ADMM for Deterministic and Convex Optimization

Hailiang ZHAO @ ZJU-CS

http://hliangzhao.me

November 1, 2022

Outline

Outline

Review the Vanilla ADMM

The vanilla version of ADMM is for solving the following problem:

$$\min_{\boldsymbol{x},\boldsymbol{y}} \quad f(\boldsymbol{x}) + g(\boldsymbol{y}), \quad s.t. \quad \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} = \boldsymbol{b}. \tag{1}$$

ADMM solves it with the following iterations:

$$\boldsymbol{x}^{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} L_\beta(\boldsymbol{x}, \boldsymbol{y}^k, \boldsymbol{\lambda}) \tag{2}$$

$$\boldsymbol{y}^{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} L_\beta(\boldsymbol{x}^{k+1}, \boldsymbol{y}, \boldsymbol{\lambda}) \tag{3}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \beta\big(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}\big). \tag{4}$$

Building Blocks for Convergence Analysis

Lemma I.1

Suppose that $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex. Let $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*)$ be a KKT point of (1), then $\forall \boldsymbol{x}, \boldsymbol{y}$, we have

$$f(\boldsymbol{x}) + g(\boldsymbol{y}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) + \langle \boldsymbol{\lambda}^*, \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b} \rangle \geq 0. \quad (5)$$

Proof.

The result is immediate with Proposition B.10.1 (every KKT point is a saddle point of the Lagrangian function). $\square$

Building Blocks for Convergence Analysis

Lemma I.2

Suppose that $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex. Let $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*)$ be a KKT point of (1). If

$$f(\boldsymbol{x}) + g(\boldsymbol{y}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) + \langle \boldsymbol{\lambda}^*, \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b} \rangle \leq \alpha_1 \quad (6)$$

$$\|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b}\| \leq \alpha_2, \quad (7)$$

then we have

$$-\|\boldsymbol{\lambda}^*\|\alpha_2 \leq f(\boldsymbol{x}) + g(\boldsymbol{y}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) \leq \|\boldsymbol{\lambda}^*\|\alpha_2 + \alpha_1. \quad (8)$$

Proof.

The result is immediate with Lemma I.1. $\square$

Building Blocks for Convergence Analysis

Lemma I.3

For ADMM, we have

$$
\begin{aligned}
\mathbf{0} &\in \partial f(\boldsymbol{x}^{k+1}) + \mathbf{A}^T\boldsymbol{\lambda}^k + \beta\mathbf{A}^T(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), && (9)\\
\mathbf{0} &\in \partial g(\boldsymbol{y}^{k+1}) + \mathbf{B}^T\boldsymbol{\lambda}^k + \beta\mathbf{B}^T(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}), && (10)\\
\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k &= \beta\left(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}\right), && (11)\\
\mathbf{0} &\in \partial f(\boldsymbol{x}^*) + \mathbf{A}^T\boldsymbol{\lambda}^*, && (12)\\
\mathbf{0} &\in \partial g(\boldsymbol{y}^*) + \mathbf{B}^T\boldsymbol{\lambda}^*, && (13)\\
\mathbf{A}\boldsymbol{x}^* + \mathbf{B}\boldsymbol{y}^* &= \boldsymbol{b}. && (14)
\end{aligned}
$$

Proof.

(9) and (10) can be derived from the Proximal Point Method (*formula (22) in the From Dual Descent to ADMM slide*). (11) is from (4). (12) - (14) are the KKT conditions. □

Building Blocks for Convergence Analysis

Based on Lemma I.3, we define two vectors:

$$\hat{\nabla} f(\boldsymbol{x}^{k+1}) = -\mathbf{A}^T \boldsymbol{\lambda}^k - \beta \mathbf{A}^T (\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \tag{15}$$

$$\begin{aligned} \hat{\nabla} g(\boldsymbol{y}^{k+1}) &= -\mathbf{B}^T \boldsymbol{\lambda}^k - \beta \mathbf{A}^T (\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}) \\ &= -\mathbf{B}^T \boldsymbol{\lambda}^{k+1}. \end{aligned} \tag{16}$$

Then we have

$$\hat{\nabla} f(\boldsymbol{x}^{k+1}) \in \partial f(\boldsymbol{x}^{k+1}), \tag{17}$$

$$\hat{\nabla} g(\boldsymbol{y}^{k+1}) \in \partial g(\boldsymbol{y}^{k+1}). \tag{18}$$

Lemma I.4

For ADMM, we have

$$\langle \hat{\nabla} g(\boldsymbol{y}^{k+1}), \boldsymbol{y}^{k+1} - \boldsymbol{y} \rangle = -\langle \boldsymbol{\lambda}^{k+1}, \mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y} \rangle, \forall \boldsymbol{y}, \tag{19}$$

and

$$\begin{aligned}
&\langle \hat{\nabla} f(\boldsymbol{x}^{k+1}), \boldsymbol{x}^{k+1} - \boldsymbol{x} \rangle + \langle \hat{\nabla} g(\boldsymbol{y}^{k+1}), \boldsymbol{y}^{k+1} - \boldsymbol{y} \rangle \\
&= -\langle \boldsymbol{\lambda}^{k+1}, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{A}\boldsymbol{x} - \mathbf{B}\boldsymbol{y} \rangle \\
&\quad + \beta \langle \mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^{k}, \mathbf{A}\boldsymbol{x}^{k+1} - \mathbf{A}\boldsymbol{x} \rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y}.
\end{aligned} \tag{20}$$

Proof.

