Block Symmetric Gauss-Seidel Iteration and Multi-Block Semidefinite Programming

Defeng Sun

Department of Applied Mathematics

The Hong Kong Polytechnic University
香港理工大學

School of Mathematics, Sichuan University
October 30, 2017

Semidefinite Programming

The standard semidefinite programming (SDP):

$$\min_{X \in \mathcal{S}^n} \{ \langle C, X \rangle \mid \mathcal{A}X = b,\ X \succeq 0 \}.$$

The dual problem in its equivalent minimization form:

$$\min_{y \in \mathbb{R}^m} \{ -\langle b, y \rangle + \delta_{\mathcal{S}^n_+}(S) \mid \mathcal{A}^* y + S = C \}.$$

The Lagrangian function of the dual problem:

$$\mathcal{L}(y, S; X) := -\langle b, y \rangle + \delta_{\mathcal{S}^n_+}(S) + \langle X, \mathcal{A}^* y + S - C \rangle.$$

The augmented Lagrangian function of the dual problem ($\sigma > 0$):

$$\mathcal{L}_\sigma(y, S; X) := \mathcal{L}(y, S; X) + \frac{\sigma}{2} \| \mathcal{A}^* y + S - C \|^2.$$

An inexact augmented Lagrangian method (ALM) framework was used in SDPNAL:

$$\left\{\begin{array}{ll} y^{k+1} \approx \underset{y\in\mathbb{R}^m}{\arg\min}\, \Phi_{\sigma_k}(y; X^k), & \\ X^{k+1} = \Pi_{\mathcal{S}^n_+}[X^k + \sigma(\mathcal{A}^* y^{k+1} - C)], & k = 0, 1, 2, \ldots, \\ \sigma_{k+1} = \rho\sigma_k \ \text{ or } \sigma_{k+1} = \sigma_k, & \end{array}\right.$$

where for a given X

$$\begin{aligned} \Phi_\sigma(y; X) : &= \min_{S\in\mathcal{S}^n_+} \mathcal{L}_\sigma(y, S; X) \\ &= -\langle b, y\rangle + \frac{1}{2\sigma}\left(\|\Pi_{\mathcal{S}^n_+}[X + \sigma(\mathcal{A}^* y - C)]\|^2 - \|X\|^2\right). \end{aligned}$$

$\Phi_{\sigma_k}(y; X^k)$ is continuously differentiable with respect to y and $\nabla_y \Phi_{\sigma_k}$ is strongly semismooth.

Newton-CG: y^{k+1} is computed via a semismooth Newton method in which each linear system is solved by a conjugate gradient method.

Pros and cons of ALM

ALM: fast local linear convergence (arbitrary linear convergence rate) when the penalty parameter exceeds a certain threshold. But

- Sometimes can be hard and expensive to solve the inner subproblems exactly or to high accuracy, especially in high-dimensional settings;
- Computationally, it is not economical to use the ALM during the early stage of solving the problem when the fast local linear convergence of ALM has not kicked in.

In SDPNAL, the boundary-point method of Rendl et al. [Computing, 78 (2006)] was used to warm-start the second-order method, i.e., one modified gradient step was used instead of solving the inner subproblem:

$$\begin{aligned} y^{k+1} &= y^k - (\sigma_k \mathcal{A}\mathcal{A}^*)^{-1} \nabla_y \Phi_\sigma(y; X^k) \\ &= y^k - (\sigma_k \mathcal{A}\mathcal{A}^*)^{-1} \nabla_y \mathcal{L}(y^k, \widetilde{X}^{k+1}), \end{aligned}$$

with $\widetilde{X}^{k+1} = \Pi_{\mathcal{S}^n_+}[X^k + \sigma(\mathcal{A}^* y^k - C)]$. One can deduce that

$$\sigma_k \mathcal{A}\mathcal{A}^* y^{k+1} = \sigma_k \mathcal{A}\mathcal{A}^* y^k - \mathcal{A}\left(\Pi_{\mathcal{S}^n_+}[X^k + \sigma(\mathcal{A}^* y^k - C)]\right) + b,$$

which implies that

$$-b + \mathcal{A}X^k + \sigma_k \mathcal{A}(\mathcal{A}^* y^{k+1} + S^{k+1} - C) = 0,$$

with $\sigma S^{k+1} := \Pi_{\mathcal{S}^n_+}[-(X^k + \sigma(\mathcal{A}^* y^k - C)]$. Therefore,

$$\begin{cases} S^{k+1} = \arg\min_S \mathcal{L}_{\sigma_k}(y^k, S; X^k), \\ y^{k+1} = \arg\min_y \mathcal{L}_{\sigma_k}(y, S^{k+1}; X^k). \end{cases}$$

The doubly nonnegative SDP

$$\min_{X \in \mathcal{S}^n} \{\langle C, X\rangle \mid \mathcal{A}_E X = b_E, \mathcal{A}_I X \geq b_I,\ X \succeq 0,\ X \geq 0\}.$$

The more general convex quadratic SDP

$$\min \left\{ \frac{1}{2}\langle X,\ \mathcal{Q}X\rangle + \langle C,\ X\rangle | \mathcal{A}_E X = b_E, \mathcal{A}_I X \geq b_I, X \in \mathcal{S}^n_+ \cap \mathcal{N} \right\}. \tag{1}$$

- $\mathcal{Q} : \mathcal{S}^n \to \mathcal{S}^n$: self-adjoint positive semidefinite;
- $\mathcal{A}_E : \mathcal{S}^n \to \mathbb{R}^{m_E}$ and $\mathcal{A}_I : \mathcal{S}^n \to \mathbb{R}^{m_I}$ are linear maps;
- $C \in \mathcal{S}^n$, $b_E \in \mathbb{R}^{m_E}$ and $b_I \in \mathbb{R}^{m_I}$ are given data;
- $\mathcal{N}$: a closed convex set (e.g. $\mathcal{N} = \{X \in \mathcal{S}^n \mid L \leq X \leq U\}$).

[Fazel, Pong, Sun and Tseng, SIMAX, 34 (2013)]:

- The introduction of the semiproximal ADMM (alternating direction methods of multipliers).

[Sun, Yang and Toh, SIOPT, 25 (2015)]:

- A convergent 3-block ADMM (ADMM3c) for doubly nonnegative SDP: only requires an inexpensive extra step per iteration but it is theoretically convergent and practically even faster.
- The precursor of the block symmetric Gauss Seidel iteration technique.

Fig. 1 from [Sun, Yang and Toh (2015)]

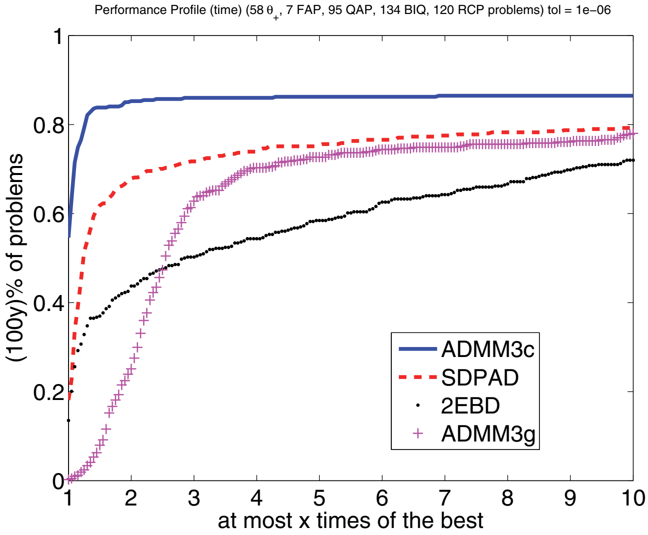

Figure: ADMM3c performs the best among a few first order methods (no inequality constraints).

Fig. 3 from [Sun, Yang and Toh (2015)]

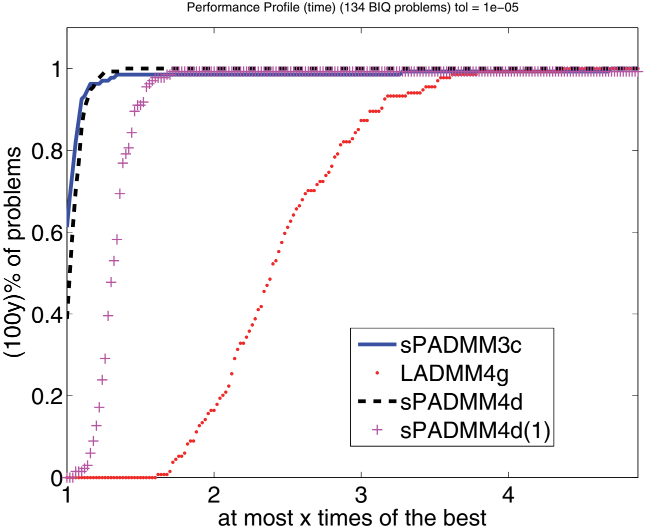

Figure: ADMM3c performs as good as the directly extended 4-block ADMM.

