A 2-phase augmented Lagrangian approach for large scale matrix optimization

Defeng Sun

Department of Mathematics, National University of Singapore

September 5, 2014 (Presentation at 2014 Workshop on Optimization for Modern Computation, Beijing University)

Joint work with: Kim-Chuan Toh, National University of Singapore

Students/postdocs: Caihua Chen (Nanjing), Junfeng Yang (Nanjing), Chao Ding (CAS), Kaifeng Jiang (DBS), Yongjin Liu (Shengyang Aerospace U), Chengjing Wang (Southwest Jiaotong U), Liuqin Yang (NUS), Xudong Li (NUS), Xinyuan Zhao (Beijing U Tech.)

- Matrix optimization problem (MOP)
- Examples: linear semidefinite programming (SDP), etc
- General framework of proximal-point algorithm (PPA)
- 2-phase PPA applied to SDP and SDP+ (matrix variable is positive semidefinite and nonnegative)
- A majorized semismooth Newton-CG (SNCG) method for solving PPA subproblems
- SDPNAL+: practical implementation of PPA for SDP+
- Numerical experiments

$\mathcal{X} = \mathbb{R}^{p\times n}$ or $\mathcal{S}^n$ ($n \times n$ symmetric matrices) endowed trace inner product $\langle \cdot, \cdot \rangle$ and Frobenius norm $\| \cdot \|$

$$\text{(MOP)} \quad \min \Big\{ f(X) \mid \mathcal{A}(X) - b \in \mathcal{Q},\ X \in \mathcal{X} \Big\}$$

$f : \mathcal{X} \to (-\infty, \infty]$ is a proper closed convex function

$\mathcal{Q}$ is a closed convex cone in $\mathbb{R}^m$

$b \in \mathbb{R}^m$

$\mathcal{A} : \mathcal{X} \to \mathbb{R}^m$ is a given (onto) linear map, e.g., $\mathcal{A}(X) = \mathrm{diag}(X)$

Define $\mathcal{A}^* =$ the adjoint of $\mathcal{A}$

Define the dual cone $\mathcal{Q}^* = \{X \in \mathcal{X} \mid \langle Y, X \rangle \geq 0\ \forall\, Y \in \mathcal{Q}\}$.

Define

$$\text{(indicator function)} \quad \delta_{\mathcal{Q}}(X) = \begin{cases} 0 & \text{if } X \in \mathcal{Q} \\ \infty & \text{otherwise} \end{cases}$$

Define

$$
\begin{aligned}
\text{(conjugate function)} \quad f^*(Z) &= \sup_{X \in \mathcal{X}} \{\langle Z, X\rangle - f(X)\} \\
\text{(subdifferential)} \quad \partial f(X) &= \mathrm{conv}\{\text{subgradients of } f \text{ at } X\}
\end{aligned}
$$

The dual problem of (MOP) is given by

$$
\max_{y \in \mathcal{Q}^*} \langle b, y\rangle - f^*(\mathcal{A}^* y)
$$

The KKT conditions for (MOP) are:

$$
\mathcal{A}X - b \in \mathcal{Q}, \quad y \in \mathcal{Q}^*, \quad \mathcal{A}^* y \in \partial f(X)
$$

MOP covers many important classes of problems

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

MOP includes linear semidefinite programming (SDP):

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

$$= \min\left\{ f(X) := \langle C,\, X\rangle + \delta_{\mathcal{S}^n_+}(X) \mid \mathcal{A}(X) - b \in \mathcal{Q} := \{0\}^m \right\}$$

$$\delta_{\mathcal{S}^n_+}(X) = \text{indicator function of } \mathcal{S}^n_+ = \begin{cases} 0 & \text{if } X \in \mathcal{S}^n_+ \\ \infty & \text{otherwise} \end{cases}$$

SDP is solvable by powerful interior-point methods if n and m are not too large, say, $n \leq 2,000$, $m \leq 10,000$.

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

SDP and MOP have lots of Applications

SDP (and more generally MOP) is a powerful modelling tool! Applications are growing rapidly, and driving developments in algorithms and software.

- LMI in control
- Combinatorial optimization
- Robust optimization: project management, revenue management
- Polynomial optimization: option pricing, queueing systems
- Moment problems, applied probability
- Engineering: Signal processing, communication, structural optimization, computer vision
- Statistics/Finance: correlation/covariance matrix estimation
- Machine learning: kernel estimation, dimensionality reduction/manifold unfolding,
- Euclidean metric embedding: sensor network localization, molecular conformation
- Quantum chemistry, quantum information
- Many others ...

