# PART III: AUGMENTED LAGRANGIAN-BASED METHODS

# Lecture 8 – Augmented Lagrangian-Type Methods

- Main problem:

$$(P) \quad H_{\rm opt} = \min\{H(\mathbf{x},\mathbf{z}) \equiv h_1(\mathbf{x}) + h_2(\mathbf{z}) : \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}\},$$

  where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{m\times p}, \mathbf{c} \in \mathbb{R}^m$ and $h_1$ and $h_2$ are proper closed and convex functions.
- Lagrangian: $L(\mathbf{x},\mathbf{z};\mathbf{y}) = h_1(\mathbf{x}) + h_2(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\rangle$.
- Dual objective function:

$$\begin{aligned} q(\mathbf{y}) &= \min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{z}\in\mathbb{R}^p} \{h_1(\mathbf{x}) + h_2(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\rangle\} \\ &= -h_1^*(-\mathbf{A}^T\mathbf{y}) - h_2^*(-\mathbf{B}^T\mathbf{y}) - \langle \mathbf{c}, \mathbf{y}\rangle, \end{aligned}$$

- Dual problem:

$$q_{\rm opt} = \max_{\mathbf{y}\in\mathbb{R}^m}\{-h_1^*(-\mathbf{A}^T\mathbf{y}) - h_2^*(-\mathbf{B}^T\mathbf{y}) - \langle \mathbf{c}, \mathbf{y}\rangle\},$$

  or

$$\min_{\mathbf{y}\in\mathbb{R}^m} \left\{h_1^*(-\mathbf{A}^T\mathbf{y}) + h_2^*(-\mathbf{B}^T\mathbf{y}) + \langle \mathbf{c}, \mathbf{y}\rangle\right\}.$$

# Derivation Contd.

- The proximal point method employed on the dual problem:

$$\mathbf{y}^{k+1} = \underset{\mathbf{y}\in\mathbb{R}^m}{\operatorname{argmin}} \left\{ h_1^*(-\mathbf{A}^T\mathbf{y}) + h_2^*(-\mathbf{B}^T\mathbf{y}) + \langle \mathbf{c}, \mathbf{y}\rangle + \frac{1}{2\rho}\|\mathbf{y}-\mathbf{y}^k\|^2 \right\}$$

- By Fermat's optimality condition,

$$(*) \quad \mathbf{0} \in -\mathbf{A}\partial h_1^*(-\mathbf{A}^T\mathbf{y}^{k+1}) - \mathbf{B}\partial h_2^*(-\mathbf{B}^T\mathbf{y}^{k+1}) + \mathbf{c} + \frac{1}{\rho}(\mathbf{y}^{k+1} - \mathbf{y}^k).$$

- By the conjugate subgradient theorem, $\mathbf{y}^{k+1}$ satisfies (*) if and only if $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})$, where $\mathbf{x}^{k+1}$ and $\mathbf{z}^{k+1}$ satisfy

$$\begin{aligned} \mathbf{x}^{k+1} &\in \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}}\{\langle \mathbf{A}^T\mathbf{y}^{k+1}, \mathbf{x}\rangle + h_1(\mathbf{x})\}, \\ \mathbf{z}^{k+1} &\in \underset{\mathbf{z}\in\mathbb{R}^p}{\operatorname{argmin}}\{\langle \mathbf{B}^T\mathbf{y}^{k+1}, \mathbf{z}\rangle + h_2(\mathbf{z})\}. \end{aligned}$$

- Plugging the update equation for $\mathbf{y}^{k+1}$ in the above, we conclude that $\mathbf{y}^{k+1}$ satisfies (*) iff

$$\begin{aligned} \mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}), \\ \mathbf{x}^{k+1} &\in \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}}\{\langle \mathbf{A}^T(\mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})), \mathbf{x}\rangle + h_1(\mathbf{x})\}, \\ \mathbf{z}^{k+1} &\in \underset{\mathbf{z}\in\mathbb{R}^p}{\operatorname{argmin}}\{\langle \mathbf{B}^T(\mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})), \mathbf{z}\rangle + h_2(\mathbf{z})\}, \end{aligned}$$

# Derivation Contd.

- iff

$$\begin{aligned} \mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}), & (1) \\ \mathbf{0} &\in \mathbf{A}^T(\mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})) + \partial h_1(\mathbf{x}^{k+1}), & (2) \\ \mathbf{0} &\in \mathbf{B}^T(\mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})) + \partial h_2(\mathbf{z}^{k+1}). & (3) \end{aligned}$$

- Conditions (2) and (3) are satisfied if and only if $(\mathbf{x}^{k+1}, \mathbf{z}^{k+1})$ is a coordinate-wise minimum of the function

$$\tilde{H}(\mathbf{x}, \mathbf{z}) \equiv h_1(\mathbf{x}) + h_2(\mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2.$$

- It is known that coordinate-wise minima points of $\tilde{H}$ are exactly the minimizers of $\tilde{H}$, and therefore, the system (1), (2), (3) leads to primal representation of the dual proximal point method, known as the augmented Lagrangian Method described in the next slide.

# The Augmented Lagrangian Method

**The Augmented Lagrangian Method**

**Initialization:** $\mathbf{y}^0 \in \mathbb{R}^m$.
**General step:** for any $k = 0, 1, 2, \ldots$ execute the following steps:

$$
\begin{aligned}
(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}) &\in \underset{\mathbf{x}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ h_1(\mathbf{x}) + h_2(\mathbf{z}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\|^2 \right\} \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}).
\end{aligned}
$$

**Remarks:**

- the first step is the primal update step and the second is the dual update step.
- The augmented Lagrangian associated with the main problem is

$$L_\rho(\mathbf{x}, \mathbf{z}; \mathbf{y}) = h_1(\mathbf{x}) + h_2(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} \rangle + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\|^2.$$

- The primal update step is $(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}) \in \underset{\mathbf{x}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^p}{\operatorname{argmin}} L_\rho(\mathbf{x}, \mathbf{z}; \mathbf{y}^k)$. which is the reason for the name of the method.

# Alternating Direction Method of Multipliers (ADMM)

- The augmented Lagrangian method is in general not an implementable method since the primal update step is as hard as the original problem.
- One source of difficulty is the coupling term between the $\mathbf{x}$ and $\mathbf{z}$ .
- The alternating direction method of multipliers (ADMM) tackles this difficulty by replacing the exact minimization in the primal update step by one iteration of the alternating minimization method.