[Li, Sun and Toh, MP, 155 (2016)]:

- A Schur complement based (SCB) multi-block ADMM for convex quadratic conic programming;
- The block symmetric Gauss-Seidel (sGS) iteration technique.

[Li, Sun and Toh, arXiv:1512.08872 (2015)][1]

- The block sGS decomposition theorem;
- Its equivalence to the SCB reduction procedure;
- The quadratic part is not necessarily separable;
- Allows the updates of the blocks to be **inexact**.

[1] Currently available at arXiv:1703.06629 (2017)

Dual of the convex QSDP

The dual of the convex QSDP problem (1) in its equivalent minimization form:

$$\begin{aligned} \min \quad & \delta^*_{\mathcal{N}}(-Z) + \frac{1}{2}\langle W,\, \mathcal{Q}W\rangle - \langle b_E,\, y_E\rangle - \langle b_I,\, y_I\rangle \\ \text{s.t.} \quad & Z - \mathcal{Q}W + S + \mathcal{A}_E^* y_E + \mathcal{A}_I^* y_I = C, \\ & S \in \mathcal{S}^n_+,\ y_I \geq 0,\ W \in \mathcal{W}. \end{aligned} \tag{2}$$

$\mathcal{W}$ is an arbitrary subspace of $\mathcal{S}^n$ containing $\mathrm{Range}(\mathcal{Q})$

- Generally, $\mathcal{W}$ is $\mathcal{S}^n$ or $\mathrm{Range}(\mathcal{Q})$.
- For first-order methods, $\mathcal{W} = \mathcal{S}^n$.

Fig. 1 from [Li, Sun and Toh (2016)]

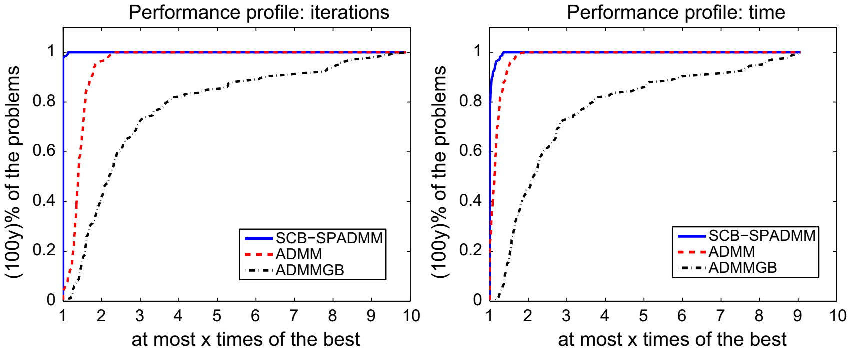

Fig. 1 Performance profiles of SCB- SPADMM, ADMM and ADMMGB for the tested large scale QSDP

Figure: SCB-ADMM performs the best for solving the tested QSDP problems (without inequality constraints).

An inexact block symmetric Gauss-Seidel (sGS) iteration

Let $s \geq 2$ be a given integer and $\mathcal{U} := \mathcal{U}_1 \times \cdots \times \mathcal{U}_s$ with all $\mathcal{U}_i$ being finite dimensional real Euclidean spaces. For any $u \in \mathcal{U}$ we write $u \equiv (u_1, \ldots, u_s)$. Let $\mathcal{H} : \mathcal{U} \to \mathcal{U}$ be a given self-adjoint positive semidefinite linear operator and

$$\mathcal{H}u := \begin{pmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} & \cdots & \mathcal{H}_{1s} \\ \mathcal{H}_{12}^* & \mathcal{H}_{22} & \cdots & \mathcal{H}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{H}_{1s}^* & \mathcal{H}_{2s}^* & \cdots & \mathcal{H}_{ss} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_s \end{pmatrix},$$

where $\mathcal{H}_{ii}$ are self-adjoint positive definite linear operators, $\mathcal{H}_{ij} : \mathcal{U}_j \to \mathcal{U}_i, i = 1, \ldots, s-1, j > i$, are linear maps. We denote

$$\mathcal{H}_u := \begin{pmatrix} 0 & \mathcal{H}_{12} & \cdots & \mathcal{H}_{1s} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \mathcal{H}_{(s-1)s} \\ & & & 0 \end{pmatrix}, \mathcal{H}_d := \begin{pmatrix} \mathcal{H}_{11} & & & \\ & \mathcal{H}_{22} & & \\ & & \ddots & \\ & & & \mathcal{H}_{ss} \end{pmatrix}. \tag{3}$$

Note that $\mathcal{H} = \mathcal{H}_d + \mathcal{H}_u + \mathcal{H}_u^*$ and $\mathcal{H}_d$ is positive definite.

Define the self-adjoint positive semidefinite linear operator $\mathrm{sGS}(\mathcal{H}) : \mathcal{U} \to \mathcal{U}$ by

$$\mathrm{sGS}(\mathcal{H}) := \mathcal{H}_u \mathcal{H}_d^{-1} \mathcal{H}_u^*.$$

For any $u \in \mathcal{U}$, denote

$$u_{\leq i} := \{u_1, \ldots, u_i\} \text{ and } u_{\geq i} := \{u_i, \ldots, u_s\}, i = 1, \ldots, s.$$

Let $\widetilde{\delta}_i, \delta_i \in \mathcal{U}_i, \quad i = 1, \ldots, s$ be given error tolerance vectors with $\widetilde{\delta}_1 = \delta_1$. Define

$$d(\widetilde{\delta}, \delta) := \delta + \mathcal{H}_u \mathcal{H}_d^{-1}(\delta - \widetilde{\delta}). \tag{4}$$

Let $\theta : \mathcal{U}_1 \to (-\infty, \infty]$ be a given closed proper convex function and $b \in \mathcal{U}$ be a given vector. Consider the quadratic function

$$h(u) := \frac{1}{2}\langle u, \mathcal{H}u\rangle - \langle b, u\rangle \quad \forall u \in \mathcal{U}.$$

Suppose that $u^- \in \mathcal{U}$ is a given vector. We want to compute

$$u^+ := \arg\min_{u \in \mathcal{U}} \left\{\theta(u_1) + h(u) + \frac{1}{2}\|u - u^-\|^2_{\mathrm{sGS}(\mathcal{H})} - \langle d(\widetilde{\delta}, \delta),\, u\rangle\right\}. \tag{5}$$

Proposition 1 (Inexact block sGS decomposition)

Assume that $\mathcal{H}_{ii}, i = 1, \ldots, s$ are positive definite. Then

$$\widehat{\mathcal{H}} := \mathcal{H} + \mathrm{sGS}(\mathcal{H}) = (\mathcal{H}_d + \mathcal{H}_u)\mathcal{H}_d^{-1}(\mathcal{H}_d + \mathcal{H}_u^*) \succ 0.$$

Furthermore, for $i = s, s-1, \ldots, 2$ (the backwark sGS sweep), define

$$\widetilde{u}_i \quad := \quad \arg\min_{u_i} \left\{ \theta(u_1^-) + h(u_{\leq i-1}^-, u_i, \widetilde{u}_{\geq i+1}) - \langle \widetilde{\delta}_i, u_i \rangle \right\}. \qquad (6)$$

Then, the optimal solution u^+ defined by (5) *can be obtained exactly via*

$$\begin{cases} u_1^+ & := \quad \arg\min_{u_1} \left\{ \theta(u_1) + h(u_1, \widetilde{u}_{\geq 2}) - \langle \delta_1, u_1 \rangle \right\}, \\ u_i^+ & := \quad \arg\min_{u_i} \left\{ \theta(u_1^+) + h(u_{\leq i-1}^+, u_i, \widetilde{u}_{\geq i+1}) - \langle \delta_i, u_i \rangle \right\}, \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 2, \ldots, s. \end{cases} \qquad (7)$$

Exact v.s. Inexact

- One should interpret $\widetilde{u}_i$ and u_i^+ as approximate solutions to the minimization problems without the terms involving $\widetilde{\delta}_i$ and δ_i.
- Once these approximate solutions have been computed, they would generate the error vectors $\widetilde{\delta}_i$ and δ_i.
- With these known error vectors, we know that $\widetilde{u}_i$ and u_i^+ are actually the exact solutions to the minimization problems in (6) and (7).

