On the Relationships of ADMM and Proximal ALM for Convex Optimization Problems

Defeng Sun

Department of Applied Mathematics

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Institute of Applied Physics and Computational Mathematics, Beijing

April 12, 2019

Based on joint works with
L. Chen (PolyU/HNU), X.D. Li (Fudan), K-C. Toh (NUS), N. Zhang (PolyU/DGUT)

Multi-block convex programming

$$\min_{x,y} \Big\{ p_1(x_1) + f(\underbrace{x_1, \ldots, x_m}_{x}) + q_1(y_1) + g(\underbrace{y_1, \ldots, y_n}_{y}) \,|\, \mathcal{A}^* x + \mathcal{B}^* y = c \Big\} \tag{P}$$

- $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$: finite-dim. real Hilbert spaces endowed with $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$
- $p_1 : \mathcal{X}_1 \to (-\infty, \infty]$ and $q_1 : \mathcal{Y}_1 \to (-\infty, \infty]$ are closed and proper convex functions. Denote $p(x) := p_1(x_1)$ and $q(y) := q_1(y_1)$
- $f : \mathcal{X} \to (-\infty, \infty)$ and $g : \mathcal{Y} \to (-\infty, \infty)$ are convex, continuously differentiable with Lipschitz continuous gradients
- $\mathcal{A}^*$ and $\mathcal{B}^*$ are the adjoints of the linear mappings $\mathcal{A} : \mathcal{Z} \to \mathcal{X}$ and $\mathcal{B} : \mathcal{Z} \to \mathcal{Y}$, $c \in \mathcal{Z}$

Notation

- Let $\mathcal{U}$ and $\mathcal{V}$ be two finite dimensional real Hilbert spaces. For any given linear map $\mathcal{H} : \mathcal{U} \to \mathcal{V}$, we use $\|\mathcal{H}\|$ to denote its spectral norm and $\mathcal{H}^* : \mathcal{V} \to \mathcal{U}$ to denote its adjoint linear operator
- If $\mathcal{U} = \mathcal{V}$ and $\mathcal{H}$ is self-adjoint, for any $u, v \in \mathcal{U}$, define $\langle u, v\rangle_{\mathcal{H}} := \langle u, \mathcal{H}v\rangle$ and $\|u\|_{\mathcal{H}}^2 := \langle u, \mathcal{H}u\rangle$; if $\mathcal{H}$ is also positive semidefinite, there exists a unique self-adjoint positive semidefinite linear operator $\mathcal{H}^{\frac{1}{2}} : \mathcal{U} \to \mathcal{U}$ such that $\mathcal{H}^{\frac{1}{2}}\mathcal{H}^{\frac{1}{2}} = \mathcal{H}$.
- For a closed proper convex function $\theta : \mathcal{U} \to (-\infty, +\infty]$, denote by $\operatorname{dom}\theta$ and $\partial\theta$ for the effective domain and the subdifferential mapping of θ, respectively

Decomposition

Decompose $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_s$, with each $\mathcal{U}_i$ being a finite dimensional real Hilbert space endowed with $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$

Decompose the self-adjoint and positive semidefinite $\mathcal{H}$ as

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

where $\mathcal{H}_{ij} : \mathcal{U}_j \to \mathcal{U}_i$, $i, j = 1, \ldots, s$ are linear maps and $\mathcal{H}_{ii}$ are self-adjoint positive definite linear operators ($\mathcal{H}_{ii} \succ 0$), $i = 1, \ldots, s$

We use $\mathcal{H}_d := \mathrm{Diag}(\mathcal{H}_{11}, \ldots, \mathcal{H}_{ss})$ to denote the block-diagonal part of $\mathcal{H}$, and denote the symbolically strictly upper triangular part of $\mathcal{H}$ by $\mathcal{H}_u$. Thus, $\mathcal{H} = \mathcal{H}_d + \mathcal{H}_u + \mathcal{H}_u^*$

One cycle of the block symmetric Gauss-Seidel

Let $\theta_1 : \mathcal{U}_1 \to (-\infty, \infty]$ be a given closed and proper convex function, $b \in \mathcal{U}$ be a given vector, and $h : \mathcal{U} \to (-\infty, \infty)$ be defined by

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

Suppose that $u^- \in \mathcal{U}$ is a given vector. Define

$$\begin{cases} u_i^{\frac{1}{2}} := \arg\min_{u_i} \Big\{\theta(u_1^-) + h(u_{<i}^-, u_i, u_{>i}^{\frac{1}{2}}) - \langle \tilde{\delta}_i, u_i\rangle\Big\}, \quad i = s, \ldots, 2 \\ u_1^+ := \arg\min_{u_1} \Big\{\theta(u_1) + h(u_1, u_{>1}^{\frac{1}{2}}) - \langle \delta_1, u_1\rangle\Big\}, \\ u_i^+ := \arg\min_{u_i} \Big\{\theta(u_1^+) + h(u_{<i}^+, u_i, u_{>i}^{\frac{1}{2}}) - \langle \delta_i, u_i\rangle\Big\}, \quad i = 2, \ldots, s \end{cases} \tag{sGS}$$

where for any $u = (u_1, \ldots, u_s) \in \mathcal{U}$ and $i \in \{1, \ldots, s\}$, we denote $u_{<i} := \{u_1, \ldots, u_{i-1}\}$, $u_{>i} := \{u_{i+1}, \ldots, u_s\}$

One cycle of the block sGS

Define

$$d(\tilde{\delta}, \delta) := \delta + \mathcal{H}_u \mathcal{H}_d^{-1}(\delta - \tilde{\delta})$$

with $\tilde{\delta}_1 = \delta_1$, $\delta := (\delta_1, \ldots, \delta_s)$ and $\tilde{\delta} := (\tilde{\delta}_1, \ldots, \tilde{\delta}_s)$

Define the self-adjoint positive semidefinite linear operator on $\mathcal{U}$ by

$$\mathrm{sGS}(\mathcal{H}) := \mathcal{H}_u \mathcal{H}_d^{-1} \mathcal{H}_u^* \quad \text{(sGS Splitting Operator)}$$

Consider the following convex composite quadratic programming:

$$\min_{u \in \mathcal{U}} \left\{ \theta(u_1) + h(u) + \frac{1}{2}\|u - u^-\|^2_{\mathrm{sGS}(\mathcal{H})} - \langle d(\tilde{\delta}, \delta),\, u \rangle \right\} \qquad \text{(CQP)}$$

Block sGS decomposition theorem

Theorem

Suppose that $\mathcal{H}_d = \mathrm{Diag}(\mathcal{H}_{11}, \ldots, \mathcal{H}_{ss}) \succ 0$. *Then,*

