \delta_{\mathbb{S}_{++}^n}(S) + \delta_\mathcal{N}(Z) \mid A^*y + S + Z = C \right\}
\]

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

\[
\mathcal{L}_\sigma(y, S, Z; X) = -\langle b, y \rangle + \delta_{\mathbb{S}_{++}^n}(S) + \delta_\mathcal{N}(Z)
+ \langle A^*y + S + Z - C, X \rangle + \frac{\sigma}{2} \|A^*y + S + Z - C\|^2
\]

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

**KKT conditions:**

\[
\mathcal{R}_{\text{KKT}}(y, S, Z; X) := \begin{pmatrix}
AX - b \\
S - \Pi_{\mathbb{S}_{++}^n}(S - X) \\
Z - \Pi_\mathcal{N}(Z - X) \\
A^*y + S + Z - C
\end{pmatrix} = \begin{pmatrix}
0 \\
0 \\
0 \\
0
\end{pmatrix}.
\]
A directly extended ADMM for dual SDP+

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

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

(1b) \(S^{k+1} = \arg\min_{S \in S_+^n} \mathcal{L}_\sigma(y^{k+1}, S, Z^k; X^k) = \Pi_{S_+^n}(\hat{C}^k - A^*y^{k+1} - Z^k)\)

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

(3) \(X^{k+1} = X^k + \tau\sigma(A^*y^{k+1} + S^{k+1} + Z^{k+1} - C)\), 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]
A convergent symmetric Gauss-Seidel (sGS) ADMM for dual SDP

But sGS-ADMM is guaranteed to converge!

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

\[
\begin{align*}
(1a) \quad & \hat{y}^{k+1} \approx \arg\min_{y \in \mathbb{R}^m} \mathcal{L}_\sigma(y, S^k, Z^k; X^k) \\
(1b) \quad & S^{k+1} = \arg\min_{S \in S_+^n} \mathcal{L}_\sigma(\hat{y}^{k+1}, S, Z^k; X^k) = \Pi_{S_+^n}(\hat{C}^k - A^*\hat{y}^{k+1} - Z^k) \\
(1c) \quad & y^{k+1} \approx \arg\min_{y \in \mathbb{R}^m} \mathcal{L}_\sigma(y, S^{k+1}, Z^k; X^k)
\end{align*}
\]

\[
\begin{align*}
(2) \quad & Z^{k+1} = \arg\min_{Z \in \mathcal{N}} \mathcal{L}_\sigma(y^{k+1}, S^{k+1}, Z; X^k) = \Pi_{\mathcal{N}}(\hat{C}^k - A^*y^{k+1} - S^{k+1}) \\
(3) \quad & X^{k+1} = X^k + \tau\sigma(A^*y^{k+1} + S^{k+1} + Z^{k+1} - C)
\end{align*}
\]

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

\[
\mathcal{L}_\sigma(y, S) \equiv \delta_{S_+^n}(S) + \frac{\sigma}{2}\|A^*y + S + Z^k + \hat{C}^k\|^2 - \langle b, y \rangle
\]

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

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

**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.


**Theorem** [Han-Sun-Zhang, MOR 2018: exact version]
Let $\Omega_{KKT} \neq \emptyset$ be the KKT solution set. Suppose that an error bound condition holds for $R_{KKT}$ 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.

$$
\text{dist}(u, \Omega_{KKT}) \leq \eta \|R_{KKT}(u)\| \quad \forall u \in B_r(u^*).
$$

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

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

Inexact version can be established via the analysis in [Chen-Sun-Toh, MP 2017] and [Han-Sun-Zhang, MOR 2018].
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. $p : \mathbb{R}^{n_1} \rightarrow (-\infty, +\infty]$, 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 $sGS(\mathcal{H}) := \mathcal{U}\mathcal{D}^{-1}\mathcal{U}^*$ (symmetric Gauss-Seidel decomp)