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:

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

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

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

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 $X \geq 0$

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 $X \in \mathcal{S}_+^{n^2}$ and $X \geq 0$

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

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

Relaxations of rank-1 tensor approximations

Consider symmetric 4-tensor [Nie, Lasserre, Lim, De Lathauwer et al]:

$$\mathbf{f}(x) = \sum_{1 \le i,j,k,l \le n} F_{ijkl}\, x_i x_j x_k x_l \;\rightarrow F \approx \lambda(u \otimes u \otimes u \otimes u)$$

for some scalar λ and $u \in \mathbb{R}^n$ with $\|u\| = 1$.
Need to solve: $\max_{x \in \mathbb{R}^n} \{\pm \mathbf{f}(x) \mid \mathbf{g}(x) := x_1^2 + \cdots + x_n^2 = 1\}$. Let

$$\begin{aligned}
[x]_d &= \text{monomial vector of degree at most } d \\
[x]_d [x]_d^T &= \sum_{|\alpha| \le 2d} A_\alpha x^\alpha \;\Rightarrow\; M_d(y) := \sum_\alpha A_\alpha y_\alpha \\
\mathbf{f}(x) &= \sum f_\alpha x^\alpha \;\Rightarrow\; \langle f,\, y\rangle \\
\mathbf{g}(x) &= \sum g_\alpha x^\alpha \;\Rightarrow\; \langle g,\, y\rangle
\end{aligned}$$

SDP relaxation is given by:

$$\max\{\langle f,\, y\rangle \mid \langle g,\, y\rangle = 1,\; M_d(y) \succeq 0\}$$

Relaxation is tight if rank$(M_d(y^*))$=1.

Given sparse and noisy distance data $\{d_{ij} \mid (i,j) \in \mathcal{E}\}$ for n atoms, find coordinates $v_1, \ldots, v_n$ in $\mathbb{R}^3$ such that $\|v_i - v_j\| \approx d_{ij}$. Typically $\mathcal{E}$ consists of 20–50% of all pairs of atoms which are ≤ 6Å apart. Consider the model:

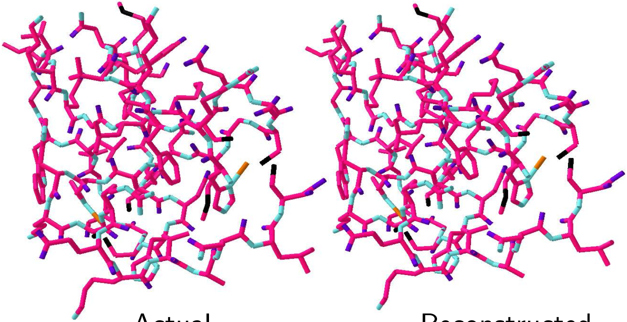
$$\min \left\{ \sum_{(ij)\in\mathcal{E}} \left| \|v_i - v_j\|^2 - d_{ij}^2 \right| \;\mid\; v_1, \ldots, v_n \in \mathbb{R}^3 \right\}$$

Let $V = [v_1, \ldots, v_n]$ and $X = V^T V$. Relaxing $X = V^T V$ to $X \succeq 0$ lead to an SDP:

$$\min_X \left\{ \sum_{(i,j)\in\mathcal{E}} |\langle A_{ij}, X\rangle - d_{ij}^2| \;:\; \langle E, X\rangle = 0, X \succeq 0 \right\}$$

where $A_{ij} = e_i e_i^T + e_j e_j^T - e_i e_j^T - e_j e_i^T$

Protein molecule `1PTQ` from Protein Data Bank:
number of atoms $n = 402$
number of pairwise distances given $|\mathcal{E}| \approx 3700$ (50% of distances $\leq 6\text{Å} \approx$ 4.5% of all pairwise distances)



Actual Reconstructed

Nuclear norm minimization problem

Given a partially observed matrix of $M \in \mathbb{R}^{n\times n}$, find a min-rank matrix $Y \in \mathbb{R}^{n\times n}$ to complete M:

$$\min_{Y\in\mathbb{R}^{n\times n}} \left\{ \text{rank}(Y) \mid Y_{ij} = M_{ij} \quad \forall\, (i,j) \in \mathcal{E} \right\} \quad \text{(NP-hard)}$$