The results are immediate with (11), (15), and (16). □

Lemma I.5

Suppose that $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex. Then for ADMM, we have

$$
\begin{aligned}
\langle \hat{\nabla} f(\boldsymbol{x}^{k+1}), \boldsymbol{x}^{k+1} - \boldsymbol{x}^* \rangle + \langle \hat{\nabla} g(\boldsymbol{y}^{k+1}), \boldsymbol{y}^{k+1} - \boldsymbol{y}^* \rangle \\
+ \langle \boldsymbol{\lambda}^*, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b} \rangle \\
\leq \frac{1}{2\beta}\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 - \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2 \\
+ \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^k - \mathbf{B}\boldsymbol{y}^*\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^*\|^2 \\
- \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^k\|^2. \quad (21)
\end{aligned}
$$

Proof.

Use above lemmas and the monotonicity of ∂g to prove it. $\square$

Building Blocks for Convergence Analysis

Lemma I.6

Suppose that $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex. Then for ADMM, we have

$$\begin{aligned} f(\boldsymbol{x}^{k+1}) + g(\boldsymbol{y}^{k+1}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) + \langle \boldsymbol{\lambda}^*, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b} \rangle \\ \leq \frac{1}{2\beta}\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 - \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2 \\ + \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^k - \mathbf{B}\boldsymbol{y}^*\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^*\|^2 \\ - \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^k\|^2. \end{aligned} \tag{22}$$

Building Blocks for Convergence Analysis

Lemma I.6 (Cont'd)

If we further assume that $g(\boldsymbol{y})$ is μ-strongly convex, then we have

$$\begin{aligned} f(\boldsymbol{x}^{k+1}) + g(\boldsymbol{y}^{k+1}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) + \langle \boldsymbol{\lambda}^*, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b} \rangle & \\ \leq \frac{1}{2\beta}\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 - \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2 & \\ + \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^k - \mathbf{B}\boldsymbol{y}^*\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^*\|^2 & \\ - \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^k\|^2 & \\ - \frac{\mu}{2}\|\boldsymbol{y}^{k+1} - \boldsymbol{y}^*\|^2. & \quad (23) \end{aligned}$$

Building Blocks for Convergence Analysis

Lemma I.6 (Cont'd)

If we further assume that $g(\boldsymbol{y})$ is L-smooth convex, then we have

$$
\begin{aligned}
f(\boldsymbol{x}^{k+1}) + g(\boldsymbol{y}^{k+1}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) + \langle \boldsymbol{\lambda}^*, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b} \rangle \\
\leq \frac{1}{2\beta}\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 - \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2 \\
+ \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^k - \mathbf{B}\boldsymbol{y}^*\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^*\|^2 \\
- \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^k\|^2 \\
- \frac{1}{2L}\|\nabla g(\boldsymbol{y}^{k+1}) - \nabla g(\boldsymbol{y}^*)\|^2. \qquad (24)
\end{aligned}
$$

Building Blocks for Convergence Analysis

Proof Skecth of Lemma I.6

- (22) is immediate with (17), (18), Lemma I.5, and the definition of subgradient of convex functions.
- (23) and (24) can be obtained based on (22) and Proposition B.2 and Proposition B.4, respectively. Each of them adds a special term to the LHS of (22).

The Convergence of ADMM

When $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex, the convergenece of ADMM exists.

Theorem I.1

Suppose that *$f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex*. Then for ADMM, we have

$$f(\boldsymbol{x}^{k+1}) - f(\boldsymbol{x}^*) + g(\boldsymbol{y}^{k+1}) - g(\boldsymbol{y}^*) \to 0, \tag{25}$$

$$\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b} \to \mathbf{0}. \tag{26}$$

The Convergence of ADMM

Proof Sketch of Theorem I.1

Combing Lemma I.1 and (22) we have

$$
\begin{aligned}
&\frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^{k}\|^2+\frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1}-\mathbf{B}\boldsymbol{y}^{k}\|^2\\
&\leq \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k}-\boldsymbol{\lambda}^{*}\|^2-\frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^{*}\|^2\\
&+\frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k}-\mathbf{B}\boldsymbol{y}^{*}\|^2-\frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1}-\mathbf{B}\boldsymbol{y}^{*}\|^2.
\end{aligned}
\tag{27}
$$

Summing over $k=0,...,\infty$, we have

$$
\begin{aligned}
&\sum_{k=0}^{\infty}\left(\frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^{k}\|^2+\frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1}-\mathbf{B}\boldsymbol{y}^{k}\|^2\right)\\
&\qquad\leq \frac{1}{2\beta}\|\boldsymbol{\lambda}^{0}-\boldsymbol{\lambda}^{*}\|^2+\frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{0}-\mathbf{B}\boldsymbol{y}^{*}\|^2.
\end{aligned}
\tag{28}
$$

The Convergence of ADMM

Proof Sketch of Theorem I.1

Note that the RHS of (28) is a constant, thus we have

$$\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k \to \mathbf{0}, \quad \mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^k \to 0. \tag{29}$$

And, $\frac{1}{2\beta}\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 + \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^k - \mathbf{B}\boldsymbol{y}^*\|^2$ must be a non-increasing sequence. So $\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2$ and $\|\mathbf{B}\boldsymbol{y}^k - \mathbf{B}\boldsymbol{y}^*\|^2$ are bounded for all k. Then we have $\|\boldsymbol{\lambda}^k\|$ is bounded for all k. Since

$$\begin{aligned}\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k &= \beta\left(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}\right) \\ &= \beta(\mathbf{A}\boldsymbol{x}^{k+1} - \mathbf{A}\boldsymbol{x}^*) + \beta(\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^*),\end{aligned} \tag{30}$$

We know that $\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}$ and $\mathbf{A}\boldsymbol{x}^{k+1} - \mathbf{A}\boldsymbol{x}^*$ are also bounded.