**ADMM**

**Initialization:** $\mathbf{x}^0 \in \mathbb{R}^n$, $\mathbf{z}^0 \in \mathbb{R}^p$, $\mathbf{y}^0 \in \mathbb{R}^m$.
**General step:** for any $k = 0, 1, \ldots$ execute the following:

(a) $\mathbf{x}^{k+1} \in \operatorname*{argmin}_{\mathbf{x}} \left\{ h_1(\mathbf{x}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\|^2 \right\}$;

(b) $\mathbf{z}^{k+1} \in \operatorname*{argmin}_{\mathbf{z}} \left\{ h_2(\mathbf{z}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\|^2 \right\}$;

(c) $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})$.

# Example: $\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x})$

- The same as

$$\min\{f(\mathbf{x}) + g(\mathbf{z}) : \mathbf{x} - \mathbf{z} = \mathbf{0}\}.$$

- ADMM is

$$\begin{aligned}
\mathbf{x}^{k+1} &= \text{argmin}\left\{f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{z}^k + \mathbf{u}^k\|_2^2\right\} = \operatorname{prox}_{\frac{1}{\rho}f}(\mathbf{z}^k - \mathbf{u}^k) \\
\mathbf{z}^{k+1} &= \text{argmin}\left\{g(\mathbf{z}) + \frac{\rho}{2}\| - \mathbf{z} + \mathbf{x}^{k+1} + \mathbf{u}_k\|_2^2\right\} = \operatorname{prox}_{\frac{1}{\rho}g}(\mathbf{x}^{k+1} + \mathbf{u}^k) \\
\mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}
\end{aligned}$$

- Specifically, if $f = \delta_C$ and $g = \delta_D$ for nonempty closed and convex sets $C, D$, the algorithm becomes

$$\begin{aligned}
\mathbf{x}^{k+1} &= P_C(\mathbf{z}^k - \mathbf{u}^k) \\
\mathbf{z}^{k+1} &= P_D(\mathbf{x}^{k+1} + \mathbf{u}^k) \\
\mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}.
\end{aligned}$$

# Comparison to Dykstra's Method (1986)

Let's rewrite ADMM (starting with the $\mathbf{z}$-update):

$$\begin{aligned} \mathbf{x}^{k+1} &= P_C(\mathbf{z}^k - \mathbf{u}^k) \\ \mathbf{z}^{k+1} &= P_D(\mathbf{x}^{k+1} + \mathbf{u}^k) \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}. \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \mathbf{z}^{k+1} &= P_D(\mathbf{x}^k + \mathbf{u}^k) \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{x}^k - \mathbf{z}^{k+1} \\ \mathbf{x}^{k+1} &= P_C(\mathbf{z}^{k+1} - \mathbf{u}^{k+1}) \end{aligned}$$

Dykstra's method (1986):

$$\begin{aligned} \mathbf{z}^{k+1} &= P_D(\mathbf{x}^k + \mathbf{u}^k) \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{x}^k - \mathbf{z}^{k+1} \\ \mathbf{x}^{k+1} &= P_C(\mathbf{z}^{k+1} + \mathbf{v}^k) \\ \mathbf{v}^{k+1} &= \mathbf{v}^k + \mathbf{z}^{k+1} - \mathbf{x}^{k+1} \end{aligned}$$

**Similar, but not the same...**

# Alternating Direction Proximal Method of Multipliers (AD-PMM)

The analysis actually considers a more general method than ADMM in which a quadratic proximity term is added to the objective in the minimization problems of steps (a) and (b). Assume that $\mathbf{G} \in \mathbb{S}^n_+, \mathbf{Q} \in \mathbb{S}^p_+$.

**AD-PMM**

**Initialization:** $\mathbf{x}^0 \in \mathbb{R}^n$, $\mathbf{z}^0 \in \mathbb{R}^p$, $\mathbf{y}^0 \in \mathbb{R}^m$.
**General step:** for any $k = 0, 1, \ldots$ execute the following:

(a) $\mathbf{x}^{k+1} \in \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ h_1(\mathbf{x}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\|^2 + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|^2_{\mathbf{G}} \right\}$;

(b) $\mathbf{z}^{k+1} \in$
$\underset{\mathbf{z} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ h_2(\mathbf{z}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\| + \frac{1}{2}\|\mathbf{z} - \mathbf{z}^k\|^2_{\mathbf{Q}} \right\}$;

(c) $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})$.

# Alternating Direction Linearized Prox Method of Multipliers (AD-LPMM)

- Choose $\mathbf{G} = \alpha\mathbf{I} - \rho\mathbf{A}^T\mathbf{A}$ with $\alpha \geq \rho\lambda_{\max}(\mathbf{A}^T\mathbf{A})$ and $\mathbf{Q} = \beta\mathbf{I} - \rho\mathbf{B}^T\mathbf{B}$ with $\beta \geq \rho\lambda_{\max}(\mathbf{B}^T\mathbf{B})$.
- $\mathbf{G}, \mathbf{Q} \in \mathbb{S}^n_+$.
- At the $\mathbf{x}$-step the following function is minimized:

$$
\begin{aligned}
& h_1(\mathbf{x}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2 + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|^2_{\mathbf{G}} \\
= \quad & h_1(\mathbf{x}) + \frac{\rho}{2}\left\|\mathbf{A}(\mathbf{x} - \mathbf{x}^k) + \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2 + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|^2_{\mathbf{G}} \\
= \quad & h_1(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{A}(\mathbf{x} - \mathbf{x}^k)\|^2 + \left\langle \rho\mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\rangle \\
& + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{x}^k\|^2 - \frac{\rho}{2}\|\mathbf{A}(\mathbf{x} - \mathbf{x}^k)\|^2 + \text{constant} \\
= \quad & h_1(\mathbf{x}) + \rho\left\langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\rangle + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{x}^k\|^2 + \text{constant},
\end{aligned}
$$

# AD-LPMM

- The $\mathbf{x}$-step can be written

$$\mathbf{x}^{k+1} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ h_1(\mathbf{x}) + \rho \left\langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\rangle + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{x}^k\|^2 \right\},$$

- or as

$$\mathbf{x}^{k+1} \in \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{\alpha} h_1(\mathbf{x}) + \frac{1}{2} \left\| \mathbf{x} - \left( \mathbf{x}^k - \frac{\rho}{\alpha}\mathbf{A}^T(\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k) \right) \right\|^2 \right\}.$$

- That is, $\mathbf{x}^{k+1} = \mathrm{prox}_{\frac{1}{\alpha}h_1}\left[ \mathbf{x}^k - \frac{\rho}{\alpha}\mathbf{A}^T \left( \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right) \right]$
- Similar expression for the $\mathbf{z}$-step.

