An Introduction to Non-smooth Optimization

Lecture 05 - Alternating Direction Method of Multipliers

Jingwei LIANG

Institute of Natural Sciences, Shanghai Jiao Tong University

Email: optimization.sjtu@gmail.com

Office: Room 355, No. 6 Science Building

Previously

Principal component pursuit Let $\mu > 0$,

$$\min_{\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^{m \times n}} \|\boldsymbol{x}\|_* + \mu \|\boldsymbol{y}\|_1,$$

$$\text{such that } \boldsymbol{x} + \boldsymbol{y} = \boldsymbol{f}.$$

Previously

Principal component pursuit Let $\mu > 0$,

$$\min_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{m\times n}} \quad \|\boldsymbol{x}\|_* + \mu\|\boldsymbol{y}\|_1,$$

$$\text{such that} \quad \boldsymbol{x} + \boldsymbol{y} = \boldsymbol{f}.$$

Problem - Non-smooth optimization problem

Consider

$$\min_{\boldsymbol{x}\in\mathbb{R}^m,\boldsymbol{y}\in\mathbb{R}^n} \quad F(\boldsymbol{x}) + R(\boldsymbol{y}),$$

$$\text{such that} \quad \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{f}.$$

with

- $F \in \Gamma_0(\mathbb{R}^m)$, $R \in \Gamma_0(\mathbb{R}^n)$
- $\boldsymbol{f} \in \mathbb{R}^p$, $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^p$ and $\boldsymbol{B} : \mathbb{R}^n \to \mathbb{R}^p$ are bounded linear.

Definition - Primal and dual problem

Let $F \in \Gamma_0(\mathbb{R}^n)$ and $R \in \Gamma_0(\mathbb{R}^m)$, let $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear. Then

- The *primal problem*

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} F(\boldsymbol{x}) + R(\boldsymbol{Kx}).$$

- The *dual problem*

$$\max_{\boldsymbol{u} \in \mathbb{R}^m} -F^*(-\boldsymbol{K}^*\boldsymbol{u}) - R^*(\boldsymbol{u}).$$

Definition - Primal and dual problem

Let $F \in \Gamma_0(\mathbb{R}^n)$ and $R \in \Gamma_0(\mathbb{R}^m)$, let $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear. Then

- The *primal problem*

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}).$$

- The *dual problem*

$$\max_{\boldsymbol{u}\in\mathbb{R}^m} -F^*(-\boldsymbol{K}^*\boldsymbol{u}) - R^*(\boldsymbol{u}).$$

Let $S = \{\boldsymbol{f}\} \subset \mathbb{R}^m$ and $R(\boldsymbol{y}) = \iota_S(\boldsymbol{y})$, then

- The *primal problem*

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \quad \Longleftrightarrow \quad \min_{\boldsymbol{x}\in\mathbb{R}^n} F(\boldsymbol{x}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}.$$

- The *dual problem*

$$\max_{\boldsymbol{u}\in\mathbb{R}^m} -F^*(-\boldsymbol{K}^*\boldsymbol{u}) - \langle \boldsymbol{f} \mid \boldsymbol{u} \rangle.$$

Duality

Consider the *primal problem*: let $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^p$ and $\boldsymbol{B} : \mathbb{R}^m \to \mathbb{R}^p$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m} F(\boldsymbol{x}) + R(\boldsymbol{y}) \quad \text{such that} \quad \boldsymbol{Ax} + \boldsymbol{By} = \boldsymbol{f}. \tag{$\mathscr{P}$}$$

Consider the *primal problem*: let $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^p$ and $\boldsymbol{B} : \mathbb{R}^m \to \mathbb{R}^p$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m} F(\boldsymbol{x}) + R(\boldsymbol{y}) \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{f}. \tag{$\mathscr{P}$}$$

Define

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}, \quad H(\boldsymbol{z}) = F(\boldsymbol{x}) + R(\boldsymbol{y}) \quad \text{and} \quad \boldsymbol{K} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}.$$

Then ($\mathscr{P}$) is equivalent to

$$\min_{\boldsymbol{z}\in\mathbb{R}^{m+n}} H(\boldsymbol{z}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{z} = \boldsymbol{f}$$

Duality

Consider the *primal problem*: let $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^p$ and $\boldsymbol{B} : \mathbb{R}^m \to \mathbb{R}^p$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m} F(\boldsymbol{x}) + R(\boldsymbol{y}) \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{f}. \tag{$\mathscr{P}$}$$

Define

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}, \quad H(\boldsymbol{z}) = F(\boldsymbol{x}) + R(\boldsymbol{y}) \quad \text{and} \quad \boldsymbol{K} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}.$$

Then ($\mathscr{P}$) is equivalent to

$$\min_{\boldsymbol{z}\in\mathbb{R}^{m+n}} H(\boldsymbol{z}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{z} = \boldsymbol{f}$$

Dual problem

$$\max_{\boldsymbol{u}\in\mathbb{R}^p} -H^*(-\boldsymbol{K}^*\boldsymbol{u}) - \langle \boldsymbol{f} \mid \boldsymbol{u} \rangle.$$

Consider the *primal problem*: let $\boldsymbol{A}:\mathbb{R}^n\to\mathbb{R}^p$ and $\boldsymbol{B}:\mathbb{R}^m\to\mathbb{R}^p$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^m} F(\boldsymbol{x})+R(\boldsymbol{y}) \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{y}=\boldsymbol{f}. \tag{$\mathscr{P}$}$$

Define

$$\boldsymbol{z}=\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}, \quad H(\boldsymbol{z})=F(\boldsymbol{x})+R(\boldsymbol{y}) \quad \text{and} \quad \boldsymbol{K}=\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\end{bmatrix}.$$

Then ($\mathscr{P}$) is equivalent to

$$\min_{\boldsymbol{z}\in\mathbb{R}^{m+n}} H(\boldsymbol{z}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{z}=\boldsymbol{f}$$

Dual problem

$$\max_{\boldsymbol{u}\in\mathbb{R}^p} -H^*(-\boldsymbol{K}^*\boldsymbol{u})-\langle \boldsymbol{f}\mid\boldsymbol{u}\rangle.$$