The Convergence of ADMM

Proof Sketch of Theorem I.1

From (17), (18), (20), and the definition of subgradient of convex functions, we have

$$\begin{aligned} & f(\boldsymbol{x}^{k+1}) - f(\boldsymbol{x}^*) + g(\boldsymbol{y}^{k+1}) - g(\boldsymbol{y}^*) \\ & \leq -\langle \boldsymbol{\lambda}^{k+1}, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b} \rangle \\ & + \beta \langle \mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^k, \mathbf{A}\boldsymbol{x}^{k+1} - \mathbf{A}\boldsymbol{x}^* \rangle \to 0. \end{aligned} \tag{31}$$

On the other hand, from (12), (13), (14), and the definition of subgradient of convex functions, we have

$$\begin{aligned} & f(\boldsymbol{x}^{k+1}) - f(\boldsymbol{x}^*) + g(\boldsymbol{y}^{k+1}) - g(\boldsymbol{y}^*) \\ & \geq \langle -\mathbf{A}^T \boldsymbol{\lambda}^*, \boldsymbol{x}^{k+1} - \boldsymbol{x}^* \rangle + \langle -\mathbf{B}^T \boldsymbol{\lambda}^*, \boldsymbol{y}^{k+1} - \boldsymbol{y}^* \rangle \\ & = -\langle \boldsymbol{\lambda}^*, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b} \rangle \to 0. \end{aligned} \tag{32}$$

Then we have (25).

Convergence Rates of ADMM

The following several pages include:

- **Sublinear Convergence Rate**
  (1) Non-ergodic convergence rate
  (2) Ergodic convergence rate
- **Linear Convergence Rate**
  (1) Under strong convexity and smoothness assumption
  (2) Under error bound condition

Sublinear Non-Ergodic Convergence Rate

Lemma I.7

Suppose that $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex, then for ADMM we have

$$
\begin{aligned}
&\frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^{k}\|^2+\frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1}-\mathbf{B}\boldsymbol{y}^{k}\|^2\\
&\le \frac{1}{2\beta}\|\boldsymbol{\lambda}^{k}-\boldsymbol{\lambda}^{k-1}\|^2+\frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k}-\mathbf{B}\boldsymbol{y}^{k-1}\|^2.
\end{aligned} \tag{33}
$$

This lemma will be used for the following Theorem I.2 and Theorem I.3.

Sublinear Non-Ergodic Convergence Rate

Theorem I.2

Suppose that *$f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex*, then for ADMM we have

$$-\|\boldsymbol{\lambda}^*\|\sqrt{\frac{C}{\beta(K+1)}} \le f(\boldsymbol{x}^{K+1}) + g(\boldsymbol{y}^{K+1}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) \le \frac{C}{K+1} + \frac{2C}{\sqrt{K+1}} + \|\boldsymbol{\lambda}^*\|\sqrt{\frac{C}{\beta(K+1)}}, \tag{34}$$

where

$$C := \frac{1}{\beta}\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|^2 + \beta\|\mathbf{B}\boldsymbol{y}^0 - \mathbf{B}\boldsymbol{y}^*\|^2. \tag{35}$$

Sublinear Ergodic Convergence Rate

Theorem I.3

Suppose that *$f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex*, then for ADMM we have

$$\begin{aligned} & |f(\hat{\boldsymbol{x}}^{K+1}) + g(\hat{\boldsymbol{y}}^{K+1}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*)| \\ & \qquad \leq \frac{C}{2(K+1)} + \frac{2\sqrt{C}\|\boldsymbol{\lambda}^*\|}{\sqrt{\beta}(K+1)}, \end{aligned} \tag{36}$$

$$\|\mathbf{A}\hat{\boldsymbol{x}}^{K+1} + \mathbf{B}\hat{\boldsymbol{y}}^{K+1} - \boldsymbol{b}\| \leq \frac{2\sqrt{C}}{\sqrt{\beta}(K+1)}, \tag{37}$$

where

$$\hat{\boldsymbol{x}}^{K+1} = \frac{1}{K+1}\sum_{k=1}^{K+1} \boldsymbol{x}^k, \quad \hat{\boldsymbol{y}}^{K+1} = \frac{1}{K+1}\sum_{k=1}^{K+1} \boldsymbol{y}^k. \tag{38}$$

Linear Convergence Rate with Assumption #1

Theorem I.4

Suppose that *$f(\boldsymbol{x})$ is convex and $g(\boldsymbol{y})$ is μ-strongly convex and L-smooth*. Assume that $\forall \boldsymbol{\lambda}$, $\|\mathbf{B}^T\boldsymbol{\lambda}\| \geq \sigma\|\boldsymbol{\lambda}\|$, where $\sigma > 0$. Let $\beta = \frac{\sqrt{\mu L}}{\sigma\|\mathbf{B}\|}$. Then we have

$$\begin{aligned} &\frac{1}{2\beta}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^*\|^2 + \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^*\|^2 \\ \leq &\left(1 + \frac{1}{2}\sqrt{\frac{\mu}{L}}\frac{\sigma}{\|\mathbf{B}\|}\right)^{-1}\left(\frac{1}{2\beta}\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 + \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y}^k - \mathbf{B}\boldsymbol{y}^*\|^2\right). \end{aligned} \tag{39}$$

Linear Convergence Rate with Assumption #2

Now we demonstrate the linear convergence rate of ADMM under the error bound condition. Firstly, we define

$$\phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) := \begin{pmatrix} \partial f(\boldsymbol{x}) + \mathbf{A}^T \boldsymbol{\lambda} \\ \partial g(\boldsymbol{y}) + \mathbf{B}^T \boldsymbol{\lambda} \\ \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b} \end{pmatrix}. \tag{40}$$

Correspondingly,

$$\phi^{-1}(\boldsymbol{s}) = \left\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) \mid \boldsymbol{s} \in \phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})\right\}. \tag{41}$$

Obviously, $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$ is a KKT point *iff* $\mathbf{0} \in \phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$.