Highlight

- When solving the subproblems in the forward GS sweep in (7) for $i = 2, \ldots, s$, we may try to estimate u_i^+ by using $\widetilde{u}_i$, and in this case the corresponding error vector δ_i would be given by

$$\delta_i = \widetilde{\delta}_i + \sum_{j=1}^{i-1} \mathcal{H}_{ji}^* (u_j^+ - u_j^-).$$

  In order to avoid solving the i-th problem in (7), one may accept such an approximate solution $u_i^+ = \widetilde{u}_i$ if the corresponding error vector satisfies an admissible condition such as $\|\delta_i\| \leq c\|\widetilde{\delta}_i\|$ for some constant $c > 1$, say $c = 10$.

Error estimation in block sGS

Proposition 1 (Li-Sun-Toh)

Let $d(\widetilde{\delta}, \delta)$ be defined by (4). Then it holds that

$$\|\widehat{\mathcal{H}}^{-\frac{1}{2}} d(\widetilde{\delta}, \delta)\| \leq \|\mathcal{H}_d^{-\frac{1}{2}}(\delta - \widetilde{\delta})\| + \|\mathcal{H}_d^{\frac{1}{2}}(\mathcal{H}_d + \mathcal{H}_u)^{-1}\widetilde{\delta}\|. \tag{8}$$

Recall that

$$\mathcal{H} = \mathcal{H}_d + \mathcal{H}_u + \mathcal{H}_u^* \succeq 0,$$

$$\mathrm{sGS}(\mathcal{H}) := \mathcal{H}_u \mathcal{H}_d^{-1} \mathcal{H}_u^* \succeq 0,$$

$$\widehat{\mathcal{H}} := \mathcal{H} + \mathrm{sGS}(\mathcal{H}) = (\mathcal{H}_d + \mathcal{H}_u)\mathcal{H}_d^{-1}(\mathcal{H}_d + \mathcal{H}_u^*) \succ 0.$$

The block sGS decomposition theorem allows us to design a convergent (inexact) sGS-ADMM for solving convex multi-block composite programming problems including convex quadratic SDPs with doubly nonnegative constraints.

[Chen, Sun and Toh, MP, 161 (2017) 327–343]:
An inexact multi-block ADMM-type first-order method for solving high-dimensional multi-block convex composite optimization problems to medium accuracy with the essential flexibility that the inner subproblems are allowed to be solved only approximately, which is a combination of

- An inexact 2-block majorized semi-proximal ADMM
- Inexact block symmetric Gauss-Seidel iteration with a non-smooth block

- **Only one cycle** of an inexact sGS iteration instead of an unknown number of cycles, as the BCD-type methods.
- The freedom to solve large scale linear systems of equations approximately by an iterative solver such as the CG method.
- Without such a flexibility, one would be forced to modify the corresponding subproblem by adding an appropriately chosen "large" semi-proximal term so as to get a closed-form solution for the modified subproblem. But such a modification can sometimes significantly slow down the outer iteration.

High-Dimensional Composite Conic Optimization

$$\min_{x} \left\{ \theta(x) + \frac{1}{2}\langle x,\, \mathcal{Q}x \rangle + \langle c,\, x \rangle \mid \mathcal{A}x - b = 0,\, x \in \mathcal{K} \right\} \qquad (9)$$

- $\mathcal{X}$, $\mathcal{Y}$: finite-dimensional real Euclidean spaces endowed with inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$
- $\theta : \mathcal{X} \to (-\infty, +\infty]$: closed proper convex
- $\mathcal{Q} : \mathcal{X} \to \mathcal{X}$: self-adjoint positive semidefinite
- $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$: linear mapping
- $c \in \mathcal{X}$, $b \in \mathcal{Y}$ are given data, $\mathcal{K} \subseteq \mathcal{X}$: closed convex cone

"High-dimensional": $\mathcal{A}\mathcal{A}^*$ or $\mathcal{Q}$ is extremely large to be explicitly stored or decomposed by Cholesky factorization.

Example: QSDP, QP, Robust PCA......

One can recast (9) (by introducing a slack variables $u \in \mathcal{X}$) as

$$\min \left\{ \theta(u) + \frac{1}{2}\langle x,\, \mathcal{Q}x\rangle + \langle c,\, x\rangle \mid \mathcal{A}x - b = 0,\, u - x = 0,\, x \in \mathcal{K} \right\}. \tag{10}$$

Soving the dual of problems (9) is equivalent to

$$\begin{aligned} &\min\ \theta^*(-s) + \tfrac{1}{2}\langle w,\, \mathcal{Q}w\rangle - \langle b,\, \xi\rangle \\ &\text{s.t. } s + z - \mathcal{Q}w + \mathcal{A}^*\xi = c,\, z \in \mathcal{K}^*,\, w \in \mathcal{W}, \end{aligned} \tag{11}$$

$\mathcal{W} \subseteq \mathcal{X}$ is a subspace containing $\mathrm{Range}(\mathcal{Q})$, θ^* is the Fenchel conjugate of θ, $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$.

General Form of the Problem

Let m, n be two nonnegative integers, $\mathcal{Z}$, $\mathcal{X}_i$, $1 \le i \le m$ and $\mathcal{Y}_j$, $1 \le j \le n$ are finte dimensional real Euclidean spaces each endowed with $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. Define $\mathcal{X} := \mathcal{X}_1 \times \ldots \times \mathcal{X}_m$ and $\mathcal{Y} := \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_n$. Problem (11) belongs to

$$\begin{array}{ll} \min\limits_{x \in \mathcal{X}, y \in \mathcal{Y}} & p_1(x_1) + f(x_1, \ldots, x_m) + q_1(y_1) + g(y_1, \ldots, y_n) \\ \text{s.t.} & \mathcal{A}^* x + \mathcal{B}^* y = c. \end{array} \tag{12}$$

- $p_1 : \mathcal{X}_1 \to (-\infty, \infty]$ and $q_1 : \mathcal{Y}_1 \to (-\infty, \infty]$: closed proper convex;
- $f : \mathcal{X} \to (-\infty, \infty)$ and $g : \mathcal{Y} \to (-\infty, \infty)$: convex, continuously differentiable with Lipschitz continuous gradients;
- $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ and $\mathcal{B} : \mathcal{X} \to \mathcal{Z}$ are defined such that their adjoints are given by $\mathcal{A}^* x = \sum_{i=1}^m \mathcal{A}_i^* x_i$ for $x = (x_1, \ldots, x_m) \in \mathcal{X}$, and $\mathcal{B}^* y = \sum_{j=1}^n \mathcal{B}_j^* y_j$ for $y = (y_1, \ldots, y_n) \in \mathcal{Y}$ with $\mathcal{A}_i^* : \mathcal{X}_i \to \mathcal{Z}, i = 1, \ldots, m$ and $\mathcal{B}_j^* : \mathcal{Y}_j \to \mathcal{Z}, j = 1, \ldots, n$ are the adjoints of the linear maps $\mathcal{A}_i : \mathcal{Z} \to \mathcal{X}_i$ and $\mathcal{B}_j : \mathcal{Z} \to \mathcal{Y}_i$, respectively.

Majorized Augmented Lagrangian Function

Define for convenience $p(x) := p_1(x_1)$ and $q(y) := q_1(y_1)$.
There exist self-adjoint positive semidefinite linear operators $\widehat{\Sigma}_f : \mathcal{X} \to \mathcal{X}$ and $\widehat{\Sigma}_g : \mathcal{Y} \to \mathcal{Y}$, such that for any $x, x' \in \mathcal{X}$ and $y, y' \in \mathcal{Y}$,

$$
\begin{aligned}
f(x) &\leq \widehat{f}(x; x') := f(x') + \langle \nabla f(x'), x - x' \rangle + \tfrac{1}{2}\|x - x'\|^2_{\widehat{\Sigma}_f}, \\
g(y) &\leq \widehat{g}(y; y') := g(y') + \langle \nabla g(y'), y - y' \rangle + \tfrac{1}{2}\|y - y'\|^2_{\widehat{\Sigma}_g}.
\end{aligned} \tag{13}
$$

Let $\sigma > 0$. The majorized augmented Lagrangian function of problem (12) is defined by for any $(x', y') \in \mathcal{X} \times \mathcal{Y}$ and $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$,

$$
\begin{aligned}
\widehat{\mathcal{L}}_\sigma(x, y; (z, x', y')) := \; & p(x) + \widehat{f}(x; x') + q(y) + \widehat{g}(y; y') \\
& + \langle z, \mathcal{A}^* x + \mathcal{B}^* y - c \rangle + \tfrac{\sigma}{2}\|\mathcal{A}^* x + \mathcal{B}^* y - c\|^2.
\end{aligned}
$$

If f and g are quadratic functions, by taking $\widehat{\Sigma}_f = \Sigma_f$ and $\widehat{\Sigma}_g = \Sigma_g$ the majorized augmented Lagrangian function is also the augmented Lagrangian function.