The *dual problem*:

$$\max_{\boldsymbol{u}\in\mathbb{R}^p} -F^*(-\boldsymbol{A}^*\boldsymbol{u})-R^*(-\boldsymbol{B}^*\boldsymbol{u})-\langle \boldsymbol{f}\mid\boldsymbol{u}\rangle. \tag{$\mathscr{D}$}$$

Outline

1. Dual ascent
2. Alternating Direction Method of Multipliers
3. Douglas–Rachford splitting method
4. Two variants of ADMM

Problem

Primal problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\phi(\boldsymbol{x}) = \{F(\boldsymbol{x}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}\}\}.$$

Problem

Primal problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{\phi(\boldsymbol{x}) = \{F(\boldsymbol{x}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}\}\right\}.$$

Definition - Lagrangian multiplier

Let $\boldsymbol{u} \in \mathbb{R}^m$,

$$\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\rangle.$$

Problem

Primal problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{\phi(\boldsymbol{x}) = \{F(\boldsymbol{x}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}\}\right\}.$$

Definition - Lagrangian multiplier

Let $\boldsymbol{u} \in \mathbb{R}^m$,

$$\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\rangle.$$

Dual problem

$$\max_{\boldsymbol{u}\in\mathbb{R}^m} \left\{\psi(\boldsymbol{u}) = -F^*(-\boldsymbol{K}^*\boldsymbol{u}) - \langle \boldsymbol{f} \mid \boldsymbol{u}\rangle\right\}.$$

Dual ascent

Consider solving

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\max_{\boldsymbol{u}\in\mathbb{R}^m}\ \left\{\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\rangle\right\}.$$

Consider solving

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \max_{\boldsymbol{u}\in\mathbb{R}^m} \left\{ \mathscr{L}(\boldsymbol{x};\boldsymbol{u}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + \langle \boldsymbol{u} \,|\, \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f} \rangle \right\}.$$

Algorithm - Dual ascent

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &\in \operatorname{Argmin}_{\boldsymbol{x}} \mathscr{L}(\boldsymbol{x};\boldsymbol{u}^{(k)}) \\ &\in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \langle \boldsymbol{u}^{(k)} \,|\, \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f} \rangle \\ \boldsymbol{u}^{(k+1)} &= \boldsymbol{u}^{(k)} + \gamma_k(\boldsymbol{K}\boldsymbol{x}^{(k+1)} - \boldsymbol{f}), \quad \gamma_k > 0 \end{aligned}$$

Dual ascent

Consider solving

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \max_{\boldsymbol{u}\in\mathbb{R}^m} \left\{\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\rangle\right\}.$$

Algorithm - Dual ascent

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &\in \operatorname{Argmin}_{\boldsymbol{x}} \mathscr{L}(\boldsymbol{x};\boldsymbol{u}^{(k)}) \\
&\in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \langle \boldsymbol{u}^{(k)} \mid \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\rangle \\
\boldsymbol{u}^{(k+1)} &= \boldsymbol{u}^{(k)} + \gamma_k(\boldsymbol{K}\boldsymbol{x}^{(k+1)} - \boldsymbol{f}), \quad \gamma_k > 0
\end{aligned}$$

- Gradient ascent for dual problem

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \gamma_k \psi(\boldsymbol{u}^{(k)}).$$

- $\nabla\psi(\boldsymbol{u}) = \boldsymbol{K}\bar{\boldsymbol{x}} - \boldsymbol{f}$ when $\bar{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x}}\mathscr{L}(\boldsymbol{x};\boldsymbol{u})$.
- Works, but needs many strong conditions.

Dual decomposition

Suppose F is separable

$$F(\boldsymbol{x}) = F_1(\boldsymbol{x}_1) + \cdots + F_\ell(\boldsymbol{x}_\ell), \quad \boldsymbol{x} = (\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_\ell^T)^T.$$

Dual decomposition

Suppose F is separable

$$F(\boldsymbol{x}) = F_1(\boldsymbol{x}_1) + \cdots + F_\ell(\boldsymbol{x}_\ell), \quad \boldsymbol{x} = (\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_\ell^T)^T.$$

- $\mathscr{L}$ is then separable in $\boldsymbol{x}$: $\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) = \mathscr{L}_1(\boldsymbol{x}_1;\boldsymbol{u}) + \cdots + \mathscr{L}_\ell(\boldsymbol{x}_\ell;\boldsymbol{u})$ with

$$\mathscr{L}_i(\boldsymbol{x}_i;\boldsymbol{u}) = F_i(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}_i \boldsymbol{x}_i \rangle, \quad i = 1, ..., \ell.$$

Dual decomposition

Suppose F is separable

$$F(\boldsymbol{x}) = F_1(\boldsymbol{x}_1) + \cdots + F_\ell(\boldsymbol{x}_\ell), \quad \boldsymbol{x} = (\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_\ell^T)^T.$$

- $\mathscr{L}$ is then separable in $\boldsymbol{x}$: $\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) = \mathscr{L}_1(\boldsymbol{x}_1;\boldsymbol{u}) + \cdots + \mathscr{L}_\ell(\boldsymbol{x}_\ell;\boldsymbol{u})$ with

$$\mathscr{L}_i(\boldsymbol{x}_i;\boldsymbol{u}) = F_i(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}_i\boldsymbol{x}_i\rangle, \quad i = 1, \ldots, \ell.$$

- x-minimization in dual ascent splits into ℓ separate minimizations

$$\boldsymbol{x}_i^{(k+1)} \in \mathrm{Argmin}_{\boldsymbol{x}_i} \mathscr{L}_i(\boldsymbol{x}_i;\boldsymbol{u}^{(k)}).$$

  which can be done in parallel fashion

Dual decomposition

Suppose F is separable

$$F(\boldsymbol{x}) = F_1(\boldsymbol{x}_1) + \cdots + F_\ell(\boldsymbol{x}_\ell), \quad \boldsymbol{x} = (\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_\ell^T)^T.$$