Linear Convergence Rate with Assumption #2

Definition I.1 *

The set-value mapping $\phi(\boldsymbol{w})$ is called as satisfying the (global) error bound condition, if there exists constant $\kappa > 0$ such that

$$\mathrm{dist}_{\mathbf{H}}(\boldsymbol{w}, \phi^{-1}(\mathbf{0})) \leq \kappa \mathrm{dist}(\mathbf{0}, \phi(\boldsymbol{w})), \quad \forall \boldsymbol{w}, \tag{42}$$

where

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta \mathbf{B}^T \mathbf{B} & 0 \\ 0 & 0 & \frac{1}{\beta}\mathbf{I} \end{pmatrix} \tag{43}$$

and

$$\mathrm{dist}_{\mathbf{H}}(\boldsymbol{w}, \phi^{-1}(\mathbf{0})) = \min_{\boldsymbol{w}^* \in \phi^{-1}(0)} \|\boldsymbol{w} - \boldsymbol{w}^*\|_{\mathbf{H}}. \tag{44}$$

Note that $\|\boldsymbol{x}\|_{\mathbf{A}}^2 := \boldsymbol{x}^T \mathbf{A} \boldsymbol{x}$.

Linear Convergence Rate with Assumption #2

Theorem I.5

Suppose that *$f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex and $\phi(\boldsymbol{w})$ satisfies the error bound condition*. Then for ADMM, we have

$$\operatorname{dist}^2_{\mathbf{H}}\left(\left(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{\lambda}^{k+1}\right), \phi^{-1}(\mathbf{0})\right)$$
$$\leq \left[1 + \frac{1}{\kappa^2(\beta\|\mathbf{A}\|_2^2 + \frac{1}{\beta})}\right]^{-1} \operatorname{dist}^2_{\mathbf{H}}\left((\boldsymbol{x}^k, \boldsymbol{y}^k, \boldsymbol{\lambda}^k), \phi^{-1}(\mathbf{0})\right). \quad (45)$$

Outline

Bregman ADMM

ADMM solves two time-consuming subproblems to update $\boldsymbol{x}$ and $\boldsymbol{y}$. The Bregman ADMM uses the linearization technique to make the subproblems compuatationally efficient. It works with the following iterations:

$$
\begin{aligned}
\boldsymbol{x}^{k+1} &= \operatorname*{argmin}_{\boldsymbol{x}} \Big(f(\boldsymbol{x}) + \mathrm{g}(\boldsymbol{y}^k) + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b} \rangle \\
&\qquad + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 + D_\phi(\boldsymbol{x}, \boldsymbol{x}^k) \Big), \qquad (46) \\
\boldsymbol{y}^{k+1} &= \operatorname*{argmin}_{\boldsymbol{y}} \Big(f(\boldsymbol{x}^{k+1}) + \mathrm{g}(\boldsymbol{y}) + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b} \rangle \\
&\qquad + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b}\|^2 + D_\Psi(\boldsymbol{y}, \boldsymbol{y}^k) \Big), \qquad (47) \\
\boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k + \beta\big(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}\big). \qquad \text{(unchanged)} \qquad (48)
\end{aligned}
$$

$D_f(\cdot, \cdot)$ is the Bregman distance w.r.t. f.

Linearized ADMM (LADMM-1)

When

$$\phi(\boldsymbol{x}) = \frac{\beta\|\mathbf{A}\|_2^2}{2}\|\boldsymbol{x} - \boldsymbol{u}_1\|^2 - \frac{\beta}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{u}_2\|^2, \tag{49}$$

$$\Psi(\boldsymbol{y}) = \frac{\beta\|\mathbf{B}\|_2^2}{2}\|\boldsymbol{y} - \boldsymbol{v}_1\|^2 - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y} - \boldsymbol{v}_2\|^2, \tag{50}$$

where $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ ($i = 1, 2$) are any constant vectors, we have

$$D_\phi(\boldsymbol{x}, \boldsymbol{x}') = \frac{\beta\|\mathbf{A}\|_2^2}{2}\|\boldsymbol{x} - \boldsymbol{x}'\|^2 - \frac{\beta}{2}\|\mathbf{A}(\boldsymbol{x} - \boldsymbol{x}')\|^2, \tag{51}$$

$$D_\Psi(\boldsymbol{y}, \boldsymbol{y}') = \frac{\beta\|\mathbf{B}\|_2^2}{2}\|\boldsymbol{y} - \boldsymbol{y}'\|^2 - \frac{\beta}{2}\|\mathbf{B}(\boldsymbol{y} - \boldsymbol{y}')\|^2, \tag{52}$$

Linearized ADMM (LADMM-1)

Based on (51) and (52), (46) is reduced to

$$\begin{aligned}\boldsymbol{x}^{k+1} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \Big(& f(\boldsymbol{x}) + g(\boldsymbol{y}^k) + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b} \rangle \\ & + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 + \beta \langle \mathbf{A}^T(\mathbf{A}\boldsymbol{x}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \boldsymbol{x} - \boldsymbol{x}^k \rangle \\ & + \frac{\beta \|\mathbf{A}\|_2^2}{2} \|\boldsymbol{x} - \boldsymbol{x}^k\|^2 \Big), \end{aligned} \tag{53}$$

which is equal to

$$\boldsymbol{x}^{k+1} = \operatorname{prox}_{(\beta \|\mathbf{A}\|_2^2)^{-1} f} \left(\boldsymbol{x}^k - \frac{\mathbf{A}^T}{\beta \|\mathbf{A}\|_2^2} \tilde{\boldsymbol{\lambda}}_1^k \right), \tag{54}$$

where $\tilde{\boldsymbol{\lambda}}_1^k := \boldsymbol{\lambda}^k + \beta(\mathbf{A}\boldsymbol{x}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b})$, and $\operatorname{prox}_{tf}(\boldsymbol{v}) := \operatorname{argmin}_{\boldsymbol{x}} (f(x) + \frac{1}{2t} \|\boldsymbol{x} - \boldsymbol{v}\|^2)$ is the proximal operator.