An inexact majorized semi-Proximal ADMM

Let $\mathcal{S} : \mathcal{X} \to \mathcal{X}$ and $\mathcal{T} : \mathcal{Y} \to \mathcal{Y}$ being two self-adjoint positive semidefinite linear operators and define

$$\mathcal{M} := \widehat{\Sigma}_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \quad \text{and} \quad \mathcal{N} := \widehat{\Sigma}_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*. \tag{14}$$

Suppose that $\{w^k := (x^k, y^k, z^k)\}$ is a sequence in $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. For convenience, we define the two functions $\psi_k : \mathcal{X} \to (-\infty, \infty]$ and $\varphi_k : \mathcal{Y} \to (-\infty, \infty]$ by

$$\psi_k(x) := p(x) + \tfrac{1}{2}\langle x,\, \mathcal{M}x\rangle - \langle l_x^k,\, x\rangle,$$
$$\varphi_k(y) := q(y) + \tfrac{1}{2}\langle y,\, \mathcal{N}y\rangle - \langle l_y^k,\, y\rangle,$$

where

$$-l_x^k := \nabla f(x^k) + \mathcal{A}z^k - \mathcal{M}x^k + \sigma\mathcal{A}(\mathcal{A}^*x^k + \mathcal{B}^*y^k - c),$$
$$-l_y^k := \nabla g(y^k) + \mathcal{B}z^k - \mathcal{N}y^k + \sigma\mathcal{B}(\mathcal{A}^*x^{k+1} + \mathcal{B}^*y^k - c).$$

Let $\{\varepsilon_k\}$ be a summable sequence of nonnegative numbers, and define

$$\mathcal{E} := \sum_{k=0}^{\infty} \varepsilon_k < \infty, \quad \mathcal{E}' := \sum_{k=0}^{\infty} \varepsilon_k^2 < \infty.$$

An inexact 2-block Majorized sPADMM

Algorithm (**imsPADMM**)

Let $\tau \in (0, (1+\sqrt{5})/2)$ *be the step-length. Let* $w^0 := (x^0, y^0, z^0) \in \operatorname{dom} p \times \operatorname{dom} q \times \mathcal{Z}$ *be the initial point. For* $k = 0, 1, \ldots$ *. Choose* $\mathcal{S}$ *and* $\mathcal{T}$ *such that* $\mathcal{M} \succ 0$ *and* $\mathcal{N} \succ 0$.

1. *Compute* x^{k+1} *and* $d_x^k \in \partial \psi_k(x^{k+1})$ *s.t.* $\|\mathcal{M}^{-\frac{1}{2}} d_x^k\| \le \varepsilon_k$ *and*

$$x^{k+1} \approx \bar{x}^{k+1} := \arg\min_{x \in \mathcal{X}} \left\{ \psi_k(x) = \widehat{\mathcal{L}}_\sigma(x, y^k; w^k) + \frac{1}{2}\|x - x^k\|_{\mathcal{S}}^2 \right\}. \tag{15}$$

2. *Compute* y^{k+1} *and* $d_y^k \in \partial \varphi_k(y^{k+1})$ *s.t.* $\|\mathcal{N}^{-\frac{1}{2}} d_y^k\| \le \varepsilon_k$ *and*

$$\begin{aligned} y^{k+1} \approx \bar{y}^{k+1} &:= \arg\min_{y \in \mathcal{Y}} \left\{ \widehat{\mathcal{L}}_\sigma(\bar{x}^{k+1}, y; w^k) + \frac{1}{2}\|y - y^k\|_{\mathcal{T}}^2 \right\} \\ &= \arg\min_{y \in \mathcal{Y}} \left\{ \varphi_k(y) + \langle \sigma \mathcal{B}\mathcal{A}^*(\bar{x}^{k+1} - x^{k+1}),\ y \rangle \right\}. \end{aligned} \tag{16}$$

3. *Compute* $z^{k+1} := z^k + \tau\sigma(\mathcal{A}^* x^{k+1} + \mathcal{B}^* y^{k+1} - c)$.

In imsPADMM, the main issue is how to choose $\mathcal{S}$ and $\mathcal{T}$, and how to compute x^{k+1} and y^{k+1}.

Decomposition of $\widehat{\Sigma}_f$ and $\widehat{\Sigma}_g$, consistent with the decompositions of $\mathcal{X}$ and $\mathcal{Y}$:

$$\widehat{\Sigma}_f = \begin{pmatrix} (\widehat{\Sigma}_f)_{11} & (\widehat{\Sigma}_f)_{12} & \cdots & (\widehat{\Sigma}_f)_{1m} \\ (\widehat{\Sigma}_f)_{12}^* & (\widehat{\Sigma}_f)_{22} & \cdots & (\widehat{\Sigma}_f)_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ (\widehat{\Sigma}_f)_{1m}^* & (\widehat{\Sigma}_f)_{2m}^* & \cdots & (\widehat{\Sigma}_f)_{mm} \end{pmatrix},$$

$$\widehat{\Sigma}_g = \begin{pmatrix} (\widehat{\Sigma}_g)_{11} & (\widehat{\Sigma}_g)_{12} & \cdots & (\widehat{\Sigma}_g)_{1n} \\ (\widehat{\Sigma}_g)_{12}^* & (\widehat{\Sigma}_g)_{22} & \cdots & (\widehat{\Sigma}_g)_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (\widehat{\Sigma}_g)_{1n}^* & (\widehat{\Sigma}_g)_{2n}^* & \cdots & (\widehat{\Sigma}_g)_{nn} \end{pmatrix}.$$

Choose two self-adjoint positive semidefinite linear operators $\widetilde{\mathcal{S}}_1 : \mathcal{X}_1 \to \mathcal{X}_1$ and $\widetilde{\mathcal{T}}_1 : \mathcal{Y}_1 \to \mathcal{Y}_1$ satisfying

$$\widetilde{\mathcal{M}}_{11} := \widetilde{\mathcal{S}}_1 + (\widehat{\Sigma}_f)_{11} + \sigma \mathcal{A}_1 \mathcal{A}_1^* \succ 0, \; \widetilde{\mathcal{N}}_{11} := \widetilde{\mathcal{T}}_1 + (\widehat{\Sigma}_g)_{11} + \sigma \mathcal{B}_1 \mathcal{B}_1^* \succ 0,$$

for making the subproblems involving p_1 and q_1 easier to solve.

We can assume that the well-defined optimization problems

$$\min_{x_1} \{ p(x_1) + \frac{1}{2} \|x_1 - x_1'\|^2_{\widetilde{\mathcal{M}}_{11}} \} \text{ and } \min_{y_1} \{ q(y_1) + \frac{1}{2} \|y_1 - y_1'\|^2_{\widetilde{\mathcal{N}}_{11}} \}$$

can be solved to arbitrary accuracy for any given $x_1' \in \mathcal{X}_1$ and $y_1' \in \mathcal{Y}_1$.