# AD-LPMM

**AD-LPMM**

**Initialization:** $\mathbf{x}^0 \in \mathbb{R}^n$, $\mathbf{z}^0 \in \mathbb{R}^p$, $\mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$, $\alpha \geq \rho\lambda_{\max}(\mathbf{A}^T\mathbf{A}), \beta \geq \rho\lambda_{\max}(\mathbf{B}^T\mathbf{B})$.

**General step:** for any $k = 0, 1, \ldots$ execute the following:

(a) $\mathbf{x}^{k+1} = \mathrm{prox}_{\frac{1}{\alpha}h_1}\left[\mathbf{x}^k - \frac{\rho}{\alpha}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right)\right]$;

(b) $\mathbf{z}^{k+1} = \mathrm{prox}_{\frac{1}{\beta}h_2}\left[\mathbf{z}^k - \frac{\rho}{\beta}\mathbf{B}^T\left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right)\right]$;

(c) $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})$.

# Underlying Assumptions

(A) $h_1 : \mathbb{R}^n \to (-\infty, \infty]$ and $h_2 : \mathbb{R}^p \to (-\infty, \infty]$ are proper closed convex.

(B) $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{m \times p}$, $\mathbf{c} \in \mathbb{R}^m$, $\rho > 0$.

(C) $\mathbf{G} \in \mathbb{S}^n_+$, $\mathbf{Q} \in \mathbb{S}^p_+$.

(D) For any $\mathbf{a} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^p$ the problems

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ h_1(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}\|^2 + \frac{1}{2}\|\mathbf{x}\|^2_{\mathbf{G}} + \langle \mathbf{a}, \mathbf{x} \rangle \right\}$$

$$\min_{\mathbf{z} \in \mathbb{R}^p} \left\{ h_2(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{B}\mathbf{z}\|^2 + \frac{1}{2}\|\mathbf{z}\|^2_{\mathbf{Q}} + \langle \mathbf{b}, \mathbf{z} \rangle \right\}$$

possess an optimal solution.

(E) There exists $\hat{\mathbf{x}} \in \mathrm{ri}(\mathrm{dom}(h_1))$ and $\hat{\mathbf{z}} \in \mathrm{ri}(\mathrm{dom}(h_2))$ for which $\mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\hat{\mathbf{z}} = \mathbf{c}$.

(F) Problem (P) has a nonempty optimal set, denoted by $X^*$, and the corresponding optimal value is $H_{\mathrm{opt}}$.

**Result (strong duality):** $H_{\mathrm{opt}} = q_{\mathrm{opt}}$ and the dual problem possesses an optimal solution.

# $O(1/k)$ Rate of Convergence of AD-PMM

Theorem. Let $\{(\mathbf{x}^k, \mathbf{z}^k)\}_{k\geq 0}$ be the sequence generated by AD-PMM. Let $(\mathbf{x}^*, \mathbf{z}^*)$ be an optimal solution of problem (P) and $\mathbf{y}^*$ be an optimal solution of the dual problem. Suppose that $\gamma > 0$ is any constant satisfying $\gamma \geq 2\|\mathbf{y}^*\|$. Then for all $n \geq 0$,

$$\begin{aligned} H(\mathbf{x}^{(n)}, \mathbf{z}^{(n)}) - H_{\mathrm{opt}} &\leq \frac{\|\mathbf{x}^* - \mathbf{x}^0\|_{\mathbf{G}}^2 + \|\mathbf{z}^* - \mathbf{z}^0\|_{\mathbf{C}}^2 + \frac{1}{\rho}(\gamma + \|\mathbf{y}^0\|)^2}{2(n+1)}, \\ \|\mathbf{A}\mathbf{x}^{(n)} + \mathbf{B}\mathbf{z}^{(n)} - \mathbf{c}\| &\leq \frac{\|\mathbf{x}^* - \mathbf{x}^0\|_{\mathbf{G}}^2 + \|\mathbf{z}^* - \mathbf{z}^0\|_{\mathbf{C}}^2 + \frac{1}{\rho}(\gamma + \|\mathbf{y}^0\|)^2}{\gamma(n+1)}, \end{aligned}$$

where $\mathbf{C} = \rho\mathbf{B}^T\mathbf{B} + \mathbf{Q}$ and

$$\mathbf{x}^{(n)} = \frac{1}{n+1}\sum_{k=0}^{n} \mathbf{x}^{k+1}, \mathbf{z}^{(n)} = \frac{1}{n+1}\sum_{k=0}^{n} \mathbf{z}^{k+1}.$$

# Example: Convex Feasibility Problem

- **Problem:** find $\mathbf{x} \in C \cap D$ for two closed and convex sets.
- **Alternative Formulation:** $\min_x \delta_C(\mathbf{x}) + \delta_D(\mathbf{x})$
- ADMM reads as

$$
\begin{aligned}
\mathbf{x}^{k+1} &= P_C(\mathbf{z}^k - \mathbf{u}^k) \\
\mathbf{z}^{k+1} &= P_D(\mathbf{x}^{k+1} + \mathbf{u}^k) \\
\mathbf{u}^{k+1} &= \mathbf{u}^k + \mathbf{x}^{k+1} - \mathbf{z}^{k+1}.
\end{aligned}
$$

- Convergence rate result ($\mathbf{x}^k \in C, \mathbf{z}^k \in D$):

$$\|\mathbf{x}^k - \mathbf{z}^k\| \leq O(1/k)$$

- Other form: take the output sequence as $\mathbf{w}^k = \frac{\mathbf{x}^k + \mathbf{z}^k}{2}$. Then

$$\max\{d_C(\mathbf{w}^k), d_D(\mathbf{w}^k)\} \leq O(1/k)$$

- Better than alternating projections: $\mathbf{v}^{k+1} = P_C(P_D(\mathbf{v}^k))$ which is a subgradient method and has the rate

$$\max\{d_C(\mathbf{w}^k), d_D(\mathbf{w}^k)\} \leq O(1/\sqrt{k})$$

# Minimizing $f_1(\mathbf{x}) + f_2(\mathbf{A}\mathbf{x})$

- Consider the model

$$(Q) \quad \min_{\mathbf{x}\in\mathbb{R}^n} \{f_1(\mathbf{x}) + f_2(\mathbf{A}\mathbf{x})\},$$