- $\mathscr{L}$ is then separable in $\boldsymbol{x}$: $\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) = \mathscr{L}_1(\boldsymbol{x}_1;\boldsymbol{u}) + \cdots + \mathscr{L}_\ell(\boldsymbol{x}_\ell;\boldsymbol{u})$ with

$$\mathscr{L}_i(\boldsymbol{x}_i;\boldsymbol{u}) = F_i(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}_i\boldsymbol{x}_i \rangle, \quad i = 1, ..., \ell.$$

- x-minimization in dual ascent splits into ℓ separate minimizations

$$\boldsymbol{x}_i^{(k+1)} \in \mathrm{Argmin}_{\boldsymbol{x}_i} \mathscr{L}_i(\boldsymbol{x}_i;\boldsymbol{u}^{(k)}).$$

  which can be done in parallel fashion

Dual decomposition

$$\boldsymbol{x}_i^{(k+1)} \in \mathrm{Argmin}_{\boldsymbol{x}_i} L_i(\boldsymbol{x}_i, \boldsymbol{u}^{(k)}), \ i = 1, ..., \ell,$$
$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \gamma_k \left(\textstyle\sum_{i=1}^{\ell} \boldsymbol{K}_i \boldsymbol{x}_i^{(k+1)} - \boldsymbol{f}\right).$$

- Scatter $\boldsymbol{u}^{(k)}$, update $\boldsymbol{x}_i$ in parallel, and gather $\boldsymbol{K}_i\boldsymbol{x}_i^{(k+1)}$.
- Waiting for the slowest $\boldsymbol{x}_i$ update.

Augmented Lagrangian

Primal problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\phi(\boldsymbol{x}) = F(\boldsymbol{x}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}\}.$$

Augmented Lagrangian

Primal problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{\phi(\boldsymbol{x}) = F(\boldsymbol{x}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}\right\}.$$

Definition - Augmented Lagrangian

Let $\rho > 0$ and $\boldsymbol{u} \in \mathbb{R}^m$,

$$\mathscr{L}_\rho(\boldsymbol{x};\boldsymbol{u}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\rangle + \frac{\rho}{2}\|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\|^2.$$

Augmented Lagrangian

Primal problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{\phi(\boldsymbol{x}) = F(\boldsymbol{x}) \quad \text{such that} \quad \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}\right\}.$$

Definition - Augmented Lagrangian

Let $\rho > 0$ and $\boldsymbol{u} \in \mathbb{R}^m$,

$$\mathscr{L}_\rho(\boldsymbol{x};\boldsymbol{u}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + \langle \boldsymbol{u} \mid \boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\rangle + \frac{\rho}{2}\|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\|^2.$$

Alternative formulation Let $\rho > 0$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\|^2$$
$$\text{such that } \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}.$$

Method of multipliers

Algorithm - Method of multipliers

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &\in \operatorname{Argmin}_{\boldsymbol{x}} \mathscr{L}_\rho(\boldsymbol{x}; \boldsymbol{u}^{(k)}) \\ &\in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho\|^2 \\ \boldsymbol{u}^{(k+1)} &= \boldsymbol{u}^{(k)} + \rho(\boldsymbol{K}\boldsymbol{x}^{(k+1)} - \boldsymbol{f}) \end{aligned}$$

- Specific step-size for dual update.
- Weaker conditions for convergence: non-smooth F.
- How $\|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{f}\|^2$ destroy the separable structure of $\boldsymbol{x}$.

Alternating Direction Method of Multipliers

ADMM, dual formulation

Primal problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m} \quad & F(\boldsymbol{x}) + R(\boldsymbol{y}), \\ \text{such that} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{f}. \end{aligned} \tag{$\mathscr{P}$}$$

Primal problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^m} \quad & F(\boldsymbol{x}) + R(\boldsymbol{y}), \\ \text{such that} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{f}. \end{aligned} \tag{$\mathscr{P}$}$$

Lagrangian multiplier

$$\mathscr{L}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{u}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + R(\boldsymbol{y}) + \langle \boldsymbol{u} \,|\, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} \rangle.$$

$$\begin{aligned} \max_{\boldsymbol{u}} \min_{\boldsymbol{x},\boldsymbol{y}} \mathscr{L}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{u}) &= \max_{\boldsymbol{u}} \min_{\boldsymbol{x},\boldsymbol{y}} F(\boldsymbol{x}) + R(\boldsymbol{y}) + \langle \boldsymbol{u} \,|\, \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} \rangle \\ &= \max_{\boldsymbol{u}} \min_{\boldsymbol{x},\boldsymbol{y}} F(\boldsymbol{x}) + \langle \boldsymbol{A}^*\boldsymbol{u} \,|\, \boldsymbol{x} \rangle + R(\boldsymbol{y}) + \langle \boldsymbol{B}^*\boldsymbol{u} \,|\, \boldsymbol{y} \rangle - \langle \boldsymbol{u} \,|\, \boldsymbol{f} \rangle \\ &= \max_{\boldsymbol{u}} \min_{\boldsymbol{x}} F(\boldsymbol{x}) - \langle -\boldsymbol{A}^*\boldsymbol{u} \,|\, \boldsymbol{x} \rangle + \min_{\boldsymbol{y}} R(\boldsymbol{y}) - \langle -\boldsymbol{B}^*\boldsymbol{u} \,|\, \boldsymbol{y} \rangle - \langle \boldsymbol{u} \,|\, \boldsymbol{f} \rangle \\ &= \max_{\boldsymbol{u}} - \max_{\boldsymbol{x}} \langle -\boldsymbol{A}^*\boldsymbol{u} \,|\, \boldsymbol{x} \rangle - F(\boldsymbol{x}) - \max_{\boldsymbol{y}} \langle -\boldsymbol{B}^*\boldsymbol{u} \,|\, \boldsymbol{y} \rangle - R(\boldsymbol{y}) - \langle \boldsymbol{u} \,|\, \boldsymbol{f} \rangle \\ &= \max_{\boldsymbol{u}} -F^*(-\boldsymbol{A}^*\boldsymbol{u}) - R^*(-\boldsymbol{B}^*\boldsymbol{u}) - \langle \boldsymbol{u} \,|\, \boldsymbol{f} \rangle \end{aligned}$$