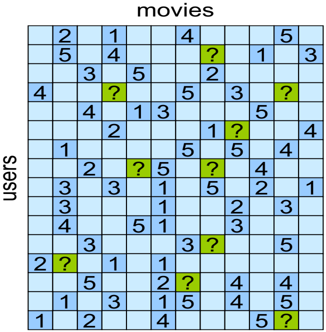
[Candes, Parrilo, Recht, Tao,...] For a given rank-r matrix $M \in \mathbb{R}^{n\times n}$ that satisfies certain properties, if enough entries $(\propto r\, n\, \text{polylog}(n))$ are sampled randomly, then with very high probability, M can be recovered from the following nuclear norm minimization problem:

$$\min_{Y\in\mathbb{R}^{n\times n}} \left\{ \|Y\|_* \mid Y_{ij} = M_{ij} \quad \forall\, (i,j) \in \mathcal{E} \right\}$$

easier problem, but still nontrivial to solve!

where $\|Y\|_* =$ sum of singular values of Y.

Based on partially observed matrix, predict unobserved entries: will customer i like movie j?



Sparse covariance selection problems

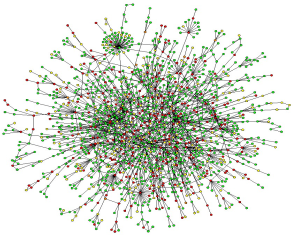
Given i.i.d. observations drawn from an n-dimensional Gaussian distribution $\mathcal{N}(x, \mu, \Sigma)$, let $\widehat{\Sigma}$ be the sample covariance matrix.

- Want to estimate Σ, whose inverse $X := \Sigma^{-1}$ is sparse.
- Dempster (1972) proved that x_i and x_j are conditionally independent (given all other x_k) if and only if $X_{ij} = 0$.

Typically, we estimate X via the log-likelihood function:

$$\max \left\{ \log\det X - \langle \widehat{\Sigma}, X \rangle - \langle W, |X| \rangle \mid X \succ 0 \right\}$$

where the weighted L_1-term is added to encourage sparsity in X. Many papers: d'Aspremont, M. Yuan, Lu, Meinshausen, Bühlmann, Wang-Sun-Toh, Yang-Sun-Toh



(MOP) also contains the important case of convex quadratic SDP:

$$\text{(QSDP)} \quad \min_{X \in \mathcal{S}^n} \left\{ \frac{1}{2}\langle X,\, \mathcal{Q}(X)\rangle + \langle C,\, X\rangle \;\middle|\; \mathcal{A}(X) - b = 0, X \in \mathcal{S}^n_+ \right\}$$

$\mathcal{Q} : \mathcal{S}^n \to \mathcal{S}^n$ is a self-adjoint positive semidefinite linear operator.

A well-studied example is the nearest correlation matrix problem, where given data matrix $U \in \mathcal{S}^n$ and weight matrix $W \succ 0$, we want to solve the W-weighted NCM problem:

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

1. The alternating projection method [Higham 02]
2. The quasi-Newton method [Malick 04]
3. An inexact semismooth Newton-CG method [Qi and Sun 06]
4. An inexact interior-point method [Toh, Tütüncü and Todd 07]

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

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

1. An inexact IPM for convex QSDP [Toh 08]
2. An ALM [Qi and Sun 10]
3. A semismooth Newton-CG ALM for convex quadratic programming over symmetric cones [Zhao 09]
4. A modified alternating direction method for convex quadratically constrained QSDPs [J. Sun and Zhang 10]

Consider a proper closed convex function f. Given $\beta > 0$, the Moreau-Yosida (MY) regularization of f is defined by

$$F_\beta(X) := \min_Y f(Y) + \frac{1}{2\beta} \|Y - X\|^2$$

Denote the unique minimizer by $\mathcal{P}_\beta(X)$ (known as the proximal mapping of f).