  $f_1, f_2$ are proper closed convex functions and "proximable". $\mathbf{A} \in \mathbb{R}^{m\times n}$.
- Equivalent formulation:

$$(Q') \quad \min_{\mathbf{x}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^m} \{f_1(\mathbf{x}) + f_2(\mathbf{z}) : \mathbf{A}\mathbf{x} - \mathbf{z} = \mathbf{0}\}.$$

  This fits the general model (P) with $h_1 = f_1, h_2 = f_2$, $\mathbf{B} = -\mathbf{I}$ and $\mathbf{c} = \mathbf{0}$.

Direct implementation of ADMM:

**Algorithm 1 [ADMM – version 1]**

- **Initialization:** $\mathbf{x}^0 \in \mathbb{R}^n, \mathbf{z}^0, \mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$.
- **General Step ($k \geq 0$):**
  (a) $\mathbf{x}^{k+1} \in \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left[ f_1(\mathbf{x}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k \right\|^2 \right];$
  (b) $\mathbf{z}^{k+1} = \operatorname{prox}_{\frac{1}{\rho} f_2} \left( \mathbf{A}\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}^k \right);$
  (c) $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}).$

# Second Version of ADMM

- Step (a) of Algorithm 1 might be difficult to compute since the minimization involves a quadratic term of the form $\frac{\rho}{2}\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}$.
- We can rewrite problem (Q) as

$$\min_{\mathbf{x},\mathbf{w}\in\mathbb{R}^n,\mathbf{z}\in\mathbb{R}^m} \left\{ f_1(\mathbf{w}) + f_2(\mathbf{z}) : \mathbf{A}\mathbf{x} - \mathbf{z} = \mathbf{0}, \mathbf{x} - \mathbf{w} = \mathbf{0} \right\}.$$

Fits model (P) with $h_1 \equiv 0, h_2(\mathbf{z},\mathbf{w}) = f_1(\mathbf{z}) + f_2(\mathbf{w}), \mathbf{B} = -\mathbf{I}$ and $\mathbf{A} \leftarrow \begin{pmatrix} \mathbf{A} \\ \mathbf{I} \end{pmatrix}$

- The dual vector $\mathbf{y} \in \mathbb{R}^{m+n}$ is of the form $\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}$, where $\mathbf{y}_1 \in \mathbb{R}^m$ and $\mathbf{y}_2 \in \mathbb{R}^n$.
- We will consider two blocks: $\mathbf{x}$ and $(\mathbf{z}, \mathbf{w})$.
- The $\mathbf{x}$-step is given by

$$\begin{aligned} \mathbf{x}^{k+1} &= \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left[ \left\| \mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}_1^k \right\|^2 + \left\| \mathbf{x} - \mathbf{w}^k + \frac{1}{\rho}\mathbf{y}_2^k \right\|^2 \right] \\ &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1} \left( \mathbf{A}^T \left[ \mathbf{z}^k - \frac{1}{\rho}\mathbf{y}_1^k \right] + \mathbf{w}^k - \frac{1}{\rho}\mathbf{y}_2^k \right). \end{aligned}$$

# Second Version of ADMM for Solving (P)

**Algorithm 2 [ADMM – version 2]**

- **Initialization:** $\mathbf{x}^0, \mathbf{w}^0, \mathbf{y}_2^0 \in \mathbb{R}^n, \mathbf{z}^0, \mathbf{y}_1^0 \in \mathbb{R}^m, \rho > 0$.
- **General step ($k \geq 0$):**

$$
\begin{aligned}
\mathbf{x}^{k+1} &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k - \frac{1}{\rho}\mathbf{y}_1^k\right] + \mathbf{w}^k - \frac{1}{\rho}\mathbf{y}_2^k\right), \\
\mathbf{z}^{k+1} &= \mathrm{prox}_{\frac{1}{\rho}f_2}\left(\mathbf{A}\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_1^k\right), \\
\mathbf{w}^{k+1} &= \mathrm{prox}_{\frac{1}{\rho}f_1}\left(\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_2^k\right), \\
\mathbf{y}_1^{k+1} &= \mathbf{y}_1^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}), \\
\mathbf{y}_2^{k+1} &= \mathbf{y}_2^k + \rho(\mathbf{x}^{k+1} - \mathbf{w}^{k+1}).
\end{aligned}
$$

# AD-LPMM for Solving (Q)

Algorithm 2 might still be too computationally demanding since it involves the evaluation of the inverse of $\mathbf{I} + \mathbf{A}^T\mathbf{A}$ We can alternatively employ AD-LPMM on problem (Q')

**Algorithm 3 [AD-LPMM ]**

- **Initialization:** $\mathbf{x}^0 \in \mathbb{R}^n, \mathbf{z}^0, \mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$, $\alpha \geq \rho\lambda_{\max}(\mathbf{A}^T\mathbf{A}), \beta \geq \rho$.
- **General Step ($k \geq 0$):**

$$\begin{aligned}
\mathbf{x}^{k+1} &= \operatorname{prox}_{\frac{1}{\alpha}f_1}\left[\mathbf{x}^k - \frac{\rho}{\alpha}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right)\right],\\
\mathbf{z}^{k+1} &= \operatorname{prox}_{\frac{1}{\beta}f_2}\left[\mathbf{z}^k + \frac{\rho}{\beta}\left(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right)\right],\\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}).
\end{aligned}$$

# Example: $l_1$-regularized least squares

- Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1 \right\}, \tag{4}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$ and $\lambda > 0$.