Primal problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m} \ F(\boldsymbol{x}) + R(\boldsymbol{y}), \quad \text{such that } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{f}. \tag{$\mathscr{P}$}$$

Augmented Lagrangian Let $\rho > 0$

$$\begin{aligned}\mathscr{L}_\rho(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}) &\stackrel{\text{def}}{=} F(\boldsymbol{x}) + R(\boldsymbol{y}) + \langle \boldsymbol{u} \mid \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} \rangle + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f}\|^2 \\ &= F(\boldsymbol{x}) + R(\boldsymbol{y}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} + \boldsymbol{u}/\rho\|^2 - \frac{\|\boldsymbol{u}\|^2}{2\rho}.\end{aligned}$$

Alternating direction method of multipliers

Primal problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m} \quad & F(\boldsymbol{x}) + R(\boldsymbol{y}), \\ \text{such that} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{f}. \end{aligned} \tag{$\mathscr{P}$}$$

Algorithm - Method of multiplier

$$\begin{aligned} (\boldsymbol{x}^{(k+1)}, \boldsymbol{y}^{(k+1)}) &\in \mathrm{Argmin}_{\boldsymbol{x}} \mathscr{L}_\rho(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}^{(k)}) \\ &= \mathrm{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + R(\boldsymbol{y}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho\|^2 - \frac{\|\boldsymbol{u}^{(k)}\|^2}{2\rho}, \\ \boldsymbol{u}^{(k+1)} &= \boldsymbol{u}^{(k)} + \rho(\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \boldsymbol{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{f}). \end{aligned}$$

Primal problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^m} \; F(\boldsymbol{x}) + R(\boldsymbol{y}), \quad \text{such that } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} = \boldsymbol{f}. \tag{$\mathscr{P}$}$$

Algorithm - ADMM [Gabay, Mercier, Glowinski, Marrocco '76]

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &\in \operatorname{Argmin}_{\boldsymbol{x}} \mathscr{L}_\rho(\boldsymbol{x}, \boldsymbol{y}^{(k)}; \boldsymbol{u}^{(k)}) \\
&= \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + R(\boldsymbol{y}^{(k)}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{(k)} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho\|^2 - \frac{\|\boldsymbol{u}^{(k)}\|^2}{2\rho}, \\
\boldsymbol{y}^{(k+1)} &\in \operatorname{Argmin}_{\boldsymbol{y}} \mathscr{L}_\rho(\boldsymbol{x}^{(k+1)}, \boldsymbol{y}; \boldsymbol{u}^{(k)}) \\
&= \operatorname{Argmin}_{\boldsymbol{y}} F(\boldsymbol{x}^{(k+1)}) + R(\boldsymbol{y}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho\|^2 - \frac{\|\boldsymbol{u}^{(k)}\|^2}{2\rho}, \\
\boldsymbol{u}^{(k+1)} &= \boldsymbol{u}^{(k)} + \rho(\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \boldsymbol{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{f}).
\end{aligned}$$

- In general **NO** closed form for $\boldsymbol{x}^{(k)}, \boldsymbol{y}^{(k)}$.
- $\boldsymbol{x}^{(k+1)}$ is unique if $\boldsymbol{A}$ has full column rank, same for $\boldsymbol{B}$ in $\boldsymbol{y}^{(k)}$ update.

Alternating direction method of multipliers

Basic convergence result

- Assumption
  - R, F are proper convex and closed.
  - $L_{\rho=0}$ has saddle-point.
- Convergence
  - Objective function value $F(\boldsymbol{x}^{(k)}) + R(\boldsymbol{y}^{(k)}) \to \mu^\star$.
  - Feasibility $\boldsymbol{Ax}^{(k)} + \boldsymbol{By}^{(k)} - \boldsymbol{f} \to \mathbf{0}$.
- Stronger assumption needed for the convergence of sequence.

Dual interpretation of ADMM

$$\boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{(k)} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho\|^2.$$

Dual interpretation of ADMM

$$\mathbf{x}^{(k+1)} \in \mathrm{Argmin}_{\mathbf{x}}\, F(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}^{(k)} - \mathbf{f} + \mathbf{u}^{(k)}/\rho\|^2.$$

Define $\mathbf{z}^{(k)} = \mathbf{u}^{(k)} - \rho(\mathbf{B}\mathbf{y}^{(k)} - \mathbf{f})$ and $\mathbf{w}^{(k+1)} = \rho(\mathbf{A}\mathbf{x}^{(k+1)} + \mathbf{B}\mathbf{y}^{(k)} - \mathbf{f} + \mathbf{u}^{(k)}/\rho)$

$$\begin{aligned}
\mathbf{0} \in \partial F(\mathbf{x}^{(k+1)}) + \mathbf{A}^T\mathbf{w}^{(k+1)} \quad &\Leftrightarrow \quad -\mathbf{A}^T\mathbf{w}^{(k+1)} \in \partial F(\mathbf{x}^{(k+1)}) \\
&\Leftrightarrow \mathbf{x}^{(k+1)} \in \partial F^*\left(-\mathbf{A}^T\mathbf{w}^{(k+1)}\right) \\
&\Leftrightarrow \quad -\mathbf{A}\mathbf{x}^{(k+1)} \in \partial(F^* \circ -\mathbf{A}^T)(\mathbf{w}^{(k+1)}) \\
&\Leftrightarrow \mathbf{w}^{(k+1)} - \rho\mathbf{A}\mathbf{x}^{(k+1)} \in \mathbf{w}^{(k+1)} + \rho\partial(F^* \circ -\mathbf{A}^T)(\mathbf{w}^{(k+1)}) \\
&\Leftrightarrow \mathbf{w}^{(k+1)} = \left(\mathbf{Id} + \rho\partial(F^* \circ -\mathbf{A}^T)\right)^{-1}(\mathbf{w}^{(k+1)} - \rho\mathbf{A}\mathbf{x}^{(k+1)}) \\
&\Leftrightarrow \boxed{\mathbf{w}^{(k+1)} = \left(\mathbf{Id} + \rho\partial(F^* \circ -\mathbf{A}^T)\right)^{-1}(2\mathbf{u}^{(k)} - \mathbf{z}^{(k)})}.
\end{aligned}$$