For $i = 2, \ldots, m$, choose a linear operator $\widetilde{S}_i \succeq 0$ such that

$$\widetilde{\mathcal{M}}_{ii} := \widetilde{\mathcal{S}}_i + (\widehat{\Sigma}_f)_{ii} + \sigma \mathcal{A}_i \mathcal{A}_i^* \succ 0,$$

and similarly, for $j = 2, \ldots, n$, we choose a linear operator $\widetilde{\mathcal{T}}_j \succeq 0$ such that

$$\widetilde{\mathcal{N}}_{jj} := \widetilde{\mathcal{T}}_j + (\widehat{\Sigma}_g)_{jj} + \sigma \mathcal{B}_j \mathcal{B}_j^* \succ 0.$$

Algorithm (sGS-imsPADMM)

Choose $\tau \in (0, (1+\sqrt{5})/2)$. *Let* $\{\widetilde{\varepsilon}_k\}_{k\geq 0}$ *be a nonnegative summable sequence of real numbers. Let* $(x^0, y^0, z^0) \in \mathrm{dom}\, p \times \mathrm{dom}\, q \times \mathcal{Z}$ *be the initial point. For* $k = 0, 1, \ldots,$

1a. *for* $i = m, \ldots, 2$ *compute*

$$\widetilde{x}_i^{k+1} \approx \arg\min_{x_i \in \mathcal{X}_i} \Big\{ \widehat{\mathcal{L}}_\sigma(x^k_{\leq i-1}, x_i, \widetilde{x}^{k+1}_{\geq i+1}, y^k; w^k) + \frac{1}{2}\|x_i - x_i^k\|^2_{\widetilde{S}_i} \Big\},$$

$$\widetilde{\delta}_i^k \in \partial_{x_i} \widehat{\mathcal{L}}_\sigma(x^k_{\leq i-1}, \widetilde{x}_i^{k+1}, \widetilde{x}^{k+1}_{\geq i+1}, y^k; w^k) + \widetilde{\mathcal{S}}_i(\widetilde{x}_i^{k+1} - x_i^k), \ \ \|\widetilde{\delta}_i^k\| \leq \widetilde{\varepsilon}_k.$$

1b. *For* $i = 1, \ldots, m$ *compute*

$$x_i^{k+1} \approx \arg\min_{x_i \in \mathcal{X}_i} \Big\{ \widehat{\mathcal{L}}_\sigma(x^{k+1}_{\leq i-1}, x_i, \widetilde{x}^{k+1}_{\geq i+1}, y^k; w^k) + \frac{1}{2}\|x_i - x_i^k\|^2_{\widetilde{S}_i} \Big\},$$

$$\delta_i^k \in \partial_{x_i} \widehat{\mathcal{L}}_\sigma(x^{k+1}_{\leq i-1}, x_i^{k+1}, \widetilde{x}^{k+1}_{\geq i+1}, y^k; w^k) + \widetilde{\mathcal{S}}_i(x_i^{k+1} - x_i^k), \|\delta_i^k\| \leq \widetilde{\varepsilon}_k.$$

Algorithm (sGS-imsPADMM (continued))

2a. *For* $j = n, \ldots, 2$ *compute*

$$\widetilde{y}_j^{k+1} \approx \arg\min_{y_j \in \mathcal{Y}_j} \Big\{ \widehat{\mathcal{L}}_\sigma(x^{k+1}, y_{\leq j-1}^k, y_j, \widetilde{y}_{\geq j+1}^{k+1}; w^k) + \frac{1}{2}\|y_j - y_j^k\|_{\widetilde{\mathcal{T}}_j}^2 \Big\},$$

$$\widetilde{\gamma}_j^k \in \partial_{y_j} \widehat{\mathcal{L}}_\sigma(x^{k+1}, y_{\leq j-1}^k, \widetilde{y}_j^{k+1}, \widetilde{y}_{\geq j+1}^{k+1}; w^k) + \widetilde{\mathcal{T}}_j(\widetilde{y}_j^{k+1} - y_j^k), \|\widetilde{\gamma}_j^k\| \leq \widetilde{\varepsilon}_k.$$

2b. *For* $j = 1, \ldots, n$ *compute*

$$y_j^{k+1} \approx \arg\min_{y_j \in \mathcal{Y}_j} \Big\{ \widehat{\mathcal{L}}_\sigma(x^{k+1}, y_{\leq j-1}^{k+1}, y_j, \widetilde{y}_{\geq j+1}^{k+1}; w^k) + \frac{1}{2}\|y_j - y_j^k\|_{\widetilde{\mathcal{T}}_j}^2 \Big\},$$

$$\gamma_j^k \in \partial_{y_j} \widehat{\mathcal{L}}_\sigma(x^{k+1}, y_{\leq j-1}^{k+1}, y_j^{k+1}, \widetilde{y}_{\geq j+1}^{k+1}; w^k) + \widetilde{\mathcal{T}}_j(y_j^{k+1} - y_j^k), \|\gamma_j^k\| \leq \widetilde{\varepsilon}_k.$$

3. *Compute* $z^{k+1} := z^k + \tau\sigma(\mathcal{A}^* x^{k+1} + \mathcal{B}^* y^{k+1} - c)$.

Define the linear operators

$$\begin{aligned} \widetilde{\mathcal{M}} &:= \widehat{\Sigma}_f + \sigma \mathcal{A}\mathcal{A}^* + \mathrm{Diag}(\widetilde{\mathcal{S}}_1, \ldots, \widetilde{\mathcal{S}}_m), \\ \widetilde{\mathcal{N}} &:= \widehat{\Sigma}_g + \sigma \mathcal{B}\mathcal{B}^* + \mathrm{Diag}(\widetilde{\mathcal{T}}_1, \ldots, \widetilde{\mathcal{T}}_n). \end{aligned} \tag{17}$$

Define $\widetilde{\mathcal{M}}_u$ and $\widetilde{\mathcal{N}}_u$ analogously as $\mathcal{H}_u$ in (3) for $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{N}}$, and

$$\widetilde{\mathcal{M}}_d := \mathrm{Diag}(\widetilde{\mathcal{M}}_{11}, \ldots, \widetilde{\mathcal{M}}_{mm}), \quad \widetilde{\mathcal{N}}_d := \mathrm{Diag}(\widetilde{\mathcal{N}}_{11}, \ldots, \widetilde{\mathcal{N}}_{nn}).$$

Then, $\widetilde{\mathcal{M}} := \widetilde{\mathcal{M}}_d + \widetilde{\mathcal{M}}_u + \widetilde{\mathcal{M}}_u^*$ and $\widetilde{\mathcal{N}} := \widetilde{\mathcal{N}}_d + \widetilde{\mathcal{N}}_u + \widetilde{\mathcal{N}}_u^*$.

Moreover, we define the following linear operators:

$$\text{sGS}(\widetilde{\mathcal{M}}) := \widetilde{\mathcal{M}}_u \widetilde{\mathcal{M}}_d^{-1} \widetilde{\mathcal{M}}_u^*, \qquad \text{sGS}(\widetilde{\mathcal{N}}) := \widetilde{\mathcal{N}}_u \widetilde{\mathcal{N}}_d^{-1} \widetilde{\mathcal{N}}_u^*.$$
$$\widehat{\mathcal{S}} := \text{Diag}(\widetilde{\mathcal{S}}_1, \ldots, \widetilde{\mathcal{S}}_m) + \text{sGS}(\widetilde{\mathcal{M}}), \quad \widehat{\mathcal{M}} := \widehat{\Sigma}_f + \sigma \mathcal{A}\mathcal{A}^* + \widehat{\mathcal{S}},$$
$$\widehat{\mathcal{T}} := \text{Diag}(\widetilde{\mathcal{T}}_1, \ldots, \widetilde{\mathcal{T}}_n) + \text{sGS}(\widetilde{\mathcal{N}}) \quad \text{and} \quad \widehat{\mathcal{N}} := \widehat{\Sigma}_g + \sigma \mathcal{B}\mathcal{B}^* + \widehat{\mathcal{T}}.$$

Define the two constants

$$\begin{aligned} \kappa &:= 2\sqrt{m-1}\|\widetilde{\mathcal{M}}_d^{-\frac{1}{2}}\| + \sqrt{m}\|\widetilde{\mathcal{M}}_d^{\frac{1}{2}}(\widetilde{\mathcal{M}}_d + \widetilde{\mathcal{M}}_u)^{-1}\|, \\ \kappa' &:= 2\sqrt{n-1}\|\widetilde{\mathcal{N}}_d^{-\frac{1}{2}}\| + \sqrt{n}\|\widetilde{\mathcal{N}}_d^{\frac{1}{2}}(\widetilde{\mathcal{N}}_d + \widetilde{\mathcal{N}}_u)^{-1}\|. \end{aligned} \tag{18}$$

For any $k \geq 0$, and $\widetilde{\delta}^k = (\widetilde{\delta}_1^k, \ldots, \widetilde{\delta}_m^k)$, $\delta^k = (\delta_1^k, \ldots, \delta_m^k)$, $\widetilde{\gamma}^k = (\widetilde{\gamma}_1^k, \ldots, \widetilde{\gamma}_n^k)$ and $\gamma^k = (\gamma_1^k, \ldots, \gamma_n^k)$ such that $\widetilde{\delta}_1^{k+1} := \delta_1^{k+1}$ and $\widetilde{\gamma}_1^{k+1} := \gamma_1^{k+1}$, we define

$$d_x^k := \delta^k + \widetilde{\mathcal{M}}_u \widetilde{\mathcal{M}}_d^{-1}(\delta^k - \widetilde{\delta}^k) \quad \text{and} \quad d_y^k := \gamma^k + \widetilde{\mathcal{N}}_u \widetilde{\mathcal{N}}_d^{-1}(\gamma^k - \widetilde{\gamma}^k). \tag{19}$$