- Problem (4) fits the composite model (Q) with $f_1(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ and $f_2(\mathbf{y}) \equiv \frac{1}{2}\|\mathbf{y}-\mathbf{b}\|_2^2$.
- $\mathrm{prox}_{\gamma f_1} = \mathcal{T}_{\gamma\lambda}$ and $\mathrm{prox}_{\gamma f_2}(\mathbf{y}) = \frac{\mathbf{y}+\gamma\mathbf{b}}{\gamma+1}$
- Step (a) of Algorithm 1 (first version of ADMM) has the form

$$\mathbf{x}^{k+1} \in \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left[ \lambda\|\mathbf{x}\|_1 + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right\|^2 \right],$$

which actually means that it is a useless algorithm.

# Algorithms 2 and 3 for Solving (4)

- **Algorithm 2 (second version of ADMM)**:

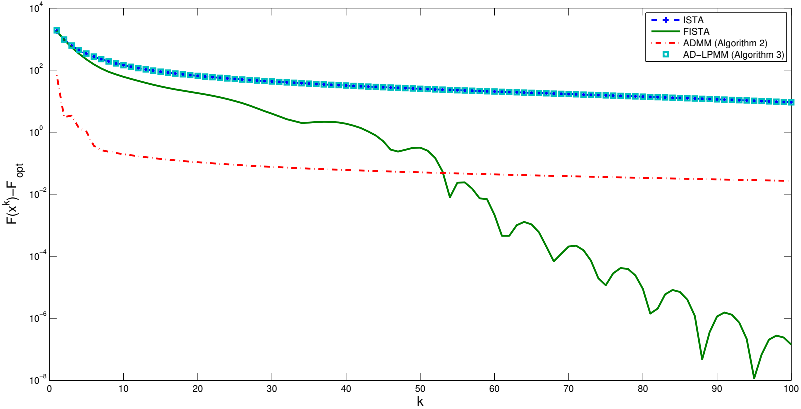
$$
\begin{aligned}
\mathbf{x}^{k+1} &= (\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k-\frac{1}{\rho}\mathbf{y}_1^k\right]+\mathbf{w}^k-\frac{1}{\rho}\mathbf{y}_2^k\right),\\
\mathbf{z}^{k+1} &= \frac{\rho\mathbf{A}\mathbf{x}^{k+1}+\mathbf{y}_1^k+\mathbf{b}}{\rho+1},\\
\mathbf{w}^{k+1} &= \mathcal{T}_{\frac{\lambda}{\rho}}\left(\mathbf{x}^{k+1}+\frac{1}{\rho}\mathbf{y}_2^k\right),\\
\mathbf{y}_1^{k+1} &= \mathbf{y}_1^k+\rho(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{z}^{k+1}),\\
\mathbf{y}_2^{k+1} &= \mathbf{y}_2^k+\rho(\mathbf{x}^{k+1}-\mathbf{w}^{k+1}).
\end{aligned}
$$

- **Algorithm 3 (AD-LPMM)** ($\alpha=\lambda_{\max}(\mathbf{A}^T\mathbf{A})\rho, \beta=\rho, L=\lambda_{\max}(\mathbf{A}^T\mathbf{A})$):

$$
\begin{aligned}
\mathbf{x}^{k+1} &= \mathcal{T}_{\frac{\lambda}{L\rho}}\left[\mathbf{x}^k-\frac{1}{L}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k-\mathbf{z}^k+\frac{1}{\rho}\mathbf{y}^k\right)\right],\\
\mathbf{z}^{k+1} &= \frac{\rho\mathbf{A}\mathbf{x}^{k+1}+\mathbf{y}^k+\mathbf{b}}{\rho+1},\\
\mathbf{y}^{k+1} &= \mathbf{y}^k+\rho(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{z}^{k+1}).
\end{aligned}
$$

# Numerical Example

$\mathbf{A} \in \mathbb{R}^{100\times 110}, \lambda = 1, \mathbf{x}_{\text{true}} = \mathbf{e}_3 - \mathbf{e}_7, \mathbf{b} = \mathbf{A}\mathbf{x}_{\text{true}}.$



# Example: Robust Regression

- Consider the problem
$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1, \tag{5}$$
- $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$.
- Problem (5) fits the composite model (Q) with $f_1 \equiv 0$ and $f_2(\mathbf{y}) = \|\mathbf{y} - \mathbf{b}\|_1$.
- For any $\gamma > 0$ $\mathrm{prox}_{\gamma f_1}(\mathbf{y}) = \mathbf{y}$ and $\mathrm{prox}_{\gamma f_2}(\mathbf{y}) = \mathcal{T}_\gamma(\mathbf{y} - \mathbf{b}) + \mathbf{b}$.
- The general step of **Algorithm 1 (first version of ADMM)** is

$$\begin{aligned}
\mathbf{x}^{k+1} &= \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^m} \left\|\mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right\|^2, \\
\mathbf{z}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left(\mathbf{A}\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}^k - \mathbf{b}\right) + \mathbf{b}, \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}).
\end{aligned}$$

- The general step of **Algorithm 2 (second version of ADMM)** is

$$
\begin{aligned}
\mathbf{x}^{k+1} &= (\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k-\frac{1}{\rho}\mathbf{y}_1^k\right]+\mathbf{w}^k-\frac{1}{\rho}\mathbf{y}_2^k\right),\\
\mathbf{z}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left(\mathbf{A}\mathbf{x}^{k+1}+\frac{1}{\rho}\mathbf{y}_1^k-\mathbf{b}\right)+\mathbf{b},\\
\mathbf{w}^{k+1} &= \mathbf{x}^{k+1}+\frac{1}{\rho}\mathbf{y}_2^k,\\
\mathbf{y}_1^{k+1} &= \mathbf{y}_1^k+\rho(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{z}^{k+1}),\\
\mathbf{y}_2^{k+1} &= \mathbf{y}_2^k+\rho(\mathbf{x}^{k+1}-\mathbf{w}^{k+1}).
\end{aligned}
$$

- Plugging the expression for $\mathbf{w}^{k+1}$ into the update formula of $\mathbf{y}_2^{k+1}$, we obtain that $\mathbf{y}_2^{k+1}=\mathbf{0}$. Thus, if we start with $\mathbf{y}_2^0=0$, then $\mathbf{y}_2^k=\mathbf{0}$ for all $k\geq 0$, and consequently $\mathbf{w}^k=\mathbf{x}^k$ for all $k$. The algorithm thus reduces to

$$
\begin{aligned}
\mathbf{x}^{k+1} &= (\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k-\frac{1}{\rho}\mathbf{y}_1^k\right]+\mathbf{x}^k\right),\\
\mathbf{z}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left(\mathbf{A}\mathbf{x}^{k+1}+\frac{1}{\rho}\mathbf{y}_1^k-\mathbf{b}\right)+\mathbf{b},\\
\mathbf{y}_1^{k+1} &= \mathbf{y}_1^k+\rho(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{z}^{k+1}).
\end{aligned}
$$