Linearized ADMM (LADMM-1)

Based on (51) and (52), (47) is reduced to

$$\begin{aligned}
\boldsymbol{y}^{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} \Big(& f(\boldsymbol{x}^{k+1}) + g(\boldsymbol{y}) + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b} \rangle \\
& + \frac{\beta}{2}\|\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 \\
& + \beta\langle \mathbf{B}^T(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \boldsymbol{y} - \boldsymbol{y}^k \rangle \\
& + \frac{\beta\|\mathbf{B}\|_2^2}{2}\|\boldsymbol{y} - \boldsymbol{y}^k\|^2 \Big),
\end{aligned} \tag{55}$$

which is equal to

$$\boldsymbol{y}^{k+1} = \operatorname{prox}_{(\beta\|\mathbf{B}\|_2^2)^{-1} f}\left(\boldsymbol{y}^k - \frac{\mathbf{B}^T}{\beta\|\mathbf{B}\|_2^2} \tilde{\boldsymbol{\lambda}}_2^k \right), \tag{56}$$

where $\tilde{\boldsymbol{\lambda}}_2^k := \boldsymbol{\lambda}^k + \beta(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b})$.

Linearized ADMM (LADMM-1)

Note that

$$\frac{\beta}{2}\|\mathbf{A}\boldsymbol{x}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 + \beta\langle \mathbf{A}^T(\mathbf{A}\boldsymbol{x}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \boldsymbol{x} - \boldsymbol{x}^k\rangle$$

is the linear approximations of $\frac{\beta}{2}\|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2$ at $\boldsymbol{x}^k$, and

$$\frac{\beta}{2}\|\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 + \beta\langle \mathbf{B}^T(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \boldsymbol{y} - \boldsymbol{y}^k\rangle$$

is the linear approximations of $\frac{\beta}{2}\|\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b}\|^2$ at $\boldsymbol{y}^k$.
Thus we call (54), (56), and (48) *linearized ADMM* (LADMM-1).

In many cases, the proximal mappings of f and g are easily computable. For example, the proximal mappings of l_1-norm, l_2-norm, and matrix operator norm and nuclear norm all have closed-form solutions.

Linearized ADMM (LADMM-2)

When the proximal mappings of f and g are not easily computable, but f and g are L_f-smooth and L_g-smooth, respectively, we may choose

$$\phi(\boldsymbol{x}) = \frac{L_f + \beta\|\mathbf{A}\|^2}{2}\|\boldsymbol{x} - \boldsymbol{u}_1\|^2 - f(\boldsymbol{x}) - \frac{\beta}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{u}_2\|^2 \quad (57)$$

$$\Psi(\boldsymbol{y}) = \frac{L_g + \beta\|\mathbf{B}\|^2}{2}\|\boldsymbol{y} - \boldsymbol{v}_1\|^2 - g(\boldsymbol{y}) - \frac{\beta}{2}\|\mathbf{B}\boldsymbol{y} - \boldsymbol{v}_2\|^\cdot \quad (58)$$

Linearized ADMM (LADMM-2)

Then we have

$$D_\phi(\boldsymbol{x}, \boldsymbol{x}') = \frac{L_f + \beta\|\mathbf{A}\|^2}{2}\|\boldsymbol{x} - \boldsymbol{x}'\|^2 - f(\boldsymbol{x}) + f(\boldsymbol{x}') \\ + \langle \nabla f(\boldsymbol{x}'), \boldsymbol{x} - \boldsymbol{x}' \rangle - \frac{\beta}{2}\|\mathbf{A}(\boldsymbol{x} - \boldsymbol{x}')\|^2, \tag{59}$$

$$D_\Psi(\boldsymbol{y}, \boldsymbol{y}') = \frac{L_g + \beta\|\mathbf{B}\|^2}{2}\|\boldsymbol{y} - \boldsymbol{y}'\|^2 - g(\boldsymbol{y}) + g(\boldsymbol{y}') \\ + \langle \nabla g(\boldsymbol{y}'), \boldsymbol{y} - \boldsymbol{y}' \rangle - \frac{\beta}{2}\|\mathbf{B}(\boldsymbol{y} - \boldsymbol{y}')\|^2, \tag{60}$$

which are also independent of $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ ($i = 1, 2$).