Proposition 2

Suppose that $\widetilde{\mathcal{M}}_d \succ 0$ *and* $\widetilde{\mathcal{N}}_d \succ 0$ *for* $\widetilde{\mathcal{M}}$ *and* $\widetilde{\mathcal{N}}$ *defined in* (17). *Let* κ *and* κ' *be defined as in* (18). *Then, the sequences* $\{w^k := (x^k, y^k, z^k)\}$, $\{\delta^k\}$, $\{\widetilde{\delta}^k\}$, $\{\gamma^k\}$ *and* $\{\widetilde{\gamma}^k\}$ *generated by the sGS-imsPADMM are well-defined and it holds that*

$$\widehat{\mathcal{M}} = \widetilde{\mathcal{M}} + \mathrm{sGS}(\widetilde{\mathcal{M}}) \succ 0, \quad \widehat{\mathcal{N}} = \widetilde{\mathcal{N}} + \mathrm{sGS}(\widetilde{\mathcal{N}}) \succ 0. \tag{20}$$

Moreover, for any $k \geq 0$, d_x^k *and* d_y^k *defined by* (19) *satisfy*

$$\begin{cases} d_x^k \in \partial_x\big(\widehat{\mathcal{L}}_\sigma^k(x^{k+1}, y^k) + \frac{1}{2}\|x^{k+1} - x^k\|_{\widehat{\mathcal{S}}}^2\big), \\ d_y^k \in \partial_y\big(\widehat{\mathcal{L}}_\sigma^k(x^{k+1}, y^{k+1}) + \frac{1}{2}\|y^{k+1} - y^k\|_{\widehat{\mathcal{T}}}^2\big), \end{cases} \tag{21}$$

$$\|\widehat{\mathcal{M}}^{-\frac{1}{2}} d_x^k\| \leq \kappa \widetilde{\varepsilon}_k, \quad \|\widehat{\mathcal{N}}^{-\frac{1}{2}} d_y^k\| \leq \kappa' \widetilde{\varepsilon}_k. \tag{22}$$

If in the imsPADMM, we choose $\mathcal{S} := \widehat{\mathcal{S}}$, $\mathcal{T} := \widehat{\mathcal{T}}$, then we have $\mathcal{M} = \widehat{\mathcal{M}} \succ 0$ and $\mathcal{N} = \widehat{\mathcal{N}} \succ 0$. Moreover, we can define the sequence $\{\varepsilon_k\}$ by $\varepsilon_k := \max\{\kappa, \kappa'\}\widetilde{\varepsilon}_k \ \forall k \geq 0$. The sequence $\{w^k\}$ generated by the sGS-imsPADMM always satisfies $\|\mathcal{M}^{-\frac{1}{2}} d_x^k\| \leq \varepsilon_k$ and $\|\mathcal{N}^{-\frac{1}{2}} d_y^k\| \leq \varepsilon_k$. Thus, $\{w^k\}$ can be viewed as a sequence generated by the imsPADMM with specially constructed semi-proximal terms.

- sGS-imsPADMM is an explicitly implementable method to handle high-dimensional convex composite conic optimization problems.
- imsPADMM has a compact formulation which can facilitate the convergence analysis of the sGS-imsPADMM.

Avoid of repetition

- We can use the $\widetilde{x}_i^{k+1}$ computed in the backward GS sweep (Step 1a) to estimate x_i^{k+1} in the forward sweep (Step 1b) for $i = 2, \ldots, m$.
- In this case, the corresponding error vector is given by

$$\delta_i^k = \widetilde{\delta}_i^k + \sum_{j=1}^{i-1} \widetilde{\mathcal{M}}_{ij}(x_j^{k+1} - x_j^k),$$

  and we may accept the approximate solution $x_i^{k+1} = \widetilde{x}_i^{k+1}$ without solving an additional subproblem if $\|\delta_i^k\| \le \widetilde{\varepsilon}_k$.
- A similar strategy also applies to the subproblems in Step 2b for $j = 2, \ldots, n$.

We only need to establish the convergence for imsPADMM!

Theorem 1

Suppose that the solution set $\overline{\mathcal{W}}$ *to the KKT system of problem (12) is nonempty and the sequence* $\{(x^k, y^k, z^k)\}$ *is generated by the imsPADMM. Assume that*[2]

$$\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0 \quad \textit{and} \quad \Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0. \tag{23}$$

Then, the sequence $\{x^k, y^k, z^k\}$ *converges to a point in* $\overline{\mathcal{W}}$.

[2] In fact, the theorem is still valid if (23) is replaced by the condition that $\widehat{\Sigma}_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0$ and $\widehat{\Sigma}_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0$.

To handle the inequality constraints in (2) we introduce a slack variable v to get

$$\begin{array}{ll} \max & \big(-\delta^*_{\mathcal{N}}(-Z) - \delta_{\mathbb{R}^{m_I}_+}(v) \big) - \frac{1}{2}\langle W,\, \mathcal{Q}W \rangle - \delta_{\mathcal{S}^n_+}(S) \\ & + \langle b_E,\, y_E \rangle + \langle b_I,\, y_I \rangle \\ \text{s.t.} & Z - \mathcal{Q}W + S + \mathcal{A}_E^* y_E + \mathcal{A}_I^* y_I = C, \\ & \mathcal{D}(v - y_I) = 0, \quad W \in \mathcal{W}, \end{array} \tag{24}$$

where $\mathcal{D} \in \mathbb{R}^{m_I \times m_I}$ is a fixed positive definite matrix.

We construct QSDP test instances based on the doubly nonnegative SDP problems arising from relaxation of binary integer quadratic (BIQ) programming with a large number of inequality constraints that was introduced by Sun et. al. for getting tighter bounds:

$$
\begin{array}{ll}
\min & \frac{1}{2}\langle X,\, \mathcal{Q}(X)\rangle + \frac{1}{2}\langle Q,\, \overline{X}\rangle + \langle c,\, x\rangle \\
\text{s.t.} & \mathrm{diag}(\overline{X}) - x = 0, \quad \alpha = 1,\; X = \begin{pmatrix} \overline{X} & x \\ x^T & \alpha \end{pmatrix} \in \mathcal{S}^n_+, \\
& \quad X \in \mathcal{N} := \{X \in \mathcal{S}^n :\; X \geq 0\},
\end{array}
$$

$$
\begin{cases}
-\overline{X}_{ij} + x_i \geq 0, & \\
-\overline{X}_{ij} + x_j \geq 0, & \forall\, i < j,\; j = 2, \ldots, n-1. \\
\overline{X}_{ij} - x_i - x_j \geq -1 &
\end{cases}
$$

For convenience, we call them as QSDP-BIQ problems. When $\mathcal{Q}$ is vacuous, we call the corresponding linear SDP problems as SDP-BIQ problems.

- The test data for Q and c are taken from the Biq Mac Library `http://biqmac.uni-klu.ac.at/biqmaclib.html`.
- We tested one group of SDP-BIQ problems and three groups of QSDP-BIQ problems with each group consisting of 80 instances with n ranging from 151 to 501.
- We compare the performance of our sGS-imsPADMM with the directly extended multi-block sPADMM with the aggressive step-length of 1.618 on solving these SDP/QSDP-BIQ problems.
- **Note:** Although its convergence is not guaranteed, such a directly extended sPADMM is currently more or less the benchmark among first-order methods for solving multi-block linear and quadratic SDPs.