# Algorithm 3 [AD-LPMM]

Algorithm 3 (which is essentially AD-LPMM) with $\alpha = \lambda_{\max}(\mathbf{A}^T\mathbf{A})\rho, \beta = \rho$ takes the form (denoting $L = \lambda(\mathbf{A}^T\mathbf{A})$),

$$
\begin{aligned}
\mathbf{x}^{k+1} &= \mathbf{x}^k - \frac{1}{L}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right), \\
\mathbf{z}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left[\left(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b} + \frac{1}{\rho}\mathbf{y}^k\right)\right] + \mathbf{b}, \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}),
\end{aligned}
$$

# Exercise

**Exercise 1:** Consider the basis pursuit problem:

$$\min\{\|\mathbf{x}\|_1 : \mathbf{C}\mathbf{x} = \mathbf{b}\},$$

where $\mathbf{C} \in \mathbb{R}^{m\times n}$ is with linearly independent rows and $\mathbf{b} \in \mathbb{R}^m$.
Write the general update step of the ADLPMM method for solving the above problem by taking

(a) $f_1(\cdot) = \|\cdot\|_1, f_2(\cdot) = \delta_{\{\mathbf{x}:\mathbf{C}\mathbf{x}=\mathbf{b}\}}(\cdot), \mathbf{A} = \mathbf{I}$

(b) $f_1(\cdot) = \|\cdot\|_1, f_2(\cdot) = \delta_{\{\mathbf{b}\}}, \mathbf{A} = \mathbf{C}$

# Chambolle-Pock (CP) method

- aims at solving the model

$$(P) \quad \min_{\mathbf{x}} \{F(\mathbf{x}) \equiv f(\mathbf{A}\mathbf{x}) + g(\mathbf{x})\}.$$

- written in the min-max form:

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \underbrace{\langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle + g(\mathbf{x}) - f^*(\mathbf{y})}_{L(\mathbf{x},\mathbf{y})}.$$

- changing min/max order we get the dual problem

$$(D) \quad \max_{\mathbf{y}} \{q(\mathbf{y}) \equiv -g^*(-\mathbf{A}^T\mathbf{y}) - f^*(\mathbf{y})\}.$$

- Underlying assumptions:
  (A) $f, g$ closed proper and convex;
  (B) $\exists$ a saddle point $(\mathbf{x}^*, \mathbf{y}^*)$ (primal-dual solution):

$$L(\mathbf{x}, \mathbf{y}^*) \geq L(\mathbf{x}^*, \mathbf{y}^*) \geq L(\mathbf{x}^*, \mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y}.$$

# Chambolle-Pock (CP) method

- **Arrow, Hurwicz and Uzawa ('58):** A natural algorithm taking proximal gradient steps w.r.t. $\mathbf{x}$ and $\mathbf{y}$:

$$
\begin{aligned}
\mathbf{x}^{k+1} &= \mathrm{prox}_{\tau g}(\mathbf{x}^k - \tau \mathbf{A}^T \mathbf{y}^k) \\
\mathbf{y}^{k+1} &= \mathrm{prox}_{\sigma f^*}(\mathbf{y}^k + \sigma \mathbf{A}\mathbf{x}^{k+1})
\end{aligned}
$$

- CP (2011) performs an over-relaxation step

$$
\begin{aligned}
\mathbf{x}^{k+1} &= \mathrm{prox}_{\tau g}(\mathbf{x}^k - \tau \mathbf{A}^T \mathbf{y}^k) \\
\mathbf{y}^{k+1} &= \mathrm{prox}_{\sigma f^*}(\mathbf{y}^k + \sigma \mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k))
\end{aligned}
$$

- Denote $\mathbf{X}^N = \frac{1}{N}\sum_{k=1}^N \mathbf{x}^k, \mathbf{Y}^N = \frac{1}{N}\sum_{k=1}^N \mathbf{y}^k$. Then if $\tau\sigma \le \frac{1}{\|\mathbf{A}\|_2^2}$, then

$$
L(\mathbf{X}^N, \mathbf{y}) - L(\mathbf{x}, \mathbf{Y}^N) \le \frac{\frac{2}{\tau}\|\mathbf{x} - \mathbf{x}^0\|^2 + \frac{2}{\sigma}\|\mathbf{y} - \mathbf{y}^0\|^2}{N}
$$

- $\mathbf{x}^k$ and $\mathbf{y}^k$ are not primal and dual feasible respectively since it might be that $\mathbf{x}^k \notin \mathrm{dom}(f \circ \mathbf{A})$, $\mathbf{y}^k \notin \mathrm{dom}(g^* \circ (-\mathbf{A}^T))$.

# Convergence of the CP method

Assumption: $\tau\sigma \le \frac{1}{\|\mathbf{A}\|_2^2}$

$$L(\mathbf{X}^N, \mathbf{y}) - L(\mathbf{x}, \mathbf{Y}^N) \le \frac{\frac{2}{\tau}\|\mathbf{x} - \mathbf{x}^0\|^2 + \frac{2}{\sigma}\|\mathbf{y} - \mathbf{y}^0\|^2}{N}$$

Assume a bit more: $\tau\sigma < \frac{1}{\|\mathbf{A}\|_2^2}$

- $(\mathbf{x}^k, \mathbf{y}^k) \to (\mathbf{x}^*, \mathbf{y}^*)$ for some saddle point $(\mathbf{x}^*, \mathbf{y}^*)$.
- If $\mathrm{dom}(f) = \mathbb{R}^m$, then plugging in $\mathbf{x} = \mathbf{x}^*$, $\mathbf{y} = \bar{\mathbf{y}}^N \in \partial f(\mathbf{A}\mathbf{X}^N)$, we obtain that

$$F(\mathbf{X}^N) - F_{\mathrm{opt}} \le \frac{\frac{2}{\tau}\|\mathbf{x}^* - \mathbf{x}^0\|^2 + \frac{2}{\sigma}\|\bar{\mathbf{y}}^N - \mathbf{y}^0\|^2}{N}$$