- (CQP) *is well-defined and admits a unique solution, which is exactly the vector* u^+ *generated by the* (sGS) *procedure*
-
$$\widehat{\mathcal{H}} := \mathcal{H} + \mathrm{sGS}(\mathcal{H}) = (\mathcal{H}_d + \mathcal{H}_u)\mathcal{H}_d^{-1}(\mathcal{H}_d + \mathcal{H}_u^*) \succ 0$$
- *the error vector* $d(\tilde{\delta}, \delta)$ *satisfies*

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

Majorization

Problem (P):

$$\min\left\{p_1(x_1) + f(x) + q_1(y_1) + g(y) \mid \mathcal{A}^*x + \mathcal{B}^*y = c\right\}$$

For the two smooth convex functions f and g in problem (P), there exist two 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

$$\begin{cases} f(x) \leq \hat{f}(x; x') := f(x') + \langle \nabla f(x'), x - x' \rangle + \frac{1}{2}\|x - x'\|^2_{\widehat{\Sigma}^f} \\ g(y) \leq \hat{g}(y; y') := g(y') + \langle \nabla g(y'), y - y' \rangle + \frac{1}{2}\|y - y'\|^2_{\widehat{\Sigma}^g} \end{cases}$$

Quadratic on the RHS

Majorized proximal augmented Lagrangian function

For any given $\sigma > 0$, the majorized proximal augmented Lagrangian function associated with problem (P) is defined by

$$\widetilde{\mathcal{L}}_\sigma(x, y; (x', y', z')) :$$

$$= p(x) + \hat{f}(x; x') + q(y) + \hat{g}(y; y') + \langle z', \mathcal{A}^* x + \mathcal{B}^* y - c\rangle$$

$$+ \frac{\sigma}{2}\|\mathcal{A}^* x + \mathcal{B}^* y - c\|^2 + \frac{1}{2}\|x - x'\|^2_{\widetilde{\mathcal{S}}} + \frac{1}{2}\|y - y'\|^2_{\widetilde{\mathcal{T}}},$$

$$\forall\, (x, y) \in \mathcal{X} \times \mathcal{Y} \quad \text{and} \quad \forall\, (x', y', z') \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z},$$

where $\widetilde{\mathcal{S}} : \mathcal{X} \to \mathcal{X}$ and $\widetilde{\mathcal{T}} : \mathcal{Y} \to \mathcal{Y}$ are self-adjoint (not necessarily positive semidefinite) linear operators

Nonsmooth+Quadratic Terms on RHS

sGS-imiPADMM

An inexact sGS decomposition based majorized indefinite-proximal ADMM

Let $\tau \in (0, (1+\sqrt{5})/2)$ [e.g., $\tau = 1.618$], $\{\tilde{\varepsilon}_k\}_{k\geq 0}$ be a summable nonnegative sequence, $(x^0, y^0, z^0) \in \mathrm{dom}\, p \times \mathrm{dom}\, q \times \mathcal{Z}$ be the initial point

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

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

$$\boxed{x_i^{k+\frac{1}{2}} \approx \arg\min_{x_i \in \mathcal{X}_i} \left\{ \widetilde{\mathcal{L}}_\sigma \left((x_{<i}^k, x_i, x_{>i}^{k+\frac{1}{2}}), y^k; (x^k, y^k, z^k) \right) \right\}},$$

$$\tilde{\delta}_i^k \in \partial_{x_i} \widetilde{\mathcal{L}}_\sigma \left((x_{<i}^k, x_i^{k+\frac{1}{2}}, x_{>i}^{k+\frac{1}{2}}), y^k; (x^k, y^k, z^k) \right) \text{ with } \|\tilde{\delta}_i^k\| \leq \tilde{\varepsilon}_k$$

1b. Compute for $i = 1, \ldots, m,$

$$\boxed{x_i^{k+1} \approx \arg\min_{x_i \in \mathcal{X}_i} \left\{ \widetilde{\mathcal{L}}_\sigma \left((x_{<i}^{k+1}, x_i, x_{>i}^{k+\frac{1}{2}}), y^k; (x^k, y^k, z^k) \right) \right\}},$$

$$\delta_i^k \in \partial_{x_i} \widetilde{\mathcal{L}}_\sigma \left((x_{<i}^{k+1}, x_i^{k+1}, x_{>i}^{k+\frac{1}{2}}), y^k; (x^k, y^k, z^k) \right) \text{ with } \|\delta_i^k\| \leq \tilde{\varepsilon}_k$$

sGS-imiPADMM

2a. Compute for $j = n, \ldots, 2$,

$$\boxed{y_j^{k+\frac{1}{2}} \approx \arg\min_{y_j \in \mathcal{Y}_j} \left\{ \widetilde{\mathcal{L}}_\sigma \left(x^{k+1}, (y_{<j}^k, y_j, y_{>j}^{k+\frac{1}{2}}); (x^k, y^k, z^k) \right) \right\}},$$

$$\tilde{\gamma}_j^k \in \partial_{y_j} \widetilde{\mathcal{L}}_\sigma \left(x^{k+1}, (y_{<j}^k, y_j^{k+\frac{1}{2}}, y_{>j}^{k+\frac{1}{2}}); (x^k, y^k, z^k) \right) \text{ with } \|\tilde{\gamma}_j^k\| \leq \tilde{\varepsilon}_k$$

2b. Compute for $j = 1, \ldots, n$,

$$\boxed{y_j^{k+1} \approx \arg\min_{y_j \in \mathcal{Y}_j} \left\{ \widetilde{\mathcal{L}}_\sigma \left(x^{k+1}, (y_{<j}^{k+1}, y_j, y_{>j}^{k+\frac{1}{2}}); (x^k, y^k, z^k) \right) \right\}},$$

$$\gamma_j^k \in \partial_{y_j} \widetilde{\mathcal{L}}_\sigma \left(x^{k+1}, (y_{<j}^{k+1}, y_j^{k+1}, y_{>j}^{k+\frac{1}{2}}); (x^k, y^k, z^k) \right) \text{ with } \|\gamma_j^k\| \leq \tilde{\varepsilon}_k$$

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

Decompositions

We symbolically decompose the positive semidefinite linear operators $\widehat{\Sigma}^f$ into

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

and decompose $\widehat{\Sigma}^g$ similarly, in consistent with the decompositions of $\mathcal{X}$ and $\mathcal{Y}$.

Define two linear operators $\widetilde{\mathcal{M}} : \mathcal{X} \to \mathcal{X}$ and $\widetilde{\mathcal{N}} : \mathcal{Y} \to \mathcal{Y}$ as follows:

$$\widetilde{\mathcal{M}} := \widehat{\Sigma}^f + \sigma \mathcal{A}\mathcal{A}^* + \widetilde{\mathcal{S}}, \quad \widetilde{\mathcal{N}} := \widehat{\Sigma}^g + \sigma \mathcal{B}\mathcal{B}^* + \widetilde{\mathcal{T}}$$

Just like the decomposition of $\widehat{\Sigma}^f$ and $\widehat{\Sigma}^g$ in (2), we can symbolically decompose $\widetilde{\mathcal{S}}$, $\widetilde{\mathcal{T}}$, $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{N}}$ accordingly.