Given $\bar{x}$, define

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

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

**Backward GS: 3 → 2.** Compute

\[
x'_3 = \arg\min_{x_3} p(\bar{x}_1) + h(\bar{x}_1, \bar{x}_2, x_3) = \mathcal{H}^{-1}_{33}(b_3 - \mathcal{H}_{13}^*\bar{x}_1 - \mathcal{H}_{23}^*\bar{x}_2)
\]

\[
x'_2 = \arg\min_{x_2} p(\bar{x}_1) + h(\bar{x}_1, x_2, x'_3) = \mathcal{H}^{-1}_{22}(b_2 - \mathcal{H}_{12}^*\bar{x}_1 - \mathcal{H}_{23}\bar{x}'_3)
\]

**Forward GS: 1 → 2 → 3.** Compute

\[
x'^+_1 = \arg\min_{x_1} p(x_1) + h(x_1, x'_2, x'_3) \quad \text{(non-smooth/non-quadratic)}
\]

\[
x'^+_2 = \arg\min_{x_2} p(x'^+_1) + h(x'^+_1, x_2, x'_3) = \mathcal{H}^{-1}_{22}(b_2 - \mathcal{H}_{12}^*x'^+_1 - \mathcal{H}_{23}x'_3)
\]

\[
x'^+_3 = \arg\min_{x_3} p(x'^+_1) + h(x'^+_1, x'^+_2, x_3) = \mathcal{H}^{-1}_{33}(b_3 - \mathcal{H}_{13}^*x'^+_1 - \mathcal{H}_{23}^*x'^+_2)
\]

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

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

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

$$x^+_1 = \underset{x_1}{\arg\min} \left\{ p(x_1) + h(x_1, x'_{\geq 2}) - \langle e^+_1, x_1 \rangle \right\},$$

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

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

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

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

\[
(y^{k+1}, S^{k+1}, Z^{k+1}) \approx \arg\min \left\{ \mathcal{L}_k(y, S, Z) := \mathcal{L}_{\sigma_k}(y, S, Z; X^k) \right\}
\]

\[
= \arg\min \left\{ -\langle b, y \rangle + \frac{\sigma}{2} \| A^* y + S + Z + \hat{C}^k \|^2 + \delta_{S^+}(S) + \delta_N(Z) \right\}
\]

2. Update

\[
X^{k+1} = X^k + \sigma_k (A^* y^{k+1} + S^{k+1} + Z^{k+1} - C);
\]

update \( \sigma_{k+1} \uparrow \sigma_\infty \leq \infty \).
Global convergence of ALM

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

$$
e^{k+1} = \begin{bmatrix}
    AX^{k+1} - b \\ 
    X^{k+1} - \Pi_{S^n_+}(X^{k+1} - S^{k+1}) \\ 
    X^{k+1} - \Pi_{N^n}(X^{k+1} - Z^{k+1})
\end{bmatrix}.
$$

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

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

(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 $P^{\text{obj}}$ satisfies a quadratic growth condition at $X^*$, i.e., $\exists$ a neighborhood $U$ of $X^*$ and $\kappa > 0$ s.t.

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

Then for $k$ large, we have

$$\text{dist}(X^{k+1}, \Omega_P) \leq \theta_k \text{dist}(X^k, \Omega_P)$$

dual feasibility at $(y^{k+1}, S^{k+1}, Z^{k+1}) \leq \tau_k \text{dist}(X^k, \Omega_P)$

dual objective gap at $(y^{k+1}, S^{k+1}, Z^{k+1}) \leq \tau'_k \text{dist}(X^k, \Omega_P)$

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

Larger $\sigma_k$ gives faster convergence, but the inner problem is harder to solve.
For simplicity, assume $\mathcal{N} = S^n$ and hence the variable $Z$ is absent.

$$\arg\min_{y, S} \left\{ \mathcal{L}_\sigma(y, S) \equiv \delta_{S^n_+}(S) + \frac{\sigma}{2} \| A^* y + S - \hat{C}^k \|_2^2 - \langle b, y \rangle \right\}$$

$$\equiv \arg\min_y \left\{ \Phi^k(y) := -\langle b, y \rangle + \frac{\sigma}{2} \| \Pi_{S^n_+}(A^* y - \hat{C}^k) \|_2^2 \right\} \text{ (project out } S \text{)}$$

Optimality condition of unconstrained subproblem in $y$ is:

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

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

$\nabla \Phi^k(y)$ is not differentiable, but is strongly semismooth [Sun-Sun, 2002]. Thus SNCG is expected to have at least superlinear convergence.
Solve $\nabla \Phi^k(y) = -b + \sigma A \Pi_{S_n^+}(U) = 0$, \( U = A^*y - \hat{C}^k \).