Linearized ADMM (LADMM-2)

Correspondingly, (46) is reduced to

$$
\begin{aligned}
\boldsymbol{x}^{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \Big(& f(\boldsymbol{x}^k) + \langle \nabla f(\boldsymbol{x}^k), \boldsymbol{x} - \boldsymbol{x}^k \rangle \\
& + \mathrm{g}(\boldsymbol{y}^k) + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b} \rangle \\
& + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 + \beta \langle \mathbf{A}^T(\mathbf{A}\boldsymbol{x}^k + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \boldsymbol{x} - \boldsymbol{x}^k \rangle \\
& + \frac{L_f + \beta\|\mathbf{A}\|_2^2}{2} \|\boldsymbol{x} - \boldsymbol{x}^k\|^2 \Big) \\
= \ & \boldsymbol{x}^k - \left(L_f + \beta\|\mathbf{A}\|_2^2 \right)^{-1} \left\{ \nabla f(\boldsymbol{x}^k) + \mathbf{A}^T \tilde{\boldsymbol{\lambda}}_1^k \right\}.
\end{aligned}
\tag{61}
$$

The purple part is the linear approximation at $\boldsymbol{x}^k$ of

$$
f(\boldsymbol{x}) + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2.
$$

Linearized ADMM (LADMM-2)

Correspondingly, (47) is reduced to

$$\begin{aligned}
\boldsymbol{y}^{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} \Big(& f(\boldsymbol{x}^{k+1}) + \mathrm{g}(\boldsymbol{y}^k) + \langle \nabla \mathrm{g}(\boldsymbol{y}^k), \boldsymbol{y} - \boldsymbol{y}^k \rangle \\
& + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b} \rangle + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 \\
& + \beta \langle \mathbf{B}^T (\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \boldsymbol{y} - \boldsymbol{y}^k \rangle \\
& + \frac{L_{\mathrm{g}} + \beta \|\mathbf{B}\|_2^2}{2} \|\boldsymbol{y} - \boldsymbol{y}^k\|^2 \Big) \\
= {} & \boldsymbol{y}^k - \left(L_{\mathrm{g}} + \beta \|\mathbf{B}\|_2^2 \right)^{-1} \left\{ \nabla \mathrm{g}(\boldsymbol{y}^k) + \mathbf{B}^T \tilde{\boldsymbol{\lambda}}_2^k \right\}.
\end{aligned} \tag{62}$$

The purple part is the linear approximation at $\boldsymbol{y}^k$ of

$$\mathrm{g}(\boldsymbol{y}) + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b}\|^2.$$

We call (61), (62), and (48) *LADMM-2*.

Sublinear Ergodic Convergence of Bregman ADMM

The ergodic convergence rate of Bregman ADMM ——

Theorem II.1

Suppose that *$f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are convex*. Then for Bregman ADMM, we have

$$|f(\hat{\boldsymbol{x}}^K) + g(\hat{\boldsymbol{y}}^K) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*)| \leq \frac{D}{2(K)} + \frac{2\sqrt{D}\|\boldsymbol{\lambda}^*\|}{\sqrt{\beta}K}, \tag{63}$$

$$\|\mathbf{A}\hat{\boldsymbol{x}}^K + \mathbf{B}\hat{\boldsymbol{y}}^K - \boldsymbol{b}\| \leq \frac{2\sqrt{D}}{\sqrt{\beta}K}, \tag{64}$$

where $\hat{x}^K = \frac{1}{K}\sum_{k=1}^K \boldsymbol{x}^k$, $\hat{y}^K = \frac{1}{K}\sum_{k=1}^K \boldsymbol{y}^k$, and

$$D = \frac{1}{\beta}\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|^2 + \beta\|\mathbf{B}\boldsymbol{y}^0 - \mathbf{B}\boldsymbol{y}^*\|^2 + 2D_\phi(\boldsymbol{x}^*, \boldsymbol{x}^0) + 2D_\Psi(\boldsymbol{y}^*, \boldsymbol{y}^0). \tag{65}$$

Sublinear Non-Ergodic Convergence of Bregman ADMM

The non-ergodic convergence rate of Bregman ADMM ——

Theorem II.2

Suppose that *f and g are both generally convex*. Let *$\Psi = 0$ and $D_\phi(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbf{M}}^2$ for some symmetric and positive semidefinite matrix $\mathbf{M}$*. Then for Bregman ADMM we have

$$
\begin{aligned}
-\|\boldsymbol{\lambda}^*\|\sqrt{\frac{C}{\beta K}} &\leq f(\boldsymbol{x}^K) + g(\boldsymbol{y}^K) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) \\
&\leq \frac{C}{K} + \frac{3C}{\sqrt{K}} + \|\boldsymbol{\lambda}^*\|\sqrt{\frac{C}{\beta K}}, \qquad (66)
\end{aligned}
$$

$$
\|\mathbf{A}\boldsymbol{x}^K + \mathbf{B}\boldsymbol{y}^K - \boldsymbol{b}\| \leq \sqrt{\frac{C}{\beta K}}, \qquad (67)
$$

where

$$
C = \frac{1}{\beta}\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|^2 + \beta\|\mathbf{B}\boldsymbol{y}^0 - \mathbf{B}\boldsymbol{y}^*\|^2 + \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_{\mathbf{M}}^2. \qquad (68)
$$

Linear Convergence of Bregman ADMM

We have two conclusions on the linear convergence of the Bregman ADMM. Note that the linear speed is achieved by adding more nice conditions.

- **Scenario #1**: $g(\boldsymbol{y})$ is μ_g-strongly convex and L_g-smooth, while $f(\boldsymbol{x})$ is only required to be convex (Theorem 3.8 of the ADMM book, p67)
- **Scenario #2**: $g(\boldsymbol{y})$ is μ_g-strongly convex and L_g-smooth, and $f(\boldsymbol{x})$ is μ_f-strongly convex (Theorem 3.9 of the ADMM book, p70)

The detailed theorems are omitted here. They have similar shapes to Theorem I.4 and Theorem I.5 above.