F_β is continuously differentiable and convex:

$$\nabla F_\beta(X) = \tfrac{1}{\beta}\,(X - \mathcal{P}_\beta(X))$$

$$\|\mathcal{P}_\beta(X) - \mathcal{P}_\beta(Y)\| \leq \|X - Y\| \;\forall\, X,\, Y$$

$$\min f(X) \Leftrightarrow \min F_\beta(X)$$

PPA is a gradient method to solve $\min F_\beta(X)$:

$$X^{k+1} \approx X^k - \beta_k \nabla F_{\beta_k}(X^k) = \mathcal{P}_{\beta_k}(X^k)$$

Given f such that $F_\beta(X), \mathcal{P}_\beta(X)$ can be computed analytically (or easily) for any X. Consider the basic problem:

$$\text{(MOP)} \quad \min\big\{ f(X) \mid \mathcal{A}(X) - b \in \mathcal{Q} \big\}.$$

The MY function at X^k is given by

$$\mathbf{F}_\beta^{\mathrm{MOP}}(X^k) = \min\Big\{ f(X) + \frac{1}{2\beta}\|X - X^k\|^2 \mid \mathcal{A}(X) - b \in \mathcal{Q} \Big\}$$

(by strong duality)

$$= \frac{1}{2\beta}\|X^k\|^2 + \max_{y \in \mathcal{Q}^*} \Big\{ \underbrace{\langle b,\, y\rangle - \frac{1}{2\beta}\|X^k + \beta\mathcal{A}^* y\|^2 + F_\beta(X^k + \beta\mathcal{A}^* y)}_{\Phi^k(y)} \Big\}$$

Optimality condition for max-subproblem is: $y = \Pi_{\mathcal{Q}^*}(y - \nabla\Phi^k(y))$,

$$\nabla\Phi^k(y) = b - \mathcal{A}\mathcal{P}_\beta(X^k + \beta\mathcal{A}^* y)$$

Linear SDP: $\min\left\{ f(X) \mid \mathcal{A}(X) - b \in \mathcal{Q} \right\}$

Here $\mathcal{Q} = \{0\}^m$, $\mathcal{Q}^* = \mathbb{R}^m$, $f(X) = \langle C,\, X\rangle + \delta_{\mathcal{S}^n_+}(X)$.

$$
\begin{aligned}
F_\beta(Y) &= \min\left\{ \langle C,\, X\rangle + \frac{1}{2\beta}\|X - Y\|^2 \mid X \in \mathcal{S}^n_+ \right\} \\
&= \frac{1}{2\beta}\|Y\|^2 - \frac{1}{2\beta}\|\Pi_{\mathcal{S}^n_+}(Y - \beta C)\|^2 \\
P_\beta(Y) &= \Pi_{\mathcal{S}^n_+}(Y - \beta C) \quad \text{(Projection of matrix onto } \mathcal{S}^n_+\text{)}.
\end{aligned}
$$

Hence the MY function at X^k is:

$$
\mathbf{F}^{\mathrm{MOP}}_\beta(X^k) = \frac{1}{2\beta}\|X^k\|^2 + \max_{y \in \mathbb{R}^m} \left\{ \Phi^k(y) := \langle b,\, y\rangle - \frac{1}{2\beta}\|P_\beta(X^k + \beta\mathcal{A}^* y)\|^2 \right\}
$$

Optimality condition of unconstrained max-subproblem is:

$$
\nabla\Phi^k(y) = b - \mathcal{A}\, P_\beta(X^k + \beta\mathcal{A}^* y) = 0.
$$

We solve it by a semismooth Newton-CG (SNCG) method.

A semismooth Newton-CG method for max-subproblem

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

$\nabla\Phi^k(y)$ is not differentiable, but is strongly semismooth [Sun^2, 2002]. At the current iteration, y_l, we solve a generalized Newton equation:

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

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

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

where $\Omega_{ij} = (d_i^+ - d_j^+)/(d_i - d_j)$. For $\gamma = \{1, \ldots, r\}$ and $\bar{\gamma} = \{r+1, \ldots, n\}$, we have

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

The structure in Ω allows for efficient computation of rhs of (2), and hence matrix-vector multiply for CG in solving (1)

Nuclear norm minimization: $\min\{f(X) \mid \mathcal{A}(X) - b \in \mathcal{Q}\}$

Here $\mathcal{Q} = \{0\}^m$, $\mathcal{Q}^* = \mathbb{R}^m$, $f(X) = \|X\|_*$.