Dual interpretation of ADMM

$$\boldsymbol{y}^{(k+1)} \in \operatorname{Argmin}_{\boldsymbol{y}} R(\boldsymbol{y}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho\|^2.$$

Dual interpretation of ADMM

$$\boldsymbol{y}^{(k+1)} \in \operatorname{Argmin}_{\boldsymbol{y}} R(\boldsymbol{y}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho\|^2.$$

Recall $\boldsymbol{z}^{(k)} = \boldsymbol{u}^{(k)} - \rho(\boldsymbol{B}\boldsymbol{y}^{(k)} - \boldsymbol{f})$ and let $\boldsymbol{u}^{(k+1)} = \rho(\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \boldsymbol{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho)$

$$\begin{aligned}
\boldsymbol{0} \in \partial R(\boldsymbol{y}^{(k+1)}) + \boldsymbol{B}^T\boldsymbol{u}^{(k+1)} \quad &\Leftrightarrow \quad -\boldsymbol{B}^T\boldsymbol{u}^{(k+1)} \in \partial R(\boldsymbol{y}^{(k+1)}) \\
&\Leftrightarrow \quad \boldsymbol{y}^{(k+1)} \in \partial R^*\left(-\boldsymbol{B}^T\boldsymbol{u}^{(k+1)}\right) \\
&\Leftrightarrow \quad -\boldsymbol{B}\boldsymbol{y}^{(k+1)} \in \partial(R^* \circ -\boldsymbol{B}^T)(\boldsymbol{u}^{(k+1)}) \\
&\Leftrightarrow \quad \boldsymbol{u}^{(k+1)} - \rho\boldsymbol{B}\boldsymbol{y}^{(k+1)} \in \boldsymbol{u}^{(k+1)} + \rho\partial(R^* \circ -\boldsymbol{B}^T)(\boldsymbol{u}^{(k+1)}) \\
&\Leftrightarrow \quad \boldsymbol{u}^{(k+1)} = \left(\boldsymbol{Id} + \rho\partial(R^* \circ -\boldsymbol{B}^T)\right)^{-1}(\boldsymbol{u}^{(k+1)} - \rho\boldsymbol{B}\boldsymbol{y}^{(k+1)}) \\
&\Leftrightarrow \quad \boxed{\boldsymbol{u}^{(k+1)} = \left(\boldsymbol{Id} + \rho\partial(R^* \circ -\boldsymbol{B}^T)\right)^{-1}(\boldsymbol{w}^{(k+1)} + \boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)} - \rho\boldsymbol{f})}.
\end{aligned}$$

Dual interpretation of ADMM

Note that

$$\begin{aligned} \boldsymbol{z}^{(k+1)} &= \boldsymbol{u}^{(k+1)} - \rho(\boldsymbol{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{f}) \\ &= \rho(\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \boldsymbol{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{f} + \boldsymbol{u}^{(k)}/\rho) - \rho(\boldsymbol{B}\boldsymbol{y}^{(k+1)} - \boldsymbol{f}) \\ &= \rho\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)} \\ &= \boldsymbol{w}^{(k+1)} + \boldsymbol{z}^{(k)} - 2\boldsymbol{u}^{(k)} + \boldsymbol{u}^{(k)} \\ &= \boldsymbol{w}^{(k+1)} + \boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)}. \end{aligned}$$

Algorithm - Dual characterization of ADMM

$$\begin{aligned} \boldsymbol{w}^{(k+1)} &= \left(\boldsymbol{Id} + \rho\partial(F^* \circ -\boldsymbol{A}^T)\right)^{-1}(2\boldsymbol{u}^{(k)} - \boldsymbol{z}^{(k)}) \\ \boldsymbol{z}^{(k+1)} &= \boldsymbol{w}^{(k+1)} + \boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)} \\ \boldsymbol{u}^{(k+1)} &= \left(\boldsymbol{Id} + \rho\partial(R^* \circ -\boldsymbol{B}^T)\right)^{-1}(\boldsymbol{z}^{(k+1)} - \rho\boldsymbol{f}) \end{aligned}$$

The *dual problem*:

$$\min_{\boldsymbol{u}\in\mathbb{R}^p} F^*(-\boldsymbol{A}^*\boldsymbol{u}) + R^*(-\boldsymbol{B}^*\boldsymbol{u}) + \langle \boldsymbol{f} \mid \boldsymbol{u}\rangle.$$

Optimality condition

$$0 \in -\boldsymbol{A}\partial F^*(-\boldsymbol{A}^*\boldsymbol{u}^\star) - \boldsymbol{B}\partial R^*(-\boldsymbol{B}^*\boldsymbol{u}^\star) + \boldsymbol{f}$$

Algorithm - Dual characterization of ADMM

$$\begin{aligned}
\boldsymbol{w}^{(k+1)} &= \left(\boldsymbol{Id} + \rho\partial(F^* \circ -\boldsymbol{A}^T)\right)^{-1}(2\boldsymbol{u}^{(k)} - \boldsymbol{z}^{(k)}) \\
\boldsymbol{z}^{(k+1)} &= \boldsymbol{w}^{(k+1)} + \boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)} \\
\boldsymbol{u}^{(k+1)} &= \left(\boldsymbol{Id} + \rho\partial(R^* \circ -\boldsymbol{B}^T)\right)^{-1}(\boldsymbol{z}^{(k+1)} - \rho\boldsymbol{f})
\end{aligned}$$

Douglas–Rachford splitting method

Convergence of ADMM

Problem - Monotone inclusion

$$\text{find } \boldsymbol{x} \in \mathbb{R}^n \quad \text{such that} \quad \mathbf{0} \in \mathcal{A}(\boldsymbol{x}) + \mathcal{B}(\boldsymbol{x}),$$

where

- $\mathcal{A}, \mathcal{B} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are maximal monotone.
- the resolvents of A, B are easy to compute.
- $\mathrm{zer}(\mathcal{A} + \mathcal{B}) \neq \emptyset$.