Decompositions

We use $\widetilde{\mathcal{M}}_d$ and $\widetilde{\mathcal{N}}_d$ to denote the corresponding diagonal parts, and $\widetilde{\mathcal{M}}_u$ and $\widetilde{\mathcal{N}}_u$ to denote the strictly upper triangular parts, respectively. Then,

$$\widetilde{\mathcal{M}} = \widetilde{\mathcal{M}}_u + \widetilde{\mathcal{M}}_d + \widetilde{\mathcal{M}}_u^*, \quad \widetilde{\mathcal{N}} = \widetilde{\mathcal{N}}_u + \widetilde{\mathcal{N}}_d + \widetilde{\mathcal{N}}_u^*$$

Decompose $\mathcal{A}$ and $\mathcal{B}$ as

$$\mathcal{A}z = (\mathcal{A}_1 z, \ldots, \mathcal{A}_m z) \quad \text{and} \quad \mathcal{B}z = (\mathcal{B}_1 z, \ldots, \mathcal{B}_n z)$$

where $\mathcal{A}_i z \in \mathcal{X}_i$ and $\mathcal{B}_j z \in \mathcal{Y}_j$, and $z \in \mathcal{Z}$

Define

$$\tilde{\delta}^k := (\tilde{\delta}_1^k, \ldots, \tilde{\delta}_m^k), \;\; \delta^k := (\delta_1^k, \ldots, \delta_m^k)$$

$$\tilde{\gamma}^k := (\tilde{\gamma}_1^k, \ldots, \tilde{\gamma}_n^k), \text{ and } \gamma^k := (\gamma_1^k, \ldots, \gamma_n^k)$$

with the convention that $\tilde{\delta}_1^k := \delta_1^k$ and $\tilde{\gamma}_1^k := \gamma_1^k$.

Decompositions

To apply the block sGS decomposition theorem, we require

$$\widetilde{\mathcal{M}}_{ii} \equiv \widehat{\Sigma}^f_{ii} + \sigma \mathcal{A}_i \mathcal{A}_i^* + \widetilde{\mathcal{S}}_{ii} \succ 0, \quad i = 1, \ldots, m$$

$$\widetilde{\mathcal{N}}_{jj} \equiv \widehat{\Sigma}^g_{jj} + \sigma \mathcal{B}_j \mathcal{B}_j^* + \widetilde{\mathcal{T}}_{jj} \succ 0, \quad j = 1, \ldots, n$$

Define the following linear operators:

$$\begin{cases} \mathcal{S}_{\rm sGS} := \widetilde{\mathcal{S}} + {\rm sGS}(\widetilde{\mathcal{M}}) = \widetilde{\mathcal{S}} + \widetilde{\mathcal{M}}_u \widetilde{\mathcal{M}}_d^{-1} \widetilde{\mathcal{M}}_u^* \\ \mathcal{T}_{\rm sGS} := \widetilde{\mathcal{T}} + {\rm sGS}(\widetilde{\mathcal{N}}) = \widetilde{\mathcal{T}} + \widetilde{\mathcal{N}}_u \widetilde{\mathcal{N}}_d^{-1} \widetilde{\mathcal{N}}_u^* \end{cases} \tag{3}$$

Theorem (via the sGS decomposition theorem)

- $\mathcal{S}_{\rm sGS}$ *and* $\mathcal{T}_{\rm sGS}$ *defined in* (3) *are well-defined, and*

$$\mathcal{M}_{\rm sGS} := \widehat{\Sigma}^f + \sigma \mathcal{A}\mathcal{A}^* + \mathcal{S}_{\rm sGS} \succ 0, \ \mathcal{N}_{\rm sGS} := \widehat{\Sigma}^g + \sigma \mathcal{B}\mathcal{B}^* + \mathcal{T}_{\rm sGS} \succ 0 \tag{4}$$

- *it holds that*

$$\begin{cases} d_x^k \in \partial_x \left\{ \widetilde{\mathcal{L}}_\sigma \left(x^{k+1}, y^k; (x^k, y^k, z^k) \right) + \frac{1}{2} \| x^{k+1} - x^k \|^2_{{\rm sGS}(\widetilde{\mathcal{M}})} \right\}, \\ d_y^k \in \partial_y \left\{ \widetilde{\mathcal{L}}_\sigma \left(x^{k+1}, y^{k+1}; (x^k, y^k, z^k) \right) + \frac{1}{2} \| y^{k+1} - y^k \|^2_{{\rm sGS}(\widetilde{\mathcal{N}})} \right\}, \end{cases}$$

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

- *one has* $\| \mathcal{M}_{\rm sGS}^{-\frac{1}{2}} d_x^k \| \le \kappa \tilde{\varepsilon}_k$ *and* $\| \mathcal{N}_{\rm sGS}^{-\frac{1}{2}} d_y^k \| \le \kappa' \tilde{\varepsilon}_k$, *where* κ *and* κ' *are some constants*

Karush-Kuhn-Tucker (KKT)

Recall that the Karush-Kuhn-Tucker (KKT) system of problem (P) is given by

$$0 \in \partial p(x) + \nabla f(x) + \mathcal{A}z, \; 0 \in \partial q(y) + \nabla g(y) + \mathcal{B}z, \; \mathcal{A}^* x + \mathcal{B}^* y = c$$

Denote the solution set of the KKT system for problem (P) by $\overline{\mathcal{W}}$.

Stopping criterion: always use the (relative) KKT residual to stop an algorithm. The (relative) distance of two consecutive iterates CANNOT be used as a reliable stopping criterion.

Assumption

- *The solution set* $\overline{\mathcal{W}}$ *to the KKT system of* (P) *is nonempty*
- $\widetilde{\mathcal{S}}$ *and* $\widetilde{\mathcal{T}}$ *are chosen such that*

$$\widetilde{\mathcal{S}} \succeq -\frac{1}{2}\widehat{\Sigma}^{f} \quad \& \quad \widetilde{\mathcal{T}} \succeq -\frac{1}{2}\widehat{\Sigma}^{g} \tag{5}$$

and

$$\widehat{\Sigma}^{f}_{ii} + \sigma\mathcal{A}_i\mathcal{A}_i^* + \widetilde{\mathcal{S}}_{ii} \succ 0, \quad i = 1, \ldots, m$$
$$\widehat{\Sigma}^{g}_{jj} + \sigma\mathcal{B}_j\mathcal{B}_j^* + \widetilde{\mathcal{T}}_{jj} \succ 0, \quad j = 1, \ldots, n$$

Theorem (Convergence of sGS-imiPADMM)

Suppose that the Assumption holds, and the linear operators $\widetilde{\mathcal{S}}$ and $\widetilde{\mathcal{T}}$ are chosen such that

$$\frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{A}\mathcal{A}^* + \mathcal{S}_{\rm sGS} \succ 0 \quad \& \quad \frac{1}{2}\widehat{\Sigma}^g + \sigma\mathcal{B}\mathcal{B}^* + \mathcal{T}_{\rm sGS} \succ 0 \tag{6}$$

Then the whole sequence $\{(x^k, y^k, z^k)\}$ converges to a solution of the KKT system

Remark: The above convergence theorem is fairly general; it covers the classic ADMM and the most recent developments for solving multi-block convex optimization problems. Below we shall use a more specific example to reveal the relations between ADMM and proxiaml ALM.