Given any X, let its SVD be $X = U\mathrm{Diag}(\sigma)V^T$. Then

$$F_\beta(X) = \min_Y \left\{ \|Y\|_* + \frac{1}{2\beta}\|Y - X\|^2 \right\} \text{ (computable via SVD of } X\text{)}$$

$$P_\beta(X) = U\mathrm{Diag}(\max\{\sigma - \beta, 0\})V^T.$$

The MY function is:

$$\mathbf{F}_\beta^{\mathrm{MOP}}(X^k) = \frac{1}{2\beta}\|X^k\|^2 + \max_{y \in Q^*} \left\{ \langle b,\, y\rangle - \frac{1}{2\beta}\|P_\beta(X^k + \beta\mathcal{A}^* y)\|^2 \right\}$$

Optimality condition for unconstrained max-subproblem is:

$$\nabla\Phi^k(y) = b - \mathcal{A}\, P_\beta(X^k + \beta\mathcal{A}^* y) = 0.$$

We solve it by a semismooth Newton-CG (SNCG) method.

For the max-subproblem, with appropriate stopping conditions including the one below:

$$\mathrm{dist}(0, \partial\Phi^k(y^{k+1})) \leq (\varepsilon_k/\beta_k)\|X^{k+1} - X^k\|, \quad \varepsilon_k \to 0,$$

then we get the following theorem based on [Rockafellar, 1976].

Theorem: Suppose primal and dual MOPs are strictly feasible, and constraint nondegeneracy hold at the optimal solution X^*, y^*. Then $\{X^k\}$, $\{y^k\}$ converge to X^*, y^*. Moreover, there exist constants θ, θ' such that for k large, we have

$$\begin{aligned} \|X^{k+1} - X^*\| &\leq \frac{\theta}{\sqrt{\theta^2 + \beta_{\max}^2}}\|X^k - X^*\| \\ \|y^{k+1} - y^*\| &\leq \frac{\theta'}{\beta_{\max}}\|X^k - X^*\|. \end{aligned}$$

Larger $\beta_{\max}$ leads to faster LINEAR convergence, but inner problem is harder to solve [we only need a decent fast linear rate, e.g., 0.95].

Large scale SDP and SDP+: a brief history

Number of constraints m is large: $m \geq 10,000 \Rightarrow m \times m$ dense "Hessian" matrix cannot be stored explicitly. For $m = 10^5$, needs 100GB RAM memory.

- Parallel implementation of IPM [Benson, Borchers, Fujisawa, ... 03-present]
- First-order gradient methods on NLP reformulation (low accuracy) [Burer-Monteiro 03]
- Inexact IPM [Kojima, Toh 04]
- Generalized Lagrangian method on shifted barrier-penalized dual [Kocvara-Stingl 03]
- ALM on primal SDP from relaxation of lift-and-project scheme [Burer-Vandenbussche 06]
- Boundary-point method: based on ALM on dual – ADMM with unit step-length [Rendl et al. 06]
- SDPNAL: SNCG ALM on dual [Zhao-Sun-Toh 10]
- SDPAD: ADMM on dual with steplength 1.618 [Wen et al. 10]
- 2EDB: hybrid proximal extra-gradient method on primal [Monteiro et al. 13]
- SDPNAL+: SNCG PPA on primal for SDP+ [Yang-Sun-Toh 14]

Define $\mathcal{N} = \{X \in \mathcal{S}^n \mid X \geq 0\}$ (cone of nonnegative matrices)

$$
\begin{array}{lrl}
\text{(SDP+)} & & \min\left\{\langle C,\, X\rangle \mid \mathcal{A}(X) = b,\ X \in \mathcal{S}^n_+,\ X \in \mathcal{N}\right\} \\
& = & \min\left\{\langle C,\, X\rangle + \delta_{\mathcal{N}}(X) \,\Big|\, \begin{bmatrix} \mathcal{A} \\ I \end{bmatrix} X - \begin{bmatrix} b \\ 0 \end{bmatrix} \in \mathcal{Q} := \{0\}^m \times \mathcal{S}^n_+\right\}
\end{array}
$$

Here $\mathcal{Q}^* = \mathbb{R}^m \times \mathcal{S}^n_+$, $f(X) = \langle C,\, X\rangle + \delta_{\mathcal{N}}(X)$.

$$
\begin{aligned}
F_\beta(Y) &= \frac{1}{2\beta}\|Y\|^2 - \frac{1}{2\beta}\|\Pi_{\mathcal{N}}(Y - \beta C)\|^2 \\
P_\beta(Y) &= \Pi_{\mathcal{N}}(Y - \beta C) \quad \text{(Projection onto } \mathcal{N}\text{)}
\end{aligned}
$$