Stop the algorithm after 500,000 iterations, or

$$\eta_{\mathsf{qsdp}} = \max\{\eta_P, \eta_D, \eta_W, \eta_{I_1}, \eta_{I_2}, \eta_{s_1}, \eta_{I_3}, \eta_{S_2}, \eta_X, \eta_Z\} < 10^{-6},$$

where

$$\eta_D = \frac{\|\mathcal{A}_E^* y_E + \mathcal{A}_I^* y_I + S + Z - \mathcal{Q}W - C\|}{1+\|C\|}, \eta_P = \frac{\|\mathcal{A}_E X - b_E\|}{1+\|b_E\|},$$

$$\eta_{I_1} = \frac{\|\min(0, y_I)\|}{1+\|y_I\|}, \ \eta_{I_2} = \frac{\|\min(0, \mathcal{A}_I X - b_I)\|}{1+\|b_I\|}, \ \eta_{I_3} = \frac{|\langle \mathcal{A}_I X - b_I, y_I\rangle|}{1+\|\mathcal{A}_I x - b_I\| + \|y_I\|}$$

$$\eta_{S1} = \frac{\|X - \Pi_{\mathcal{S}_+^n}(X)\|}{1+\|X\|}, \ \eta_{S2} = \frac{|\langle X, S\rangle|}{1+\|X\|+\|S\|}, \ \eta_W = \frac{\|\mathcal{Q}X - \mathcal{Q}W\|}{1+\|\mathcal{Q}\|},$$

$$\eta_X = \frac{\|X - \Pi_{\mathcal{N}}(X)\|}{1+\|X\|}, \ \eta_Z = \frac{\|X - \Pi_{\mathcal{N}}(X - Z)\|}{1+\|X\|+\|Z\|}.$$

In addition, we also measure the duality gap:

$$\eta_{gap} := \frac{\mathrm{Obj}_{\mathrm{primal}} - \mathrm{Obj}_{\mathrm{dual}}}{1+|\mathrm{Obj}_{\mathrm{primal}}| + |\mathrm{Obj}_{\mathrm{dual}}|},$$

where

$$\begin{cases} \mathrm{Obj}_{\mathrm{primal}} := \frac{1}{2}\langle X, \mathcal{Q}X\rangle + \langle C, X\rangle, \\ \mathrm{Obj}_{\mathrm{dual}} := -\delta_{\mathcal{N}}^*(-Z) - \frac{1}{2}\langle W, \mathcal{Q}W\rangle + \langle b_E, y_E\rangle + \langle b_I, y_I\rangle. \end{cases}$$

For example, the subproblem corresponding to the block y_I in ADMM type methods with/without semi-proximal term has to be solved:

$$\min \left\{ -\langle b_I,\, y_I \rangle + \frac{\sigma}{2} \| [\mathcal{A}_I, -\mathcal{D}]^* y_I - r \|^2 + \frac{1}{2} \| y_I - y_I^- \|_{\mathcal{T}}^2 \right\}, \quad (25)$$

where $\mathcal{T}$ is a self-adjoint positive semidefinite linear operator on $\mathbb{R}^{m_I}$, and r and y_I^- are given data.

Define $\mathcal{V} := \mathcal{A}_I \mathcal{A}_I^* + \mathcal{D}^2$ and $\widetilde{r} := b_I + \sigma(\mathcal{A}_I, -\mathcal{D})r + \mathcal{T} y_I^-$. Solving (25) is equivalent to solving the linear equation

$$\big(\sigma \mathcal{V} + \mathcal{T}\big) y_I = \widetilde{r} \quad (26)$$

Remark: It is generally very difficult to compute the solution of (26) exactly for large scale problems if $\mathcal{T}$ is the zero operator, i.e., not adding a proximal term.

Solve $(\sigma\mathcal{V} + \mathcal{T})y_I = \widetilde{r}$

Suppose that $\mathcal{V}$ admits the eigenvalue decomposition

$$\mathcal{V} = \sum_{i=1}^{n} \lambda_i \mathcal{P}_i \mathcal{P}_i^*,$$

with $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$. We can choose $\mathcal{T}$ by using the first l largest eigenvalues and the corresponding eigenvectors of $\mathcal{V}$, i.e.,

$$\mathcal{T} = \sigma \textstyle\sum_{i=l+1}^{n} (\lambda_{l+1} - \lambda_i) \mathcal{P}_i \mathcal{P}_i^*, \tag{27}$$

which is self-adjoint positive semidefinite.

Remark: it is more likely that such a $\mathcal{T}$ is "**smaller**" than the natural choice of setting it to be $\sigma(\lambda_1 \mathcal{I} - \mathcal{V})$. Indeed we have observed in our numerical experiments that the latter choice always leads to more iterations compared to the choice in (27).

Solve $(\sigma\mathcal{V} + \mathcal{T})y_I = \widetilde{r}$

$(\sigma\mathcal{V} + \mathcal{T})^{-1}$ can be obtained analytically as

$$(\sigma\mathcal{V} + \mathcal{T})^{-1} = (\sigma\lambda_{l+1})^{-1}\mathcal{I} + \sum_{i=1}^{l}((\sigma\lambda_i)^{-1} - (\sigma\lambda_{l+1})^{-1})\mathcal{P}_i\mathcal{P}_i^*.$$

Thus, we **only need to calculate the first few largest eigenvalues and the corresponding eigenvectors of $\mathcal{V}$** and this can be done efficiently via variants of the Lanczos method.

When the problem (25) is allowed to be solved **inexactly**, we can set $\mathcal{T} = 0$ in (25) and solve the linear system $\sigma\mathcal{V} = \widetilde{r}$ by a preconditioned conjugate gradient (PCG) method.

- In this setting, $(\sigma\mathcal{V} + \mathcal{T})^{-1}$ with $\mathcal{T}$ defined in (27) can serve as an effective **preconditioner**.

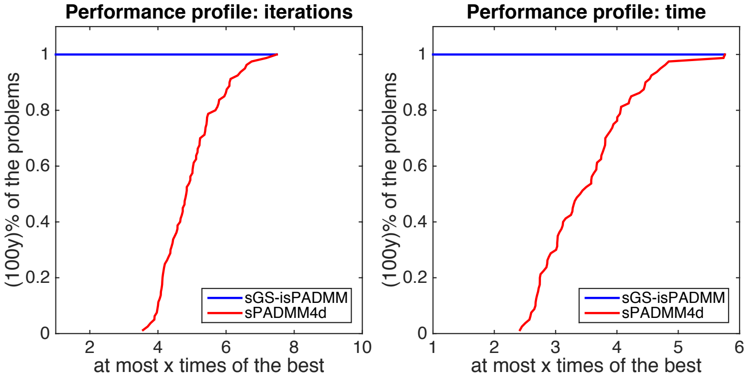

Figure: Performance profiles of sGS-isPADMM and sPADMM4d on solving the SDP-BIQ problems.

Numerical Performance for QSDP Problems

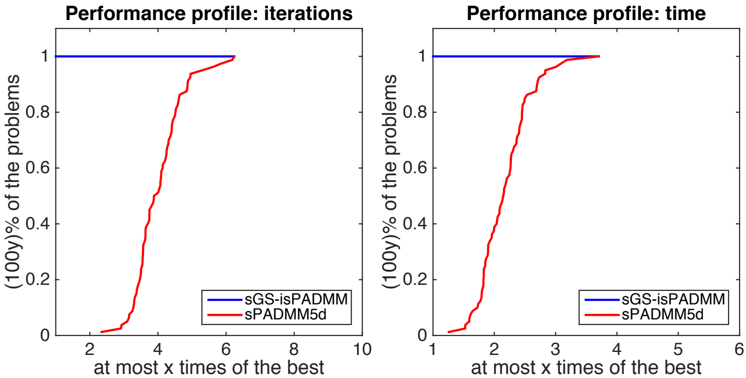

Figure: Performance profiles of sGS-isPADMM and sPADMM5d on solving the QSDP-BIQ problems (group 1).

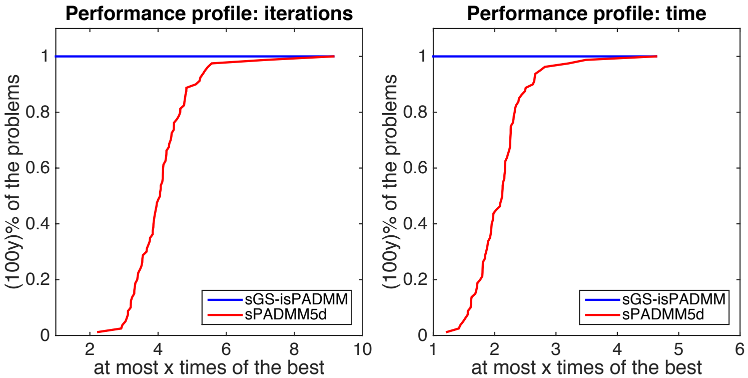

Figure: Performance profiles of sGS-isPADMM and sPADMM5d on solving the QSDP-BIQ problems (group 2).

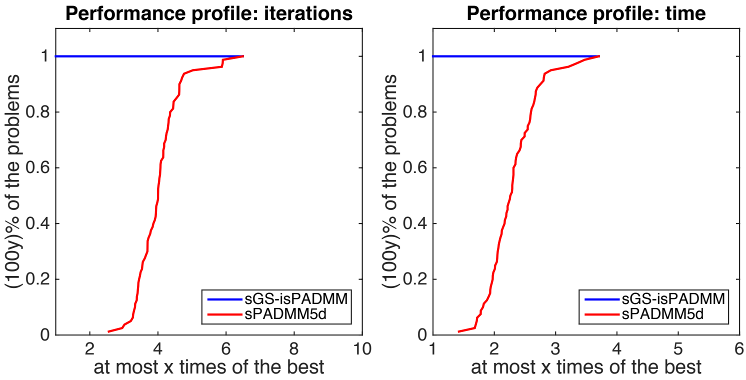

Figure: Performance profiles of sGS-isPADMM and sPADMM5d on solving the QSDP-BIQ problems (group 3).