Convex Composite Quadratic Programming

$$\min_{x\in\mathcal{X}}\left\{\psi(x)+\frac{1}{2}\langle x,\mathcal{Q}x\rangle-\langle c,x\rangle \;\Big|\; \mathcal{A}_E x=b_E,\ \mathcal{A}_I x-b_I\in\mathcal{K}\right\} \tag{7}$$

- $\psi:\mathcal{X}\to(-\infty,+\infty]$ is a closed proper convex function [simple]
- $\mathcal{Q}:\mathcal{X}\to\mathcal{X}$ satisfying $\mathcal{Q}=\mathcal{Q}^*,\ \mathcal{Q}\succeq 0$
- $\mathcal{A}_E:\mathcal{X}\to\mathcal{Z}_1$ and $\mathcal{A}_I:\mathcal{X}\to\mathcal{Z}_2$, given linear mappings
- $b=(b_E;b_I)\in\mathcal{Z}:=\mathcal{Z}_1\times\mathcal{Z}_2$, given vector
- $\mathcal{K}\subseteq\mathcal{Z}_2$ is a closed convex set (cone) [simple]

Equivalently,

$$\min_{x\in\mathcal{X},x'\in\mathcal{Z}_2}\left\{\psi(x)+\delta_{\mathcal{K}}(x')+\frac{1}{2}\langle x,\mathcal{Q}x\rangle-\langle c,x\rangle \;\Big|\; \begin{pmatrix}\mathcal{A}_E & 0\\ \mathcal{A}_I & -\mathcal{I}\end{pmatrix}\begin{pmatrix}x\\ x'\end{pmatrix}=b\right\}$$

CCQP

The dual of the above problem [or equivalently problem (7)] is

$$\min_{w,y,z} \left\{ p(w) + \frac{1}{2}\langle y, \mathcal{Q}y\rangle - \langle b, z\rangle \;\middle|\; \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} \mathcal{Q} \\ 0 \end{pmatrix} y - \begin{pmatrix} \mathcal{A}_E^* & \mathcal{A}_I^* \\ 0 & -\mathcal{I} \end{pmatrix} z = \begin{pmatrix} c \\ 0 \end{pmatrix} \right\}$$

- $w := (u, v) \in \mathcal{X} \times \mathcal{Z}_2$
- $p(w) := p(u, v) = \psi^*(u) + \delta_{\mathcal{K}}^*(v)$
- $\delta_{\mathcal{K}}(\cdot)$ is the indicator function over $\mathcal{K}$

- The dual is about (w, y, z) – three or more blocks
- Nonsmoothness only exists in one block of variables, i.e., the w-block
- It covers many important classes of convex optimization problems that are best solved in this (dual) form!

Deal with Convex Quadratic Constraints

Add additional convex quadratic constraints to problem (7):

$$\langle x, \mathcal{Q}_i x\rangle - \langle c_i', x\rangle \leq b_i', \quad i = 1, \ldots, l$$

where $\mathcal{Q}_i = \mathcal{L}_i \mathcal{L}_i^* \succeq 0$ for a certain linear operator $\mathcal{L}_i$

Write the above constraints as $\|\mathcal{L}_i^* x\|^2 - \langle c_i', x\rangle \leq b_i', \quad i = 1, \ldots, l$, or equivalently

$$\left\| \begin{pmatrix} 1 - b_i' - \langle c_i', x\rangle \\ 2\mathcal{L}_i^* x \end{pmatrix} \right\|_2 \leq 1 + b_i' + \langle c_i', x\rangle, \quad i = 1, \ldots, l$$

We can further rewrite the above as

$$\begin{pmatrix} 1 + b_i' + \langle c_i', x\rangle \\ 1 - b_i' - \langle c_i', x\rangle \\ 2\mathcal{L}_i^* x \end{pmatrix} \in \mathcal{K}_i, \quad i = 1, \ldots, l$$

where $\mathcal{K}_i$ is the second-order-cone of a proper dimension, $i = 1, \ldots, l$

Therefore, convex quadratic constraints can be added to problem (7) without changing its structure

Penalized and Constrained Regression Models

The penalized and constrained (PAC) regression often arises in high-dimensional generalized linear models with linear equality and inequality constraints, e.g.,

$$\min_{x\in\mathbb{R}^n} \left\{ p(x) + \frac{1}{2\lambda}\|\Phi x - \eta\|^2 \,\middle|\, A_E x = b_E,\ A_I x - b_I \in \mathcal{K} \right\}$$

- $p(\cdot)$ is a proper closed convex regularizer such as $p(x) = \|x\|_1,\ \|x\|_*$ [Non-convex counterparts can be dealt with via proximal DC (difference of convex functions) algorithm – another talk]
- $\lambda > 0$ is a parameter
- It is a special case of problem (7)

Multi-block convex composite optimization

$$\min_{y\in\mathcal{Y},z\in\mathcal{Z}}\Big\{\underbrace{p(y_1)+f(\underbrace{y_1,y_2,\ldots,y_s}_{y})-\langle b,z\rangle}_{\Phi(w)}\ \Big|\ \underbrace{\mathcal{F}^*y+\mathcal{G}^*z=c}_{\mathcal{A}^*w=c}\Big\}$$

with $w=(y,z)\in\mathcal{W}:=\mathcal{Y}\times\mathcal{Z}$

- $\mathcal{X}$, $\mathcal{Z}$ and $\mathcal{Y}:=\mathcal{Y}_1\times\cdots\times\mathcal{Y}_s$: finite-dimensional real Hilbert spaces, endowed with $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$
- $p:\mathcal{Y}_1\to(-\infty,+\infty]$: (nonsmooth) closed proper convex function
  $f:\mathcal{Y}\to(-\infty,+\infty)$: continuously differentiable convex function with Lipschitz gradient
- $\mathcal{F}^*$ and $\mathcal{G}^*$: the adjoints of the given linear mappings $\mathcal{F}:\mathcal{X}\to\mathcal{Y}$ and $\mathcal{G}:\mathcal{X}\to\mathcal{Z}$; $b\in\mathcal{Z}$ and $c\in\mathcal{X}$: the given data