Complexity Comparisons

Complexity comparisons between ADMM and two variants of linearized ADMM:

Method	Rate	Linearization
ADMM	$\mathcal{O}(\sqrt{\frac{L_g}{\mu_g}}\frac{\|\mathbf{B}\|_2}{\sigma}\log\frac{1}{\epsilon})$	None
LADMM-1	$\mathcal{O}((\sqrt{\frac{L_g}{\mu_g}}\frac{\|\mathbf{B}\|_2}{\sigma}+\frac{\|\mathbf{B}\|_2^2}{\sigma^2})\log\frac{1}{\epsilon})$	On aug.
LADMM-2	$\mathcal{O}((\frac{\|\mathbf{B}\|_2^2}{\sigma^2}+\frac{L_g}{\mu_g})\log\frac{1}{\epsilon})$	On f, g and aug.

P69 of the ADMM book gives the proof on the complexity of LADMM-1. The other results can be obtained with a similar approach.

Outline

Acc-LADMM-1

ADMM can be combined with Nesterov's acceleration techniques.

When both f and g are generally convex and g is L_g-smooth, we can linear g at the auxiliary variable $\boldsymbol{v}^k$ in the $\boldsymbol{y}$ update step:

$$\boldsymbol{v}^k = \theta_k \boldsymbol{y}^k + (1-\theta_k)\tilde{\boldsymbol{y}}^k, \tag{69}$$

$$\begin{aligned} \boldsymbol{x}^{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \Big(& f(\boldsymbol{x}) + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b} \rangle \\ & + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 \Big), \end{aligned} \tag{70}$$

Acc-LADMM-1

When both f and g are generally convex and g is L_g-smooth, we can linear g at the auxiliary variable $\boldsymbol{v}^k$ in the $\boldsymbol{y}$ update step:

$$\begin{aligned}
\boldsymbol{y}^{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} \Big(& g(\boldsymbol{v}^k) + \langle \nabla g(\boldsymbol{v}^k), \boldsymbol{y} - \boldsymbol{v}^k \rangle) \\
& + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b} \rangle \\
& + \beta \langle \mathbf{B}^T (\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \boldsymbol{y} - \boldsymbol{y}^k \rangle \\
& + \frac{L_g \theta_k + \beta \|\mathbf{B}\|_2^2}{2} \|\boldsymbol{y} - \boldsymbol{y}^k\|^2 \Big),
\end{aligned} \tag{71}$$

$$\tilde{\boldsymbol{x}}^{k+1} = \theta_k \boldsymbol{x}^{k+1} + (1 - \theta_k) \tilde{\boldsymbol{x}}^k, \tag{72}$$

$$\tilde{\boldsymbol{y}}^{k+1} = \theta_k \boldsymbol{y}^{k+1} + (1 - \theta_k) \tilde{\boldsymbol{y}}^k, \tag{73}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \beta (\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}) \tag{74}$$

We call (69) $\sim$ (74) *Acc-LADMM-1*.

Convergence Rate of Acc-LADMM-1

Acc-LADMM-1 has a convergence rate of $\mathcal{O}(\frac{1}{K} + \frac{L_g}{K^2})$, which is faster than LADMM-2. The result is as follows.

Theorem III.1

Suppose that *f and g are generally convex and g is L_g-smooth*. Let $\theta_k \in (0, 1]$, $k \geq 0$, satisfy: $\forall k \geq 1 [\frac{1-\theta_k}{\theta_k^2} = \frac{1}{\theta_{k-1}^2}]$, $\theta_0 = 1$, and $\theta_{-1} = \infty$. Assume that $\forall k$, $\|\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^*\|^2 \leq D_{\boldsymbol{\lambda}}$ and $\|\boldsymbol{y}^k - \boldsymbol{y}^*\|^2 \leq D_{\boldsymbol{y}}$. Then for Acc-LADMM-1, we have

$$
\begin{aligned}
&|f(\tilde{\boldsymbol{x}}^{K+1}) + g(\tilde{\boldsymbol{y}}^{K+1}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*)| \\
&\qquad \leq \mathcal{O}\left(\frac{D_{\boldsymbol{y}} + D_{\boldsymbol{\lambda}} + \|\boldsymbol{\lambda}^*\|\sqrt{D_{\boldsymbol{\lambda}}}}{K} + \frac{L_g}{K^2}\right), \qquad (75)
\end{aligned}
$$

$$
\|\mathbf{A}\tilde{\boldsymbol{x}}^{K+1} + \mathbf{B}\tilde{\boldsymbol{y}}^{K+1} - \boldsymbol{b}\| \leq \mathcal{O}\left(\frac{\sqrt{D_{\boldsymbol{\lambda}}}}{K}\right). \qquad (76)
$$

Acc-LADMM-2

Acc-LADMM-1 only linearizes the sec. subproblem. Now we introduce Acc-LADMM-2, which linearizes both subproblems and it can also solve composite problems, i.e.,

$$f(\boldsymbol{x}) = f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) \text{ and } g(\boldsymbol{y}) = g_1(\boldsymbol{y}) + g_2(\boldsymbol{y}), \tag{77}$$

with non-smooth f_1 and g_1 and smooth f_2 and g_2.

The following iterations are called *Acc-LADMM-2*:

$$\boldsymbol{u}^k = \boldsymbol{x}^k + \frac{\theta_k(1-\theta_{k-1})}{\theta_{k-1}}(\boldsymbol{x}^k - \boldsymbol{x}^{k-1}), \tag{78}$$

$$\boldsymbol{v}^k = \boldsymbol{y}^k + \frac{\theta_k(1-\theta_{k-1})}{\theta_{k-1}}(\boldsymbol{y}^k - \boldsymbol{y}^{k-1}), \tag{79}$$