  $\mathbf{X}^N$ bounded $\Rightarrow$ $\mathbf{y}^N$ bounded $\Rightarrow$ $F(\mathbf{X}^N) - F_{\mathrm{opt}} \le O(1/N)$

- Similarly, if $\mathrm{dom}(g^*) = \mathbb{R}^n$, then

$$F_{\mathrm{opt}} - q(\mathbf{Y}^N) \le O(1/N)$$

- If $\mathrm{dom}(f) = \mathbb{R}^m, \mathrm{dom}(g^*) = \mathbb{R}^n$, then we obtain **primal-dual gap convergence rate**

$$F_{\mathrm{opt}} - q(\mathbf{Y}^N) \le O(1/N)$$

# Equivalence of CP and ADLPMM

- Dual problem:
$$(D) \quad \min_{\mathbf{y}} g^*(-\mathbf{A}^T\mathbf{y}) + f^*(\mathbf{y}).$$
- First write the CP method on the dual:
$$\begin{aligned} \mathbf{y}^{k+1} &= \mathrm{prox}_{\sigma f^*}(\mathbf{y}^k + \sigma \mathbf{A}\mathbf{x}^k) \\ \mathbf{x}^{k+1} &= \mathrm{prox}_{\tau g}(\mathbf{x}^k - \tau \mathbf{A}^T(2\mathbf{y}^{k+1} - \mathbf{y}^k)) \end{aligned}$$
- Switching the order of **x** and **y** update:
$$\begin{aligned} \mathbf{x}^{k+1} &= \mathrm{prox}_{\tau g}(\mathbf{x}^k - \tau \mathbf{A}^T(2\mathbf{y}^{k} - \mathbf{y}^{k-1})) \\ \mathbf{y}^{k+1} &= \mathrm{prox}_{\sigma f^*}(\mathbf{y}^k + \sigma \mathbf{A}\mathbf{x}^{k+1}) \end{aligned}$$
- By Moreau Decomposition: $\mathbf{y}^{k+1} = \mathbf{y}^k + \sigma \mathbf{A}\mathbf{x}^{k+1} - \sigma \mathrm{prox}_{\frac{1}{\sigma}g}\left(\frac{1}{\sigma}\mathbf{y}^k + \mathbf{A}\mathbf{x}^{k+1}\right)$. The algorithm is thus,
$$\begin{aligned} \mathbf{x}^{k+1} &= \mathrm{prox}_{\tau f}(\mathbf{x}^k - \tau \mathbf{A}^T(2\mathbf{y}^{k} - \mathbf{y}^{k-1})) \\ \mathbf{z}^{k+1} &= \mathrm{prox}_{\frac{1}{\sigma}g}\left(\frac{1}{\sigma}\mathbf{y}^k + \mathbf{A}\mathbf{x}^{k+1}\right) \\ \mathbf{y}^{k+1} &= \mathbf{y}^k + \sigma(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}) \end{aligned}$$

# Equivalence of CP and ADLPMM

$$\begin{aligned}
\mathbf{x}^{k+1} &= \mathrm{prox}_{\tau f}(\mathbf{x}^k - \tau \mathbf{A}^T(2\mathbf{y}^k - \mathbf{y}^{k-1})) \\
\mathbf{z}^{k+1} &= \mathrm{prox}_{\frac{1}{\sigma} g}\left(\frac{1}{\sigma}\mathbf{y}^k + \mathbf{A}\mathbf{x}^{k+1}\right) \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \sigma(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1})
\end{aligned}$$

- The first step now reads as $\mathbf{x}^{k+1} = \mathrm{prox}_{\tau f}(\mathbf{x}^k - \tau \mathbf{A}^T(\mathbf{y}^k + \sigma(\mathbf{A}\mathbf{x}^k - \mathbf{z}^k)))$ Take $\tau = \frac{1}{\rho\|\mathbf{A}\|_2^2}, \sigma = \rho$. Then the method becomes exactly ADLPMM:

$$\begin{aligned}
\mathbf{x}^{k+1} &= \mathrm{prox}_{\frac{1}{\rho\|\mathbf{A}\|_2^2} f}\left[\mathbf{x}^k - \frac{1}{\|\mathbf{A}\|_2^2}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right)\right] \\
\mathbf{z}^{k+1} &= \mathrm{prox}_{\frac{1}{\rho} g}\left(\frac{1}{\rho}\mathbf{y}^k + \mathbf{A}\mathbf{x}^{k+1}\right) \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1})
\end{aligned}$$

# Douglas-Rachford is CP

Problem:

$$\min f(\mathbf{x}) + g(\mathbf{x})$$

- Douglas Rachford method is

$$\begin{aligned} \mathbf{x}^{k+1} &= \operatorname{prox}_f(\mathbf{y}^k) \\ \mathbf{y}^{k+1} &= \mathbf{y}^k + \operatorname{prox}_g(2\mathbf{x}^{k+1} - \mathbf{y}^k) - \mathbf{x}^{k+1}. \end{aligned}$$

- CP is ($\mathbf{A} = \mathbf{I}, \tau = \sigma = 1$):

$$\begin{aligned} \mathbf{x}^{k+1} &= \operatorname{prox}_f(\mathbf{x}^k - \mathbf{z}^k) \\ \mathbf{z}^{k+1} &= \operatorname{prox}_{g^*}(\mathbf{z}^k + 2\mathbf{x}^{k+1} - \mathbf{x}^k). \end{aligned}$$

- The same since DR can be rewritten as $(\operatorname{prox}_{g^*}(2\mathbf{x}^{k+1} - \mathbf{y}^k) = 2\mathbf{x}^{k+1} - \mathbf{y}^k - \operatorname{prox}_{g^*}(2\mathbf{x}^{k+1} - \mathbf{y}^k))$

$$\begin{aligned} \mathbf{x}^{k+1} &= \operatorname{prox}_f(\mathbf{y}^k) \\ \mathbf{y}^{k+1} &= \mathbf{x}^{k+1} - \operatorname{prox}_{g^*}(2\mathbf{x}^{k+1} - \mathbf{y}^k) \end{aligned} \quad \overset{\mathbf{z}^k = \mathbf{x}^k - \mathbf{y}^k}{\Longrightarrow} \quad \begin{aligned} \mathbf{x}^{k+1} &= \operatorname{prox}_f(\mathbf{x}^k - \mathbf{z}^k) \\ \mathbf{z}^{k+1} &= \operatorname{prox}_{g^*}(\mathbf{z}^k + 2\mathbf{x}^{k+1} - \mathbf{x}^k). \end{aligned}$$

# Extension: Addition of a Smooth Term

- aims at solving the model:

$$(P) \quad \min_{\mathbf{x}} \{F(\mathbf{x}) \equiv f(\mathbf{A}\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{x})\}.$$

$h : \mathbb{R}^n \to \mathbb{R}$ is $L_h$-smooth.