The augmented Lagrangian function[1]

Recall the problem

$$\min_{y\in\mathcal{Y},z\in\mathcal{Z}} \{p(y_1) + f(y_1, y_2, \ldots, y_s) - \langle b, z\rangle \mid \mathcal{F}^*y + \mathcal{G}^*z = c\}$$

or

$$\min_{w\in\mathcal{W}} \{\Phi(w) \mid \mathcal{A}^*w = c\}$$

Let $\sigma > 0$ be the penalty parameter. The augmented Lagrangian function:

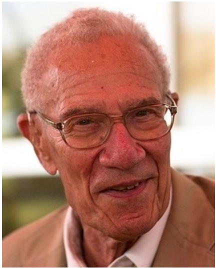
$$\begin{aligned}\mathcal{L}_\sigma(y, z; x) :=\ & \underbrace{p(y_1) + f(y_1, y_2, \ldots, y_s) - \langle b, z\rangle}_{\Phi(w)} \\ & + \underbrace{\langle x, \mathcal{F}^*y + \mathcal{G}^*z - c\rangle}_{\langle x, \mathcal{A}^*w - c\rangle} + \frac{\sigma}{2}\underbrace{\|\mathcal{F}^*y + \mathcal{G}^*z - c\|^2}_{\|\mathcal{A}^*w - c\|^2}, \\ & \qquad \forall w = (y, z) \in \mathcal{W} := \mathcal{Y}\times\mathcal{Z},\ x \in \mathcal{X}\end{aligned}$$

[1] Arrow, K.J., Solow, R.M.: Gradient methods for constrained maxima with weakened assumptions. In: Arrow, K.J., Hurwicz, L., Uzawa, H., (eds.) Studies in Linear and Nonlinear Programming. Stanford University Press, Stanford, pp. 165-176 (1958)

K. Arrow and R. Solow

Kenneth Joseph “Ken” Arrow
(23 August 1921 – 21 February 2017)

John Bates Clark Medal (1957); Nobel Prize in Economics (1972); von Neumann Theory Prize (1986); National Medal of Science (2004); ForMemRS (2006)



Robert Merton Solow
(August 23, 1924 –)

John Bates Clark Medal (1961); Nobel Memorial Prize in Economic Sciences (1987); National Medal of Science (1999); Presidential Medal of Freedom (2014); ForMemRS (2006)

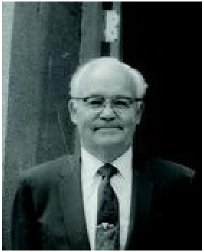
The augmented Lagrangian method[2] (ALM)

$\mathcal{L}_\sigma(y, z; x) = p(y_1) + f(y) - \langle b, z\rangle + \langle x, \mathcal{F}^* y + \mathcal{G}^* z - c\rangle + \frac{\sigma}{2}\|\mathcal{F}^* y + \mathcal{G}^* z - c\|^2$

Starting from $x^0 \in \mathcal{X}$, performs for $k = 0, 1, \ldots$

(1) $\underbrace{(y^{k+1}, z^{k+1})}_{w^{k+1}} \Leftarrow \min_{y,z} \mathcal{L}_\sigma(\underbrace{y, z}_{w}; x^k)$ (approximately)

(2) $x^{k+1} := x^k + \tau\,\sigma(\mathcal{F}^* y^{k+1} + \mathcal{G}^* z^{k+1} - c)$ with $\tau \in (0, 2)$



Magnus Rudolph Hestenes
(February 13 1906 – May 31 1991)

Michael James David Powell
(29 July 1936 – 19 April 2015)

[2]Also known as the method of multipliers

ALM to proximal ALM[3] (PALM)

Minimize the augmented Lagrangian function plus a quadratic **proximal term**:

$$w^{k+1} \approx \arg\min_{w} \mathcal{L}_\sigma(w; x^k) + \frac{1}{2}\|w - w^k\|_{\mathcal{D}}^2$$

- $\mathcal{D} = \sigma^{-1}\mathcal{I}$ in the seminal work of Rockafellar (in which inequality constraints are considered). Note that $\mathcal{D} \to 0$ as $\sigma \to \infty$, which is critical for asymptotically superlinear convergence (for $\tau = 1$)
- It is a primal-dual type proximal point algorithm (PPA)

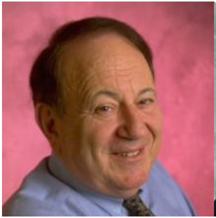
[3] Also known as the proximal method of multipliers

"Decoupling" (or "splitting") based ADMM

One the other hand, "decoupling" (or "splitting") based approach is available, i.e,

$$\begin{cases} y^{k+1} \approx \arg\min\limits_{y}\{\mathcal{L}_\sigma(y, z^k; x^k)\}, \ z^{k+1} \approx \arg\min\limits_{z}\{\mathcal{L}_\sigma(y^{k+1}, z; x^k)\}; \\ x^{k+1} := x^k + \tau\sigma(\mathcal{F}^* y^{k+1} + \mathcal{G}^* z^{k+1} - c), \quad \tau \in (\mathbf{0}, \infty) \end{cases}$$

- The **two-block ADMM**
- Allows $\tau \in (0, (1+\sqrt{5})/2)$ if the convergence of the full (primal & dual) sequence is required (first proven by Glowinski in 1977 at Tata Institute, India)
- The case with $\tau = 1$ is a kind of PPA (Gabay + Bertsekas-Eckstein)



An inexact majorized indefinite proximal ALM

Consider

$$\min_{w\in\mathcal{W}} \Phi(w) := \varphi(w) + h(w) \quad \text{s.t.} \quad \mathcal{A}^* w = c$$

- There exists a self-adjoint positive semidefinite linear operator $\widehat{\Sigma}^h : \mathcal{W} \to \mathcal{W}$, such that for any $w, w' \in \mathcal{W}$,

$$\boxed{h(w) \leq \hat{h}(w, w') := h(w') + \langle \nabla h(w'), w - w' \rangle + \frac{1}{2}\|w - w'\|^2_{\widehat{\Sigma}^h}}$$

  which is called a majorization (or surrogate) of h at w'

Prerequisites

One definition and one assumption

Let $\sigma > 0$. The majorized augmented Lagrangian function is defined, for any $(w, x, w') \in \mathcal{W} \times \mathcal{X} \times \mathcal{W}$, by

$$\widehat{\mathcal{L}}_\sigma(w; (x, w')) := \varphi(w) + \hat{h}(w, w') + \langle \mathcal{A}^* w - c, x \rangle + \frac{\sigma}{2} \|\mathcal{A}^* w - c\|^2$$