Acc-LADMM-2

The following ierations are called *Acc-LADMM-2*:

$$\begin{aligned}
\boldsymbol{x}^{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \Big(& f_1(\boldsymbol{x}) + \langle \nabla f_2(\boldsymbol{u}^k), \boldsymbol{x} \rangle + \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{u}^k\|^2 \\
& + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x} \rangle + \frac{\beta}{\theta_k}\langle \mathbf{A}^T(\mathbf{A}\boldsymbol{u}^k + \mathbf{B}\boldsymbol{v}^k - \boldsymbol{b}), \boldsymbol{x} \rangle \\
& + \frac{\beta\|\mathbf{A}\|_2^2}{2\theta_k}\|\boldsymbol{x} - \boldsymbol{u}^k\|^2 \Big).
\end{aligned} \tag{80}$$

$$\begin{aligned}
\boldsymbol{y}^{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} \Big(& g_1(\boldsymbol{y}) + \langle \nabla g_2(\boldsymbol{v}^k), \boldsymbol{y} \rangle + \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{v}^k\|^2 \\
& + \langle \boldsymbol{\lambda}^k, \mathbf{B}\boldsymbol{y} \rangle + \frac{\beta}{\theta_k}\langle \mathbf{B}^T(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{v}^k - \boldsymbol{b}), \boldsymbol{y} \rangle \\
& + \frac{\beta\|\mathbf{B}\|_2^2}{2\theta_k}\|\boldsymbol{y} - \boldsymbol{v}^k\|^2 \Big).
\end{aligned} \tag{81}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \beta\tau(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}). \tag{82}$$

Convergence Rate of Acc-LADMM-2

Acc-LADMM-2 has a better convergence rate than Acc-LADMM-1. The result is shown below in sketch:

Theorem III.2

With *non-smooth f_1 and g_1 and smooth f_2 and g_2*, we have

$$
\begin{aligned}
-\frac{2C_1\|\boldsymbol{\lambda}^*\|}{1+K(1-\tau)} &\le f(\boldsymbol{x}^{K+1}) + g(\boldsymbol{y}^{K+1}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) \\
&\le \frac{2C_1\|\boldsymbol{\lambda}^*\|}{1+K(1-\tau)} + \frac{C}{1+K(1-\tau)}, \qquad (83)
\end{aligned}
$$

$$
\|\mathbf{A}\boldsymbol{x}^{K+1} + \mathbf{B}\boldsymbol{y}^{K+1} - \boldsymbol{b}\| \le \frac{2C_1}{1+K(1-\tau)}, \qquad (84)
$$

where C and C_1 are constants.

The details are in Theorem 3.11 (p89, the ADMM book).

Acc-LADMM-3

Acc-LADMM-1 requires that f and g are generally convex and g is L_g-smooth. If we further assume that g is μ_g-strongly convex, then we further accelerate Acc-LADMM-1 with the following ierations:

$$
\begin{aligned}
\boldsymbol{w}^k &= \theta \boldsymbol{y}^k + (1-\theta)\tilde{\boldsymbol{y}}^k, && (85)\\
\boldsymbol{x}^{k+1} &= \operatorname*{argmin}_{\boldsymbol{x}} \Big(f(\boldsymbol{x}) + \langle \boldsymbol{\lambda}^k, \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b} \rangle \\
&\quad + \frac{\beta\theta}{2} \|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}\|^2 \Big). && (86)
\end{aligned}
$$

Acc-LADMM-3

We further accelerate Acc-LADMM-1 with the following ierations (cont'd):

$$
\begin{aligned}
\boldsymbol{y}^{k+1} &= \operatorname*{argmin}_{\boldsymbol{y}} \Big(\langle \nabla \mathrm{g}(\boldsymbol{w}^k), \boldsymbol{y} \rangle + \langle \boldsymbol{\lambda}^k, \mathbf{B}\boldsymbol{y} \rangle \\
&\quad + \beta\theta \langle \mathbf{B}^T (\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}), \boldsymbol{y} \rangle \\
&\quad + \frac{1}{2}\left(\frac{\theta}{\alpha} + \mu_{\mathrm{g}}\right) \left\| \boldsymbol{y} - \frac{1}{\frac{\theta}{\alpha} + \mu_{\mathrm{g}}} \left(\frac{\theta}{\alpha}\boldsymbol{y}^k + \mu_{\mathrm{g}} \boldsymbol{w}^k \right) \right\|^2 \Big). \\
&= \frac{1}{\frac{\theta}{\alpha} + \mu_{\mathrm{g}}} \Big\{ \mu_{\mathrm{g}} \boldsymbol{w}^k + \frac{\theta}{\alpha}\boldsymbol{y}^k - \big[\nabla \mathrm{g}(\boldsymbol{w}^k) + \mathbf{B}^T \boldsymbol{\lambda}^k \\
&\quad + \beta\theta \mathbf{B}^T (\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}) \big] \Big\}.
\end{aligned} \tag{87}
$$

Acc-LADMM-3

We further accelerate Acc-LADMM-1 with the following ierations (cont'd):

$$\tilde{\boldsymbol{x}}^{k+1} = \theta \boldsymbol{x}^{k+1} + (1-\theta)\tilde{\boldsymbol{x}}^{k}, \tag{88}$$

$$\tilde{\boldsymbol{y}}^{k+1} = \theta \boldsymbol{y}^{k+1} + (1-\theta)\tilde{\boldsymbol{y}}^{k}, \tag{89}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^{k} + \beta\theta(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}) \tag{90}$$

We call (85) $\sim$ (90) *Acc-LADMM-3.*

Acc-LADMM-3 has a faster convergence rate than LADMM-2, but with smaller complexity —— the same as ADMM. Its convergence rate is concluded in Theorem 3.12 (p95 of the ADMM book).

Outline

References

1. Lin, Zhouchen, Huan Li, and Cong Fang. *Alternating Direction Method of Multipliers for Machine Learning.* Springer Nature, 2022.
2. Boyd, Stephen, Stephen P. Boyd, and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.