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

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

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

\[
\Pi'_{S_n^+}(U)[M] = Q(\Omega \circ (Q^T MQ))Q^T
\]

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

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

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

\[
\text{cond}(\mathcal{H}) \leq \sigma \Theta(1) \text{cond}([AQ_\gamma \otimes Q_\gamma, AQ_\gamma \otimes Q_{\bar{\gamma}}])^2
\]
Exploiting second-order structured sparsity

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

$$Y := \Pi_{S_n^+}^\prime (U)[M] = Q(\Omega \circ (Q^T MQ))Q^T$$

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

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

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

If $r \approx n$, then use

$$Y = Q(E \circ (Q^T MQ))Q^T - Q(\overline{\Omega} \circ (Q^T MQ))Q^T = M - Q(\overline{\Omega} \circ (Q^T MQ))Q^T$$

where $\overline{\Omega} = E - \Omega$ has a similar structure as $\Omega$ 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:
\[(y^0, S^0, Z^0, X^0, \sigma_0) \leftarrow \text{ADMM+}(\bar{y}^0, \bar{S}^0, \bar{Z}^0, \bar{X}^0, \bar{\sigma}_0)\]

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 \(\sigma_{k+1}\)
   
   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+

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

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 \( \eta \leq 10^{-6} \)

<table>
<thead>
<tr>
<th>problem set (No.)</th>
<th>SDPNAL+</th>
<th>ADMM+</th>
<th>SDPAD</th>
<th>2EBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta ) (58)</td>
<td>58</td>
<td>56</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>( \theta_+ ) (58)</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>56</td>
</tr>
<tr>
<td>FAP (7)</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>QAP (95)</td>
<td>95</td>
<td>39</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>BIQ (134)</td>
<td>134</td>
<td>134</td>
<td>134</td>
<td>134</td>
</tr>
<tr>
<td>RCP (120)</td>
<td>120</td>
<td>120</td>
<td>114</td>
<td>109</td>
</tr>
<tr>
<td>R1TA (55)</td>
<td>55</td>
<td>42</td>
<td>47</td>
<td>18</td>
</tr>
<tr>
<td>Total (527)</td>
<td>527</td>
<td>456</td>
<td>443</td>
<td>393</td>
</tr>
</tbody>
</table>
Performance profile on 527 large SDPs

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

Performance Profile (58 $\theta$, 58 $\theta^+$, 7 FAP, 95 QAP, 134 BIQ, 120 RCP, 55 R1TA problems) tol = $1e^{-06}$

- SDPNAL+
- ADMM+
- SDPAD
- 2EBD

at most $x$ times of the best

(100$y$)% of problems

at most $x$ times of the best
Numerical results for SDPNAL+

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.

<table>
<thead>
<tr>
<th>Prob</th>
<th>$m; n$</th>
<th>$\eta$</th>
<th>time (hour:minute)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SDPAD</td>
<td>2EBD</td>
</tr>
<tr>
<td>1dc.2048</td>
<td>58368+$\mathcal{N}$; 2048</td>
<td>9.9-7</td>
<td>9.9-7</td>
</tr>
<tr>
<td>fap25</td>
<td>2118+$\mathcal{N}$; 2118</td>
<td>9.9-7</td>
<td>9.9-7</td>
</tr>
<tr>
<td>nug30</td>
<td>1393+$\mathcal{N}$; 900</td>
<td>1.1-5</td>
<td>1.7-5</td>
</tr>
<tr>
<td>tai30a</td>
<td>1393+$\mathcal{N}$; 900</td>
<td>4.6-6</td>
<td>1.3-5</td>
</tr>
<tr>
<td>nsym_rd[40,40,40]</td>
<td>672399; 1600</td>
<td>1.5-3</td>
<td>2.0-3</td>
</tr>
<tr>
<td>nonsym(14,4)</td>
<td>1.16M; 2744</td>
<td>1.4-2</td>
<td>5.2-3</td>
</tr>
</tbody>
</table>

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 $\theta$, $\theta_+$, 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!