Assumption

The solution set K to the KKT system is nonempty and $\mathcal{D} : \mathcal{W} \to \mathcal{W}$ is a given self-adjoint linear operator such that

$$\frac{1}{2}\widehat{\Sigma}^h + \mathcal{D} \succeq 0 \quad \& \quad \frac{1}{2}\widehat{\Sigma}^h + \mathcal{D} + \sigma \mathcal{A}\mathcal{A}^* \succ 0 \tag{8}$$

- **$\mathcal{D}$ is not necessarily to be positive semidefinite!**

Alg. an inexact majorized indefinite proximal ALM

Let $\{\varepsilon_k\}$ be a summable sequence of nonnegative numbers. Choose an initial point $(x^0, w^0) \in \mathcal{X} \times \mathcal{W}$. For $k = 0, 1, \ldots,$

1 Compute

$$w^{k+1} \approx \arg\min_{w \in \mathcal{W}} \left\{ \widehat{\mathcal{L}}_\sigma(w; (x^k, w^k)) + \frac{1}{2}\|w - w^k\|_{\mathcal{D}}^2 \right\}$$

such that there exists d_k satisfying $\|d^k\| \leq \varepsilon_k$ and

$$d^k \in \partial_w \widehat{\mathcal{L}}_\sigma(w^{k+1}; (x^k, w^k)) + \mathcal{D}(w^{k+1} - w^k)$$

2 Update $x^{k+1} := x^k + \tau\sigma(\mathcal{A}^* w^{k+1} - c)$ with $\tau \in (0, 2)$

Theorem

The sequence $\{(x^k, w^k)\}$ generated by the above Algorithm converges to a solution to the KKT system.

Multi-block: Majorization and Splitting

There exists a self-adjoint linear operator $\widehat{\Sigma}^f \succeq 0$ on $\mathcal{Y}$ such that for any $y, y' \in \mathcal{Y}$,

$$f(y) \leq \hat{f}(y, y') := f(y') + \langle \nabla f(y'), y - y' \rangle + \frac{1}{2}\|y - y'\|^2_{\widehat{\Sigma}^f}$$

- Denote $y_{<i} := (y_1; \ldots; y_{i-1})$ and $y_{>i} := (y_{i+1}; \ldots; y_s)$
- Decompose $\widehat{\Sigma}^f = \begin{pmatrix} \widehat{\Sigma}^f_{11} & \widehat{\Sigma}^f_{12} & \cdots & \widehat{\Sigma}^f_{1s} \\ (\widehat{\Sigma}^f_{12})^* & \widehat{\Sigma}^f_{22} & \cdots & \widehat{\Sigma}^f_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ (\widehat{\Sigma}^f_{1s})^* & (\widehat{\Sigma}^f_{2s})^* & \cdots & \widehat{\Sigma}^f_{ss} \end{pmatrix}$

  with

$$\widehat{\Sigma}^f_{ij} : \mathcal{Y}_j \to \mathcal{Y}_i, \ \forall 1 \leq i \leq j \leq s$$

Basic Assumptions

(a) The self-adjoint linear operator $\mathcal{S} : \mathcal{Y} \to \mathcal{Y}$ satisfies

$$\boxed{\widehat{\Sigma}^f_{ii} + \sigma \mathcal{F}_i \mathcal{F}_i^* + \mathcal{S}_{ii} \succ 0 \quad \text{and} \quad \mathcal{S} \succeq -\tfrac{1}{2}\widehat{\Sigma}^f}$$

(b) The linear operator $\mathcal{G}$ is surjective (always true if restricted to its range space)

Let $\sigma > 0$. The *majorized* proximal augmented Lagrangian function:

$$\begin{aligned}
\widetilde{\mathcal{L}}_\sigma(y, z; (x, y')) := \quad & p(y_1) + \hat{f}(y, y') - \langle b, z \rangle \\
& + \langle \mathcal{F}^* y + \mathcal{G}^* z - c, x \rangle + \frac{\sigma}{2} \|\mathcal{F}^* y + \mathcal{G}^* z - c\|^2 \\
& + \frac{1}{2}\|y - y'\|^2_{\mathcal{S}}
\end{aligned}$$

The Algorithm: sGS-imPADMM

$(x^0, y^0, z^0) \in \mathcal{X} \times \operatorname{dom} p \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_s \times \mathcal{Z}$. $\{\tilde{\varepsilon}_k\}$ nonnegative and summable. For $k = 0, 1, \ldots,$

1 Compute for $i = s, \ldots, 2,$

$$y_i^{k+\frac{1}{2}} \approx \operatorname*{arg\,min}_{y_i \in \mathcal{Y}_i} \widetilde{\mathcal{L}}_\sigma\Big(y_{<i}^k, y_i, y_{>i}^{k+\frac{1}{2}}, z^k; (x^k, y^k)\Big)$$

2 Compute for $i = 1, \ldots, s,$

$$y_i^{k+1} \approx \operatorname*{arg\,min}_{y_i \in \mathcal{Y}_i} \widetilde{\mathcal{L}}_\sigma\Big(y_{<i}^{k+1}, y_i, y_{>i}^{k+1/2}, z^k; (x^k, y^k)\Big)$$

3 Compute

$$z^{k+1} \approx \operatorname*{arg\,min}_{z \in \mathcal{Z}} \widetilde{\mathcal{L}}_\sigma(y^{k+1}, z; (x^k, y^k))$$

4 Compute $x^{k+1} := x^k + \tau\sigma(\mathcal{F}^* y^{k+1} + \mathcal{G}^* z^{k+1} - c)$, $\tau \in (0, 2)$

Criteria for inexact solutions in sGS-imPADMM

1 For $i = s, \ldots, 2$, the approximate solution $y_i^{k+\frac{1}{2}}$ is chosen such that there exists $\tilde{\delta}_i^k$ satisfying $\|\tilde{\delta}_i^k\| \leq \tilde{\varepsilon}_k$ and

$$\tilde{\delta}_i^k \in \partial_{y_i} \widetilde{\mathcal{L}}_\sigma\Big(y_{<i}^k, y_i^{k+\frac{1}{2}}, y_{>i}^{k+\frac{1}{2}}, z^k; (x^k, y^k)\Big)$$

2 For $i = 1, \ldots, s$, the approximate solution y_i^{k+1} is chosen such that there exists δ_i^k satisfying $\|\delta_i^k\| \leq \tilde{\varepsilon}_k$ and

$$\delta_i^k \in \partial_{y_i} \widetilde{\mathcal{L}}_\sigma\Big(y_{<i}^{k+1}, y_i^{k+1}, y_{>i}^{k+1/2}, z^k; (x^k, y^k)\Big)$$