Hence the MY function at X^k is:

$$
\begin{aligned}
\mathbf{F}^{\mathrm{MOP}}_\beta(X^k) = {} & \frac{1}{2\beta}\|X^k\|^2 \\
& + \max_{y \in \mathbb{R}^m,\, S \in \mathcal{S}^n_+} \left\{\langle b,\, y\rangle - \frac{1}{2\beta}\|\Pi_{\mathcal{N}}(X^k + \beta(\mathcal{A}^* y + S - C))\|^2\right\}
\end{aligned}
$$

A majorized SNCG method for subproblem: $\mathbf{F}_\beta^{\mathrm{MOP}}(X^k)$

Let $C^k = C - \beta^{-1} X^k$, and

$$\Phi^k(y, S) \;=\; -\langle b,\, y\rangle + \frac{\beta}{2}\|\Pi_{\mathcal{N}}(\mathcal{A}^* y + S - C^k)\|^2$$

At a given $(\widehat{y}, \widehat{S})$, we have the quadratic majorization:

$$\Phi^k(y, S) \;\;\le\;\; \Phi^k(\widehat{y}, \widehat{S}) - \langle b,\, y\rangle + \frac{\beta}{2}\|\mathcal{A}^* y + S - C^k + \widehat{Z}^k\|^2,$$

where $\widehat{Z}^k = \Pi_{\mathcal{N}}(C^k - \mathcal{A}^*\widehat{y} - \widehat{S})$.

Solve $\min\{\Phi^k(y, S) \mid y \in \mathbb{R}^m, S \in \mathcal{S}_+^n\}$ via majorized SNCG method:

Input $(y_0, S_0) = (y^k, S^k)$. For $l = 0, 1, \ldots,$

Compute $\widehat{Z}_l^k = \Pi_{\mathcal{N}}(C^k - \mathcal{A}^* y_l - S_l)$, solve

$(y_{l+1}, S_{l+1}) \approx \mathrm{argmin}_{y\in\mathbb{R}^m, S\in\mathcal{S}_+^n}\Big\{-\langle b,\, y\rangle + \frac{\beta}{2}\|\mathcal{A}^* y + S - C^k + \widehat{Z}_l^k\|^2\Big\}$

Output (y^{k+1}, S^{k+1})

Cheaper than majorized SNCG if only low accuracy is required.
Dual of SDP+ and its augmented Lagrangian function are given by:

$$\min\{-\langle b,\, y\rangle \mid \mathcal{A}^* y + S + Z = C,\ S \in \mathcal{S}^n_+,\ Z \in \mathcal{N}\}$$

$$\mathcal{L}_\sigma(y, S, Z; X) = -\langle b,\, y\rangle + \frac{\sigma}{2}\|\mathcal{A}^* y + S + Z + \sigma^{-1} X - C\|^2 - \frac{1}{2\sigma}\|X\|^2$$

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

(a) $Z_{l+1} = \mathrm{argmin}_{Z\in\mathcal{N}}\{\mathcal{L}_\sigma(y_l, S_l, Z; X_l)\} = \Pi_{\mathcal{N}}(\widehat{C}_l - \mathcal{A}^* y_l - S_l)$

(b) $\widehat{\mathbf{y}}_{l+1} = \mathrm{argmin}_{y\in\mathbb{R}^m}\{\mathcal{L}_\sigma(y, S_l, Z_{l+1}; X_l)\}$

(c) $S_{l+1} = \mathrm{argmin}_{S\in\mathcal{S}^n_+}\{\mathcal{L}_\sigma(\widehat{\mathbf{y}}_{l+1}, S, Z_{l+1}; X_l)\} = \Pi_{\mathcal{S}^n_+}(\widehat{C}_l - \mathcal{A}^* \widehat{\mathbf{y}}_{l+1} - Z_{l+1})$

(d) $y_{l+1} = \mathrm{argmin}_{y\in\mathbb{R}^m}\{\mathcal{L}_\sigma(y, S_{l+1}, Z_{l+1}; X_l)\}$

(e) $X_{l+1} = X_l + \tau\sigma(\mathcal{A}^* y_{l+1} + S_{l+1} + Z_{l+1} - C)$ (e.g., $\tau = 1.618$).