Table: The numerical performance of sGS-isPADMM and the directly extended multi-block ADMM with step-length $\tau = 1.618$ ($n > 500$, accuracy $= 10^{-6}$)

Problem	$m_E; m_I$	n_s	Iteration sGS-isP|sP-d	η_{qsdp} sGS-isP|sP-d	η_{gap} sGS-isP|sP-d	Time sGS-isP|sP-d
SDP-BIQ						
bqp500-2	501;374250	501	17525|82401	9.9-7|9.9-7	-6.3-7|2.3-8	42:27|2:12:29
bqp500-4	501;374250	501	15352|75995	9.9-7|9.9-7	-6.4-7|-3.2-8	36:53|1:59:52
bqp500-6	501;374250	501	17747|78119	9.9-7|9.9-7	-1.6-7|-2.4-8	45:10|2:04:23
bqp500-8	501;374250	501	20386|110825	9.9-7|9.9-7	-4.3-7|2.1-8	52:04|3:10:43
bqp500-10	501;374250	501	16407|68985	9.7-7|9.9-7	-5.6-7|3.7-9	39:30|1:46:01
gka1f	501;374250	501	9101|60073	9.9-7|9.9-7	-4.4-7|1.1-8	20:22|1:32:22
gka2f	501;374250	501	16193|74034	9.9-7|9.9-7	-2.7-7|-1.1-8	39:35|1:59:59
gka3f	501;374250	501	16323|72563	9.9-7|9.9-7	-1.3-7|3.9-8	40:38|1:56:28
gka4f	501;374250	501	15502|63285	9.6-7|9.9-7	-6.1-7|3.4-8	36:58|1:41:20
gka5f	501;374250	501	17664|76164	9.9-7|9.9-7	-1.3-7|1.1-8	43:45|2:05:14
QSDP-BIQ (group 1)						
bqp500-2	501;374250	501	19053|71380	9.9-7|9.9-7	-1.2-7|1.1-8	1:02:31|1:52:02
bqp500-4	501;374250	501	13905|67865	9.9-7|9.9-7	-8.9-7|7.8-8	43:17|1:46:07
bqp500-6	501;374250	501	17211|62562	9.9-7|9.9-7	-2.0-7|6.9-8	56:23|1:37:19
bqp500-8	501;374250	501	19742|85057	9.9-7|9.9-7	-4.9-7|7.0-8	1:05:09|2:15:52
bqp500-10	501;374250	501	17690|65484	9.9-7|9.9-7	-2.3-7|6.7-8	58:00|1:43:04
gka1f	501;374250	501	8919|55669	9.9-7|9.9-7	-8.8-7|4.1-8	26:42|1:25:01
gka2f	501;374250	501	13587|61324	9.9-7|9.9-7	-4.5-7|2.1-8	42:50|1:37:15
gka3f	501;374250	501	13786|62438	9.9-7|9.9-7	-2.2-7|3.1-8	42:55|1:37:29
gka4f	501;374250	501	13953|57164	9.6-7|9.9-7	-7.2-7|-3.4-8	44:25|1:31:14
gka5f	501;374250	501	15968|62001	9.9-7|9.9-7	-1.4-7|4.6-8	50:22|1:35:40

Table: The numerical performance of sGS-isPADMM and the directly extended multi-block ADMM with step-length $\tau = 1.618$ ($n > 500$, accuracy $= 10^{-6}$)

Problem	$m_E; m_I$	n_s	Iteration sGS-isP|sP-d	η_{qsdp} sGS-isP|sP-d	η_{gap} sGS-isP|sP-d	Time sGS-isP|sP-d
QSDP-BIQ (group 2)						
bqp500-2	501;374250	501	16506|79086	9.9-7|9.9-7	-1.2-7|4.2-8	52:46|1:52:08
bqp500-4	501;374250	501	8675|30677	9.9-7|9.9-7	2.7-8|2.3-8	25:32|41:15
bqp500-6	501;374250	501	10043|42654	9.9-7|9.9-7	-3.0-8|8.3-8	29:46|58:58
bqp500-8	501;374250	501	9410|43785	9.9-7|9.9-7	-2.5-8|2.9-8	27:37|59:05
bqp500-10	501;374250	501	10656|35213	9.9-7|9.9-7	-3.6-8|8.8-8	32:35|47:00
gka1f	501;374250	501	10939|52226	9.9-7|9.9-7	-5.8-8|3.8-8	36:10|1:16:48
gka2f	501;374250	501	7757|34660	9.9-7|9.9-7	-1.8-8|6.0-8	25:17|48:40
gka3f	501;374250	501	11241|45857	9.9-7|9.9-7	-1.2-8|2.7-8	34:55|1:02:59
gka4f	501;374250	501	11706|37466	9.9-7|9.9-7	-3.7-8|6.4-8	36:19|51:25
gka5f	501;374250	501	14229|48670	9.9-7|9.9-7	-4.8-8|9.8-8	42:37|1:06:37
QSDP-BIQ (group 3)						
bqp500-2	501;374250	501	18311|66867	9.9-7|9.9-7	-1.9-7|1.2-7	41:33|1:11:30
bqp500-4	501;374250	501	14169|65580	9.9-7|9.9-7	-7.8-7|1.1-7	30:04|1:10:29
bqp500-6	501;374250	501	16428|68301	9.9-7|9.9-7	-2.3-7|8.4-8	36:25|1:13:20
bqp500-8	501;374250	501	26308|107664	9.9-7|9.9-7	-4.0-7|9.5-9	1:01:17|2:00:06
bqp500-10	501;374250	501	16398|57221	9.9-7|9.9-7	-2.8-7|8.6-8	37:22|1:06:27
gka1f	501;374250	501	14479|51294	9.9-7|9.9-7	-3.6-7|7.0-8	31:05|59:17
gka2f	501;374250	501	9365|60799	9.9-7|9.9-7	-1.5-6|-1.9-9	18:30|1:04:14
gka3f	501;374250	501	14175|57782	9.9-7|9.9-7	-3.2-7|2.0-8	30:10|1:01:35
gka4f	501;374250	501	13356|56588	9.8-7|9.9-7	-5.8-7|-2.0-8	27:42|1:00:10
gka5f	501;374250	501	14122|58716	9.9-7|9.9-7	-1.4-7|9.3-8	29:38|1:01:13

Concluding remark

- Combining an inexact 2-block majorized sPADMM and the recent advances in the inexact block symmetric Gauss-Seidel (sGS) technique

- Only needs one cycle of an inexact sGS iteration, instead of an unknown number of cycles, to solve each of the subproblems involved.

- For the vast majority of the tested problems, the proposed sGS-imsPADMM is 2 to 3 times faster than the directly extended multi-block PADMM even with the aggressive step length of 1.618.

Conclusions

- One does not need to sacrifice speed in exchange for convergence guarantee in developing ADMM-type first order methods, at least for solving high-dimensional linear and convex quadratic SDP problems to moderate accuracy.
- The merit that is brought about by solving the original subproblems inexactly without adding proximal terms is thus evidently clear.
- More powerful algorithms are needed such as SDPNAL for solving the standard SDP.

Reference on first-order methods

- D.F. Sun, K.-C. Toh and L. Yang, A convergent 3-block semiproximal alternating direction method of multipliers for conic programming with 4-type constraints SIAM J. Optim., 25 (2015) 882-915.
- X.D. Li, D.F. Sun and K.-C. Toh, A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions, Math. Program., 155 (2016) 333–373.
- L. Chen, D.F. Sun and K.-C. Toh, An efficient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming, Math. Program., 161 (2017) 237-270.

Reference on second-order methods

- X.Y. Zhao, D.F. Sun, K.-C. Toh, A Newton-CG augmented Lagrangian method for semidefinite programming, SIAM J. Optim., 20 (2010) 1737–1765.

- L.Q. Yang, D.F. Sun and K.-C. Toh, SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints, Math. Program. Comput., 7 (2015) 331–366.

- X.D. Li, D.F. Sun and K.-C. Toh, QSDPNAL: A two-phase augmented Lagrangian method for convex quadratic semidefinite programming, arXiv:1512.08872 (2015).