3 The approximate solution z^{k+1} is chosen such that $\|\gamma^k\| \leq \tilde{\varepsilon}_k$ with

$$\begin{aligned}\gamma^k : \;&= \nabla_z \widetilde{\mathcal{L}}_\sigma\Big(y^{k+1}, z^{k+1}; (x^k, y^k)\Big) \\ &= \mathcal{G}x^k - b + \sigma\mathcal{G}(\mathcal{F}^* y^{k+1} + \mathcal{G}^* z^{k+1} - c)\end{aligned}$$

Inexact block sGS decomposition

Define $\mathcal{H} := \widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{S} = \mathcal{H}_u + \mathcal{H}_d + \mathcal{H}_u^*$ with $\mathcal{H}_d := \text{Diag}(\mathcal{H}_{11}, \ldots, \mathcal{H}_{ss})$ and

$$\mathcal{H}_u := \begin{pmatrix} 0 & \mathcal{H}_{12} & \cdots & \mathcal{H}_{1s} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \mathcal{H}_{(s-1)s} \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \mathcal{H}_{ij} = \widehat{\Sigma}^f_{ij} + \sigma\mathcal{F}_i\mathcal{F}_j^* + \mathcal{S}_{ij}$$

For convenience, we denote for each $k \geq 0$,

$$\tilde{\delta}_1^k := \delta_1^k, \quad \tilde{\delta}^k := (\tilde{\delta}_1^k, \tilde{\delta}_2^k \ldots, \tilde{\delta}_s^k), \quad \delta^k := (\delta_1^k, \ldots, \delta_s^k)$$

Define the sequence $\{\Delta^k\} \in \mathcal{Y}$ by

$$\Delta^k := \delta^k + \mathcal{H}_u\mathcal{H}_d^{-1}(\delta^k - \tilde{\delta}^k)$$

Moreover, we can define the linear operator

$$\widehat{\mathcal{H}} := \mathcal{H}_u\mathcal{H}_d^{-1}\mathcal{H}_u^* \quad \textbf{(sGS Splitting Operator)}$$

Result by the block sGS decomposition theorem [4]

The iterate y^{k+1} in Step 2 of sGS-imPADMM is the unique solution to a proximal minimization problem given by

$$y^{k+1} = \arg\min_{y} \Big\{ \underbrace{\widehat{\mathcal{L}}_\sigma(y, z^k; (x^k, y^k)) + \frac{1}{2}\|y - y^k\|^2_{\mathcal{S}+\widehat{\mathcal{H}}}}_{\text{strongly convex}} - \langle \Delta^k, y \rangle \Big\}$$

- Recall that $\mathcal{H} := \widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{S}$
- Linearly transported error: $\Delta^k = \delta^k + \mathcal{H}_u \mathcal{H}_d^{-1}(\delta^k - \tilde{\delta}^k)$

[4] X.D. Li, D.F. Sun, and K.-C Toh, A block symmetric Gauss-Seidel decomposition theorem for convex composite quadratic programming and its applications, Math Prog (2019) [DOI: 10.1007/s10107-018-1247-7]

The equivalence property

Recall that $\mathcal{W} = \mathcal{Y} \times \mathcal{Z}$. Define $\widehat{\Sigma}^h : \mathcal{W} \to \mathcal{W}$ by

$$\widehat{\Sigma}^h := \begin{pmatrix} \widehat{\Sigma}^f & \\ & 0 \end{pmatrix}$$

For $w = (y; z)$ and $w' = (y'; z')$, denote

$$\widehat{\mathcal{L}}_\sigma(w; (x, w')) := \widehat{\mathcal{L}}_\sigma(y, z; (x, y'))$$

Define the error term

$$\widehat{\Delta}^k := \Delta^k - \mathcal{F}\mathcal{G}^*(\mathcal{G}\mathcal{G}^*)^{-1}(\gamma^{k-1} - \gamma^k - \mathcal{G}(x^{k-1} - x^k)) \in \mathcal{Y}$$

with the convention that

$$\begin{cases} x^{-1} := x^0 - \tau\sigma(\mathcal{F}^* y^0 + \mathcal{G}^* z^0 - c), \\ \gamma^{-1} := -b + \mathcal{G}x^{-1} + \sigma\mathcal{G}(\mathcal{F}^* y^0 + \mathcal{G}^* z^0 - c) \end{cases}$$

The equivalence property

Define the block-diagonal linear operator

$$\mathcal{T} := \begin{pmatrix} \mathcal{S} + \widehat{\mathcal{H}} + \sigma \mathcal{F}\mathcal{G}^*(\mathcal{G}\mathcal{G}^*)^{-1}\mathcal{G}\mathcal{F}^* & \\ & 0 \end{pmatrix} \quad \boxed{\mathcal{W} \to \mathcal{W}}$$

Theorem

Let $\{(x^k, w^k)\}$ with $w^k := (y^k; z^k)$ be the sequence generated by sGS-imPADMM. Then, for any $k \geq 0$, it holds that

(i) *the linear operators $\mathcal{T}$, $\mathcal{A}$ and $\widehat{\Sigma}^h$ satisfy*

$$\mathcal{T} + \frac{1}{2}\widehat{\Sigma}^h \succeq 0$$

(ii)

$$w^{k+1} \approx \arg\min_{w \in \mathcal{W}} \left\{ \widehat{\mathcal{L}}_\sigma\Big(w; (x^k, w^k)\Big) + \frac{1}{2}\|w - w^k\|^2_{\mathcal{T}} \right\}$$

in the sense that $(\widehat{\Delta}^k; \gamma^k) \in \partial_w \widehat{\mathcal{L}}_\sigma((w^{k+1}; (x^k, w^k)) + \mathcal{T}(w^{k+1} - w^k)$ and $\|(\widehat{\Delta}^k, \gamma^k)\| \leq \widehat{\varepsilon}_k$ with $\{\widehat{\varepsilon}_k\}$ being a summable sequence of nonnegative numbers

sGS-imPADMM convergence

One can readily get the following convergence theorem

Theorem

Suppose that

$$\frac{1}{2}\widehat{\Sigma}^f + \sigma\mathcal{F}\mathcal{F}^* + \mathcal{S} + \mathcal{H}_u\mathcal{H}_d^{-1}\mathcal{H}_u^* \succ 0$$

Then,

$$\mathcal{T} + \frac{1}{2}\widehat{\Sigma}^h + \sigma\mathcal{A}\mathcal{A}^* \succ 0$$

Moreover, the sequence $\{(x^k, y^k, z^k)\}$ *generated by the Algorithm converges to a solution of the KKT system of the problem. Thus,* $\{(y^k, z^k)\}$ *converges to a solution to this problem and* $\{x^k\}$ *converges to a solution of its dual*

The two-block case

Let $\mathcal{Y} = \mathcal{Y}_1$ and f be vacuous (e.g., the dual of linear conic programming), i.e.,

$$\min\{p(y) - \langle b, z\rangle \mid \mathcal{F}^* y + \mathcal{G}^* z = c\} \tag{9}$$