Example - Non-smooth optimization

Let $F, R \in \Gamma_0(\mathbb{R}^n)$

$$\mathbf{0} \in \partial F(\boldsymbol{x}^\star) + \partial R(\boldsymbol{x}^\star) \quad \Longrightarrow \quad \boldsymbol{x}^\star \in \mathrm{Argmin}(F + R).$$

Fixed-point characterization

Given $\boldsymbol{x}^\star \in \mathrm{zer}(\mathcal{A}+\mathcal{B})$, let $\gamma > 0$, there exists $\boldsymbol{z}^\star \in \mathbb{R}^n$ such hat

$$
\mathbf{0} \in \gamma\mathcal{A}(\boldsymbol{x}^\star) + \gamma\mathcal{B}(\boldsymbol{x}^\star) \iff \begin{cases} \boldsymbol{x}^\star - \boldsymbol{z}^\star \in \gamma\mathcal{A}(\boldsymbol{x}^\star) \\ \boldsymbol{z}^\star - \boldsymbol{x}^\star \in \gamma\mathcal{B}(\boldsymbol{x}^\star) \end{cases}
$$

$$
\iff \begin{cases} 2\boldsymbol{x}^\star - \boldsymbol{z}^\star \in \boldsymbol{x}^\star + \gamma\mathcal{A}(\boldsymbol{x}^\star) \\ \boldsymbol{z}^\star \in \boldsymbol{x}^\star + \gamma\mathcal{B}(\boldsymbol{x}^\star) \end{cases}
$$

$$
\iff \begin{cases} \boldsymbol{x}^\star = (\boldsymbol{Id} + \gamma\mathcal{A})^{-1}(2\boldsymbol{x}^\star - \boldsymbol{z}^\star) \\ \boldsymbol{x}^\star = (\boldsymbol{Id} + \gamma\mathcal{B})^{-1}(\boldsymbol{z}^\star) \end{cases}
$$

Fixed-point characterization

Given $\mathbf{x}^\star \in \mathrm{zer}(\mathcal{A}+\mathcal{B})$, let $\gamma > 0$, there exists $\mathbf{z}^\star \in \mathbb{R}^n$ such hat

$$\mathbf{0} \in \gamma\mathcal{A}(\mathbf{x}^\star) + \gamma\mathcal{B}(\mathbf{x}^\star) \iff \begin{cases} \mathbf{x}^\star - \mathbf{z}^\star \in \gamma\mathcal{A}(\mathbf{x}^\star) \\ \mathbf{z}^\star - \mathbf{x}^\star \in \gamma\mathcal{B}(\mathbf{x}^\star) \end{cases}$$

$$\iff \begin{cases} 2\mathbf{x}^\star - \mathbf{z}^\star \in \mathbf{x}^\star + \gamma\mathcal{A}(\mathbf{x}^\star) \\ \mathbf{z}^\star \in \mathbf{x}^\star + \gamma\mathcal{B}(\mathbf{x}^\star) \end{cases}$$

$$\iff \begin{cases} \mathbf{x}^\star = (\mathbf{Id} + \gamma\mathcal{A})^{-1}(2\mathbf{x}^\star - \mathbf{z}^\star) \\ \mathbf{x}^\star = (\mathbf{Id} + \gamma\mathcal{B})^{-1}(\mathbf{z}^\star) \end{cases}$$

Let $\lambda > 0$, then

$$\begin{cases} \mathbf{z}^\star = \mathbf{z}^\star + \lambda\big((\mathbf{Id} + \gamma\mathcal{A})^{-1}(2\mathbf{x}^\star - \mathbf{z}^\star) - \mathbf{x}^\star\big) \\ \mathbf{x}^\star = (\mathbf{Id} + \gamma\mathcal{B})^{-1}(\mathbf{z}^\star) \end{cases}$$

Algorithm - Douglas–Rachford splitting method

Let $\gamma > 0$ and $\lambda \in]0, 2[$,

$$\begin{cases} \mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \lambda\big(\mathcal{J}_{\gamma\mathcal{A}}(2\mathbf{x}^{(k)} - \mathbf{z}^{(k)}) - \mathbf{x}^{(k)}\big) \\ \mathbf{x}^{(k+1)} = \mathcal{J}_{\gamma\mathcal{B}}(\mathbf{z}^{(k+1)}) \end{cases}$$

Douglas–Rachford splitting method

Algorithm - Douglas–Rachford splitting method

Let $\gamma > 0$ and $\lambda \in]0,2[$,

$$\begin{cases} \boldsymbol{z}^{(k+1)} = \boldsymbol{z}^{(k)} + \lambda\big(\mathcal{J}_{\gamma\mathcal{A}}(2\boldsymbol{x}^{(k)} - \boldsymbol{z}^{(k)}) - \boldsymbol{x}^{(k)}\big) \\ \boldsymbol{x}^{(k+1)} = \mathcal{J}_{\gamma\mathcal{B}}(\boldsymbol{z}^{(k+1)}) \end{cases}$$

Alternative expression

$$\begin{cases} \boldsymbol{w}^{(k+1)} = \mathcal{J}_{\gamma\mathcal{A}}(2\boldsymbol{x}^{(k)} - \boldsymbol{z}^{(k)}) \\ \boldsymbol{z}^{(k+1)} = \boldsymbol{z}^{(k)} + \lambda\big(\boldsymbol{w}^{(k+1)} - \boldsymbol{x}^{(k)}\big) \\ \boldsymbol{x}^{(k+1)} = \mathcal{J}_{\gamma\mathcal{B}}(\boldsymbol{z}^{(k+1)}) \end{cases}$$