ADMM+ is a convergent enhancement of the direct extension of ADMM, whose convergence is not guaranteed.

1. Generate a good starting point to warm-start PPA-SNCG:
   $(y^0, S^0, Z^0, X^0, \beta_0) \leftarrow$ ADMM+$(\bar{y}^0, \bar{S}^0, \bar{Z}^0, \bar{X}^0, \bar{\beta}_0)$
2. For $k = 0, 1, \ldots$

   Generate $(y^{k+1}, S^{k+1}, Z^{k+1}, \beta_{k+1})$ by majorized SNCG

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

   If progress of PPA-SNCG is slow,

   Rescale data

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

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

Robustness of SDPNAL+

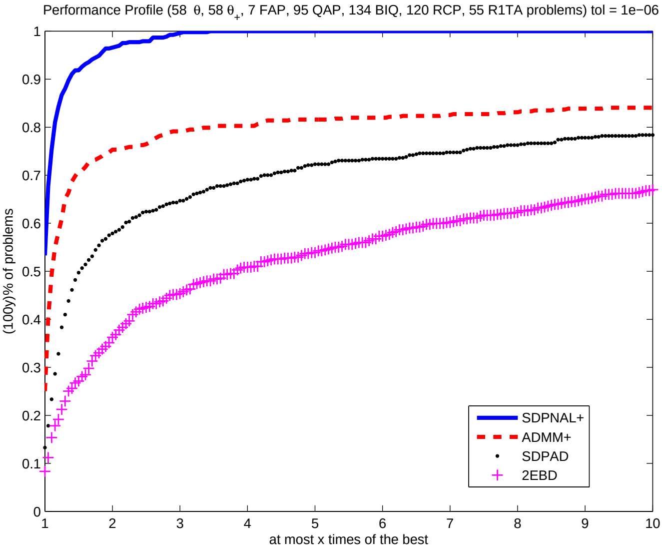
$$\eta := \max\left\{\begin{array}{c} R_P, R_D, R_{\mathcal{S}^n_+}(X), R_{\mathcal{N}}(X), R_{\mathcal{S}^n_+}(S), R_{\mathcal{N}}(Z), \\ R(\langle X,\ S\rangle), R(\langle X,\ Z\rangle) \end{array}\right\} \leq 10^{-6}.$$

We compare the performance of our SDPNAL+ and ADMM+ with the direct ADMM ($\tau = 1.618$) implemented in SDPAD [Wen et al.] and 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 profiles of SDPNAL+, ADMM+, SDPAD and 2EBD



Numerical results for SDPNAL+

Implemented the algorithms in MATLAB.
Runs perform on a 6 cores Linux Server with 12 Intel Xeon processors at 2.67 GHz and 32G RAM.

Stop SDPAD and 2EBD after 25000 iterations or 99 hours.

Prob	$m; n$	η SDPAD|2EDB|SDPNAL+	time (hour:minute)
1dc.2048	58368; 2048	9.9-7| 9.9-7| 9.9-7	14:00| 16:04| 5:50
fap36	4110+$\mathcal{N}$; 4110	9.9-7| 9.9-7| 9.5-7	78:43| 43:37| 23:07
nug30	1393+$\mathcal{N}$; 900	1.1-5| 1.7-5| 9.6-7	4:58| 5:39| 0:45
tai30a	1393+$\mathcal{N}$; 900	4.6-6| 1.3-5| 9.9-7	6:09| 6:00| 0:29
nonsym(6,5)	194480; 1296	9.9-7| **1.6-3**| 5.2-7	2:59| 11:24| 0:05
nsym_rd[40,40,40]	672399; 1600	**3.7-4**| **5.1-4**| 8.6-7	13:56| 22:41| 0:14
nonsym(12,4)	12.32M; 9261	**9.8-3**| **5.2-3**| 5.7-8	99:00| 99:00| 14:22

Results show that it is essential to use second-order information, wisely, to solve hard and/or large problems!

Summary

- 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 relative high accuracy. SDPAD and 2EDB also performed well, though SDPNAL+ is often more efficient.
- Many of the tested SDPs are degenerate, but SDPNAL+ can still solve them accurately with $\eta < 10^{-6}$. Other hand, SDPAD and 2EDB were not able to solve many such problems.
- Many more challenging problems.

Thank you for your attention!