- The two-block ADMM originates from the ALM, but it actually deviates substantially from the ALM!!!
- ADMM (decoupling) is NOT ALM (recoupling)
- Note that $\mathcal{T}$ has a term propositional to σ while in Rockafellar's proximal ALM, the corresponding proximal term is proportional to σ^{-1}. This is the $\underbrace{\text{price to pay}}$ for "decoupling" — loss of the arbitrary linear convergence rate [in the terminology of M.J.D. Powell]

Comments on the two-block case

- The assumptions we made for problem (9) are apparently much weaker than those in original work of Gabay and Mercier[5], where $\mathcal{F}$ is assumed to be the identity operator and p is assumed to be strongly convex

- In Gabay and Mercier (1976), Theorem 3.1, only the convergence of the primal sequence $\{(y^k, z^k)\}$ is obtained while the dual sequence $\{x^k\}$ is only proven to be bounded

- In S., Toh and Yang *et al.*[6], a similar result to ours has been derived with the requirements that the initial multiplier x^0 satisfies $\mathcal{G}x^0 - b = 0$ and all the subproblems are solved exactly

---

[5] Gabay, D. and Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)

[6] Sun, D.F., Toh, K.-C. and Yang, L.Q.: A convergent proximal alternating direction method of multipliers for conic programming with 4-type constraints. SIAM J. Optim. **25**(2), 882–915 (2015)

Numerical Experiments

Solving dual linear SDP problems via the two-block ADMM with step-length taking values beyond the classic restriction of $(1+\sqrt{5})/2$

- To know to what extent the numerical efficiency of the ADMM can be improved if the equivalence proved in this paper is incorporated
- To see whether a step-length that is very close to 2 will lead to better or worse numerical performance

Solving $\min_X \{\langle C, X\rangle \mid \mathcal{A}X = b, X \in \mathbb{S}^n_+\}$

The dual is

$$\min_{Y,z} \left\{\delta_{\mathbb{S}^n_+}(Y) - \langle b, z\rangle \mid Y + \mathcal{A}^* z = C\right\}$$

Here $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is linear, $b \in \mathbb{R}^m$ and $C \in \mathbb{S}^n$ are given data

ADMM has been incorporated in solving dual SDP for more than a decade:

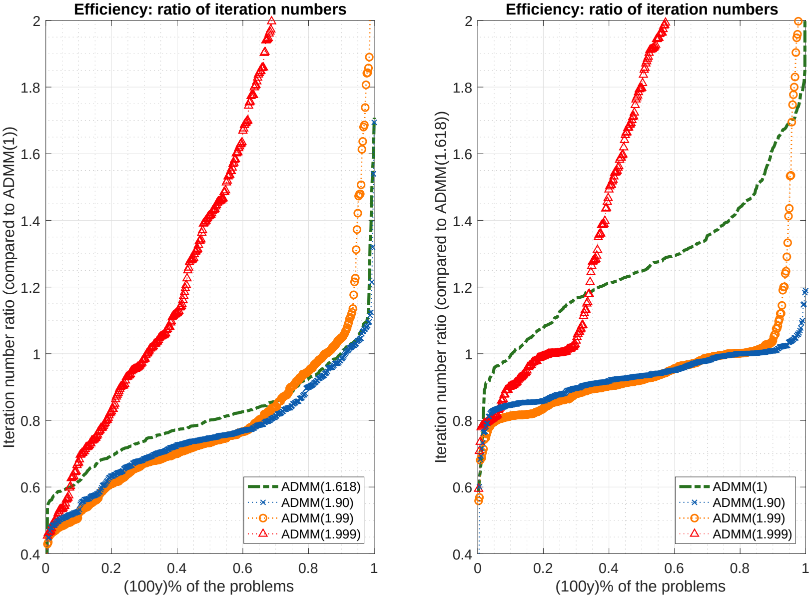
- ADMM with unit step-length was first employed in Povh *et al.* [Comput. 78 (2006)] under the name of boundary point method for solving the dual SDP (Later extended in Malick *et al.* [SIOPT 20 (2009)] with a convergence proof)
- ADMM was used in the software SDPNAL developed by Zhao et al. [SIOPT 20 (2010)] to warm-start a semismooth Newton ALM for dual SDP
- SDPAD by Wen *et al.* [MPC 2 (2010)]: ADMM solver on dual SDP (used SDPNAL template)

Numerical Experiments: details

- Five choices of the step-length, i.e., $\tau = 1$, $\tau = 1.618$, $\tau = 1.90$, $\tau = 1.99$ and $\tau = 1.999$
- Running the Matlab package SDPNAL+ (version 1.0)[7]
- 6 categories of SDP problems
- In general it is a good idea to use a step-length larger than 1, e.g., $\tau = 1.618$
- We can even set the step-length to be larger than 1.618, say $\tau = 1.9$, to get better numerical performance
- Stopping Criteria: DIMACS rule based on relative residuals of primal/dual feasibility and complementarity
- maximum number of iterations: 10^5

[7] awarded the triennial Beale-Orchard–Hays Prize for Excellence in Computational Mathematical Programming by the Mathematical Optimization Society in 2018

Numerical comparisons



Conclusions

- A block sGS decomposition based (exact or inexact) multi-block majorized (proximal or not) ADMM is equivalent to an inexact majorized proximal ALM with $\tau \in (0, 2)$

- ADMM can achieve better numerical performance if the step-length is larger than the conventional upper bound of $(1 + \sqrt{5})/2$ but not too close to 2. It also justifies the safety and effectiveness of choosing $\tau = 1.618$

- The proximal ALM interpretation of the ADMM may explain why it often converges slowly after the initial iterations [the automatically generated proximal term (hidden) is too large]

“Recoupling”?

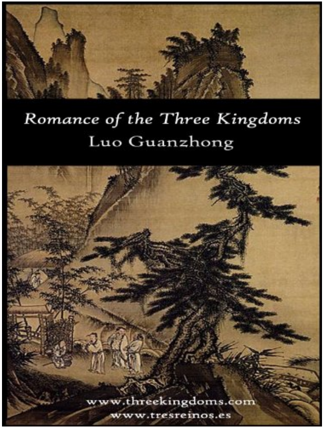
- ALM $\Longrightarrow$ ADMM $\quad\Longleftrightarrow\quad$ “Coupling” $\Longrightarrow$ “Decoupling”

- For big challenging problems, it is time for “Recoupling”?

Any Reason?

天下大事 分久必合 合久必分

罗贯中 《三国演义》



World under heaven, after a long period of division, tends to unite; after a long period of union, tend to divide. This has been so since antiquity.

From **"Romance of the Three Kingdoms"**

a 14th-century historical novel

by **Guanzhong Luo** (Author)

`www.threekingdoms.com` (Editor)

`www.tresreinos.es` (Editor)

C.H. Brewitt Taylor (Translator)