Douglas–Rachford splitting method

Algorithm - Douglas–Rachford splitting method

Let $\gamma > 0$ and $\lambda \in]0, 2[$,

$$\begin{cases} \mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \lambda\big(\mathcal{J}_{\gamma\mathcal{A}}(2\mathbf{x}^{(k)} - \mathbf{z}^{(k)}) - \mathbf{x}^{(k)}\big) \\ \mathbf{x}^{(k+1)} = \mathcal{J}_{\gamma\mathcal{B}}(\mathbf{z}^{(k+1)}) \end{cases}$$

Fixed-point iteration

$$\begin{aligned}
& \mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \lambda\big(\mathcal{J}_{\gamma\mathcal{A}}(2\mathbf{x}^{(k)} - \mathbf{z}^{(k)}) - \mathbf{x}^{(k)}\big) \\
\iff\quad & \mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \lambda\Big(\mathcal{J}_{\gamma\mathcal{A}}\big(2\mathcal{J}_{\gamma\mathcal{B}}(\mathbf{z}^{(k)}) - \mathbf{z}^{(k)}\big) - \mathcal{J}_{\gamma\mathcal{B}}(\mathbf{z}^{(k)})\Big) \\
\iff\quad & \mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \frac{\lambda}{2}\big(2\mathcal{J}_{\gamma\mathcal{A}} \circ (2\mathcal{J}_{\gamma\mathcal{B}} - \mathbf{Id})(\mathbf{z}^{(k)}) - 2\mathcal{J}_{\gamma\mathcal{B}}(\mathbf{z}^{(k)}) + \mathbf{z}^{(k)} - \mathbf{z}^{(k)}\big) \\
\iff\quad & \mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \frac{\lambda}{2}\big(2\mathcal{J}_{\gamma\mathcal{A}} \circ (2\mathcal{J}_{\gamma\mathcal{B}} - \mathbf{Id})(\mathbf{z}^{(k)}) - (2\mathcal{J}_{\gamma\mathcal{B}} - \mathbf{Id})(\mathbf{z}^{(k)}) - \mathbf{z}^{(k)}\big) \\
\iff\quad & \mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} + \frac{\lambda}{2}\big((2\mathcal{J}_{\gamma\mathcal{A}} - \mathbf{Id}) \circ (2\mathcal{J}_{\gamma\mathcal{B}} - \mathbf{Id})(\mathbf{z}^{(k)}) - \mathbf{z}^{(k)}\big) \\
\iff\quad & \mathbf{z}^{(k+1)} = (1-\lambda)\mathbf{z}^{(k)} + \lambda \boxed{\frac{1}{2}\big((2\mathcal{J}_{\gamma\mathcal{A}} - \mathbf{Id}) \circ (2\mathcal{J}_{\gamma\mathcal{B}} - \mathbf{Id}) + \mathbf{Id}\big)}(\mathbf{z}^{(k)})
\end{aligned}$$

Douglas–Rachford splitting method

Algorithm - Douglas–Rachford splitting method

Let $\gamma > 0$ and $\lambda \in]0,2[$,

$$\begin{cases} \boldsymbol{z}^{(k+1)} = \boldsymbol{z}^{(k)} + \lambda\big(\mathcal{J}_{\gamma\mathcal{A}}(2\boldsymbol{x}^{(k)} - \boldsymbol{z}^{(k)}) - \boldsymbol{x}^{(k)}\big) \\ \boldsymbol{x}^{(k+1)} = \mathcal{J}_{\gamma\mathcal{B}}(\boldsymbol{z}^{(k+1)}) \end{cases}$$

Proposition - Fixed-point operator of Douglas–Rachford

Let $\gamma > 0$, and

$$\mathcal{F}_{\mathrm{DR}} = \frac{1}{2}\big((2\mathcal{J}_{\gamma\mathcal{A}} - \boldsymbol{Id}) \circ (2\mathcal{J}_{\gamma\mathcal{B}} - \boldsymbol{Id}) + \boldsymbol{Id}\big).$$

Then

- $\mathcal{F}_{\mathrm{DR}}$ is firmly non-expansive.
- $(1-\lambda)\boldsymbol{Id} + \lambda\mathcal{F}_{\mathrm{DR}}$ is $\frac{\lambda}{2}$-averaged non-expansive.

Convergence

Proposition - Convergence

Let $\mathcal{A}, \mathcal{B} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximally monotone such that $\operatorname{zer}(\mathcal{A}+\mathcal{B}) \neq \emptyset$. Let $\lambda_k \in [0,2]$ such that $\sum_k \lambda_k(2-\lambda_k) = +\infty$, and let $\gamma > 0$. Considering

$$\begin{cases} \boldsymbol{w}^{(k+1)} = \mathcal{J}_{\gamma\mathcal{A}}(2\boldsymbol{x}^{(k)} - \boldsymbol{z}^{(k)}) \\ \boldsymbol{z}^{(k+1)} = \boldsymbol{z}^{(k)} + \lambda_k\left(\boldsymbol{w}^{(k)} - \boldsymbol{x}^{(k)}\right) \\ \boldsymbol{x}^{(k+1)} = \mathcal{J}_{\gamma\mathcal{B}}(\boldsymbol{z}^{(k+1)}) \end{cases}$$

Then there exists $\boldsymbol{z}^\star \in \operatorname{fix}(\mathcal{F}_{\mathrm{DR}})$ such that $\boldsymbol{z}^{(k)} \to \boldsymbol{z}^\star$. Set $\boldsymbol{x}^\star = \mathcal{J}_{\gamma\mathcal{B}}(\boldsymbol{z}^\star)$, then

- $\boldsymbol{w}^{(k)} - \boldsymbol{x}^{(k)} \to 0$.
- $\boldsymbol{x}^{(k)} \to \boldsymbol{x}^\star$ and $\boldsymbol{w}^{(k)} \to \boldsymbol{x}^\star$.

- For ADMM, if $\boldsymbol{A}$ and $\boldsymbol{B}$ have full column rank, then convergence of sequence can be derived.