- CP method:

$$\begin{aligned} \mathbf{x}^{k+1} &= \operatorname{prox}_{\tau g}(\mathbf{x}^k - \tau(\mathbf{A}^T\mathbf{y}^k + \nabla h(\mathbf{x}^k))) \\ \mathbf{y}^{k+1} &= \operatorname{prox}_{\sigma f^*}(\mathbf{y}^k + \sigma\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k)) \end{aligned}$$

- Same convergence results under the condition

$$\left(\frac{1}{\tau} - L_h\right)\frac{1}{\sigma} \geq \|\mathbf{A}\|_2^2$$

# Accelerated CP

- Model:

$$(P) \quad \min_{\mathbf{x}} \{F(\mathbf{x}) \equiv f(\mathbf{A}\mathbf{x}) + g(\mathbf{x})\}.$$

**Additional assumption:** $g$ (or $f^*$) is $\gamma$-strongly convex ($\gamma > 0$).

- Accelerated CP: take $\tau_0 = \sigma_0 = \frac{1}{\|\mathbf{A}\|_2}$

$$\begin{aligned}
\mathbf{y}^{k+1} &= \operatorname{prox}_{\sigma_k f^*}(\mathbf{y}^k + \sigma_k \mathbf{A}(\mathbf{x}^k + \theta_k(\mathbf{x}^k - \mathbf{x}^{k-1}))) \\
\mathbf{x}^{k+1} &= \operatorname{prox}_{\tau_k g}(\mathbf{x}^k - \tau_k \mathbf{A}^T \mathbf{y}^{k+1}) \\
\theta_{k+1} &= \frac{1}{\sqrt{1+\gamma\tau_k}}, \tau_{k+1} = \theta_{k+1}\tau_k, \sigma_{k+1} = \sigma_k/\theta_{k+1}
\end{aligned}$$

- Rate of convergence result:

$$L(\mathbf{X}^N, \mathbf{y}) - L(\mathbf{x}, \mathbf{Y}^N) \le O(1)\frac{\frac{\|\mathbf{x}-\mathbf{x}^0\|^2}{\tau} + \frac{\|\mathbf{y}-\mathbf{y}^0\|^2}{\sigma}}{N^2}$$

Also $\|\mathbf{x}^k - \mathbf{x}^*\|^2 \le O(1/k^2)$

# Exercise

**Exercise 2:** Consider the problem

$$\min\left\{-\sum_{i=1}^{m}\log(\mathbf{a}_i^T\mathbf{x}-b_i)+\sum_{i=1}^{n-2}\sqrt{(x_i-x_{i+1})^2+(x_{i+1}-x_{i+2})^2}:\mathbf{a}_i^T\mathbf{x}>b_i, i\in[m]\right\}$$

where $\mathbf{a}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for all $i \in [m]$.

(a) Write an ADLPMM method for solving the problem.

(b) Write the CP method for solving the problem.

(c) Assume that $\mathbf{A}$ and $\mathbf{b}$ are generated via

**MATLAB**

```
m=30;
n=25;
a = [0:m-1]+1;
b = [0:n-1]+0.5;
A = sin(10*(a'*b).^3);
xi = sin(31*[1:n].^3)';
b = A*xi+sin(23*[1:m].^3)'+1.5
```

**Python**

```
import numpy as np
import numpy.linalg as la
m = 30
n = 25
a = np.arange(m)+1
b = np.arange(n)+0.5
A = np.sin(10*np.outer(a,b)**3)
xi = np.sin(31*np.arange(1,n+1)**3)
b = A@xi+np.sin(23*np.arange(1,m+1)**3)+1.5
```

# Exercise 2 Contd.

Implement the CP and ADLPMM in MATLAB/Python. Run the ADLPPMM method with $\rho = 1$ and $\rho = \frac{1}{\|\mathbf{A}\|_2}$. Run the CP method with $\tau = \sigma = \frac{1}{\|\mathbf{A}\|_2}$ and also $\tau = \frac{1}{\|\mathbf{A}\|_2^2}, \sigma = 1$. Run 60 iterations of each of the methods ($4 = 2 \times 2$ runs). Initialize all the vectors with zeros. Print the first three components $(x_1, x_2, x_3)$ generated by each of the methods. Add a plot of $f(\mathbf{x}^k)$ of the four methods (in the same plot). Plot only the iterations in which the vector is feasible!

(d) Consider the problem

$$\min \left\{ \frac{\|\mathbf{x}\|_2^2}{2} - \sum_{i=1}^{m} \log(\mathbf{a}_i^T \mathbf{x} - b_i) + \sum_{i=1}^{n-2} \sqrt{(x_i - x_{i+1})^2 + (x_{i+1} - x_{i+2})^2} : \mathbf{a}_i^T \mathbf{x} > b_i \right\}$$

Write an accelerated CP (ACP) method as well as the FDPG method for solving the problem.

(e) Implement the two methods from part (d) in MATLAB/Python on the data generated in part (c). For the ACP method use $\tau_0 = \sigma_0 = \frac{1}{\|\mathbf{A}\|_2}$. Run 100 iterations of the two methods. Write the the first three components produced by each of the methods and plot the function values generated by each of the two methods (in the same plot). Plot only the iterations in which the vector is feasible.

# References


- A. Beck, *First Order Methods in Optimization*, SIAM-MOS (2017)
- S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Found. Trends Mach. Learn. (2011)
- A. Chambolle and T. Pock, *A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging*, J. Math. Imaging Vision (2011)
- X. Gao and S.-Z. Zhang, *First-order algorithms for convex optimization with nonseparable objective and coupled constraints*, J. Oper. Res. Soc. China (2017)
- B. He and X. Yuan, *On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal. (2012)