Two variants of ADMM

Scaled iteration, and precondition

Dual scaled ADMM

Augmented Lagrangian Let $\rho > 0$

$$\mathscr{L}_\rho(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{u}) = F(\boldsymbol{x}) + R(\boldsymbol{y}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} - \boldsymbol{f} + \boldsymbol{u}/\rho\|^2 - \frac{\|\boldsymbol{u}\|^2}{2\rho}.$$

Dual scaled ADMM

Augmented Lagrangian Let $\rho > 0$

$$\mathscr{L}_\rho(\mathbf{x}, \mathbf{y}; \mathbf{u}) = F(\mathbf{x}) + R(\mathbf{y}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{f} + \mathbf{u}/\rho\|^2 - \frac{\|\mathbf{u}\|^2}{2\rho}.$$

Scale dual $\mathbf{v} = \mathbf{u}/\rho$

$$\mathbf{u}^+ = \mathbf{u} + \rho(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{f}) \implies \mathbf{v}^+ = \mathbf{v} + (\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{f}).$$

Algorithm - Dual scaled ADMM

$$\begin{aligned}
\mathbf{x}^{(k+1)} &\in \mathrm{Argmin}_x\, F(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}^{(k)} - \mathbf{f} + \mathbf{v}^{(k)}\|^2, \\
\mathbf{y}^{(k+1)} &\in \mathrm{Argmin}_y\, R(\mathbf{y}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{(k+1)} + \mathbf{B}\mathbf{y} - \mathbf{f} + \mathbf{v}^{(k)}\|^2, \\
\mathbf{v}^{(k+1)} &= \mathbf{v}^{(k)} + (\mathbf{A}\mathbf{x}^{(k+1)} + \mathbf{B}\mathbf{y}^{(k+1)} - \mathbf{f}).
\end{aligned}$$

Preconditioned ADMM

$$\boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_x F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{(k)} - \boldsymbol{f} + \boldsymbol{v}^{(k)}\|^2.$$

Preconditioned ADMM

$$x^{(k+1)} \in \mathrm{Argmin}_x \, F(x) + \frac{\rho}{2}\|Ax + By^{(k)} - f + v^{(k)}\|^2.$$

- Let $w^{(k)} = -(By^{(k)} - f + v^{(k)})$

$$x^{(k+1)} \in \mathrm{Argmin}_x \, F(x) + \frac{\rho}{2}\|Ax - w^{(k)}\|^2.$$

  No closed form solution due to A.

Preconditioned ADMM

$$\boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_x F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{(k)} - \boldsymbol{f} + \boldsymbol{v}^{(k)}\|^2.$$

- Let $\boldsymbol{w}^{(k)} = -(\boldsymbol{B}\boldsymbol{y}^{(k)} - \boldsymbol{f} + \boldsymbol{v}^{(k)})$

  $$\boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{w}^{(k)}\|^2.$$

  No closed form solution due to $\boldsymbol{A}$.

- Let $\boldsymbol{Q}$ be symmetric and positive definite, and

  $$\boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{w}^{(k)}\|^2 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|_{\boldsymbol{Q}}^2.$$

Preconditioned ADMM

$$\boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_x F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{(k)} - \boldsymbol{f} + \boldsymbol{v}^{(k)}\|^2.$$

- Let $\boldsymbol{w}^{(k)} = -(\boldsymbol{B}\boldsymbol{y}^{(k)} - \boldsymbol{f} + \boldsymbol{v}^{(k)})$

$$\boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{w}^{(k)}\|^2.$$

  No closed form solution due to $\boldsymbol{A}$.

- Let $\boldsymbol{Q}$ be symmetric and positive definite, and

$$\boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{w}^{(k)}\|^2 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|_{\boldsymbol{Q}}^2.$$

- Choose $\boldsymbol{Q} = \frac{1}{\tau}\boldsymbol{Id} - \rho\boldsymbol{A}^T\boldsymbol{A}$, τ is smaller enough such that $\boldsymbol{Q}$ is SPD

$$\begin{aligned}
& \boldsymbol{x}^{(k+1)} \in \operatorname{Argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) + \tfrac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{w}^{(k)}\|^2 + \tfrac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|_{\boldsymbol{Q}}^2 \\
\Leftrightarrow\quad & \boldsymbol{0} \in \partial F(\boldsymbol{x}^{(k+1)}) + \rho\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}^{(k+1)} - \boldsymbol{w}^{(k)}) + \boldsymbol{Q}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \\
\Leftrightarrow\quad & \rho\boldsymbol{A}^T\boldsymbol{w}^{(k)} \in \partial F(\boldsymbol{x}^{(k+1)}) + \rho\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}^{(k+1)} + \tfrac{1}{\tau}\boldsymbol{x}^{(k+1)} - \rho\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}^{(k+1)} - \boldsymbol{Q}\boldsymbol{x}^{(k)} \\
\Leftrightarrow\quad & \rho\boldsymbol{A}^T\boldsymbol{w}^{(k)} + \boldsymbol{Q}\boldsymbol{x}^{(k)} \in \partial F(\boldsymbol{x}^{(k+1)}) + \tfrac{1}{\tau}\boldsymbol{x}^{(k+1)} \\
\Leftrightarrow\quad & \boldsymbol{x}^{(k+1)} = (\boldsymbol{Id}/\tau + \partial F)^{-1}\left(\rho\boldsymbol{A}^T\boldsymbol{w}^{(k)} + \boldsymbol{Q}\boldsymbol{x}^{(k)}\right)
\end{aligned}$$

References

- H. Bauschke and P. L. Combettes. “Convex Analysis and Monotone Operator Theory in Hilbert Spaces”. Springer, 2011.
- S. Boyd, et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers”. Foundations and Trends® in Machine learning 3.1 (2011): 1-122.
- J. Eckstein, and D. P. Bertsekas. “On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators”. Mathematical Programming 55.1-3 (1992): 293-318.
- E., Esser, X. Zhang, and T. F. Chan. “A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science”. SIAM Journal on Imaging Sciences 3.4 (2010): 1015-1046.