An Introduction to Non-smooth Optimization

Lecture 04 - Primal-Dual Splitting Methods

Jingwei LIANG

Institute of Natural Sciences, Shanghai Jiao Tong University

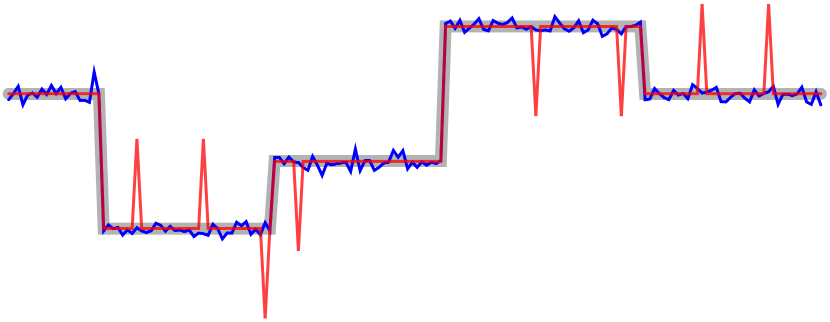
Email: optimization.sjtu@gmail.com

Office: Room 355, No. 6 Science Building

Previously

TV based image processing Let $p \in \{1, 2\}$,

$$\min_{\boldsymbol{x} \in \mathbb{R}^{m \times n}} \ \mu \|\nabla \boldsymbol{x}\|_1 + \frac{1}{2} \|\boldsymbol{f} - \mathcal{F}\boldsymbol{x}\|_p^p .$$



Previously

TV based image processing Let $p \in \{1, 2\}$,

$$\min_{\mathbf{x} \in \mathbb{R}^{m \times n}} \ \mu \|\nabla \mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{f} - \mathcal{F}\mathbf{x}\|_p^p.$$

Problem - Non-smooth optimization problem

Consider

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \Phi(\mathbf{x}) \stackrel{\text{def}}{=} F(\mathbf{x}) + R(\mathbf{K}\mathbf{x}) \right\},$$

with

- $F \in \Gamma_0(\mathbb{R}^n)$, $R \in \Gamma_0(\mathbb{R}^m)$
- $\mathbf{K} : \mathbb{R}^n \to \mathbb{R}^m$ is bounded linear.

Outline

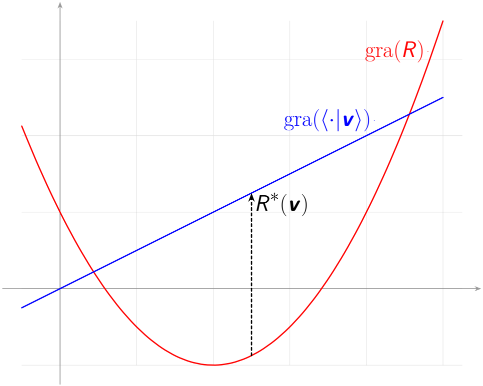
1. Conjugation
2. Duality
3. Primal-Dual splitting method
4. Variants

Definition - Conjugate

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$, the conjugate of R is defined by

$$R^*(\boldsymbol{u}) \stackrel{\text{def}}{=} \sup_{\boldsymbol{x}\in\mathbb{R}^n} \left(\langle \boldsymbol{x} \mid \boldsymbol{u}\rangle - R(\boldsymbol{x})\right).$$

- Biconjugate $R^{**} = (R^*)^*$.
- Also called *Fenchel conjugate*, *Legendre transform*, or *Legendre-Fenchel transform*.



Conjugate

Definition - Conjugate

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$, the conjugate of R is defined by

$$R^*(\boldsymbol{u}) \stackrel{\text{def}}{=} \sup_{\boldsymbol{x} \in \mathbb{R}^n} \left(\langle \boldsymbol{x} \mid \boldsymbol{u} \rangle - R(\boldsymbol{x}) \right).$$

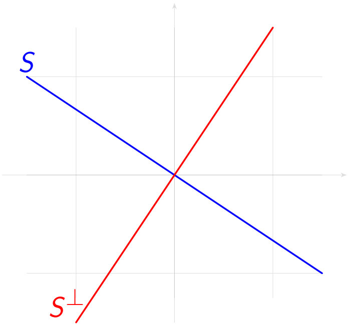
Example - Support function

Let $S \subseteq \mathbb{R}^n$ be a non-empty convex set, the support function of S is defined by

$$\sigma_S(\boldsymbol{u}) \stackrel{\text{def}}{=} \sup_{\boldsymbol{x} \in S} \langle \boldsymbol{x} \mid \boldsymbol{u} \rangle = \iota_S^*(\boldsymbol{u}).$$

Let S be a linear subspace of $\mathbb{R}^n$, then

$$\sigma_S(\boldsymbol{u}) = \iota_{S^\perp}(\boldsymbol{u}).$$



Conjugate

Definition - Conjugate

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$, the conjugate of R is defined by

$$R^*(\boldsymbol{u}) \stackrel{\text{def}}{=} \sup_{\boldsymbol{x}\in\mathbb{R}^n} \left(\langle \boldsymbol{x} \mid \boldsymbol{u}\rangle - R(\boldsymbol{x})\right).$$

Example - ℓ_2-norm square

Let $R = \frac{1}{2}\|\boldsymbol{x}\|^2$,

$$R^*(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{u}\|^2.$$

- Self-conjugacy

Conjugate

Definition - Conjugate

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$, the conjugate of R is defined by

$$R^*(\boldsymbol{u}) \stackrel{\text{def}}{=} \sup_{\boldsymbol{x} \in \mathbb{R}^n} \left(\langle \boldsymbol{x} \mid \boldsymbol{u} \rangle - R(\boldsymbol{x}) \right).$$

Example - ℓ_2-norm square

Let $R = \frac{1}{2}\|\boldsymbol{x}\|^2$,

$$R^*(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{u}\|^2.$$

- Self-conjugacy

Definition - Dual norm

Let $\|\cdot\|$ be a norm defined on $\mathbb{R}^n$. Its *dual norm*, denoted by $\|\cdot\|_*$, is defined as

$$\|\boldsymbol{u}\|_* = \sup \left\{ \langle \boldsymbol{u} \mid \boldsymbol{x} \rangle \ : \ \|\boldsymbol{x}\| \leq 1 \right\}.$$

Definition - Conjugate

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$, the conjugate of R is defined by

$$R^*(\boldsymbol{u}) \stackrel{\text{def}}{=} \sup_{\boldsymbol{x}\in\mathbb{R}^n} \left(\langle \boldsymbol{x} \,|\, \boldsymbol{u}\rangle - R(\boldsymbol{x})\right).$$

Example - ℓ_2-norm square

Let $R = \frac{1}{2}\|\boldsymbol{x}\|^2$,

$$R^*(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{u}\|^2.$$

- Self-conjugacy

Definition - Dual norm

Let $\|\cdot\|$ be a norm defined on $\mathbb{R}^n$. Its *dual norm*, denoted by $\|\cdot\|_*$, is defined as

$$\|\boldsymbol{u}\|_* = \sup\left\{\langle \boldsymbol{u} \,|\, \boldsymbol{x}\rangle \ : \ \|\boldsymbol{x}\| \le 1\right\}.$$

Example - Conjugate of norm

Let $R = \|\boldsymbol{x}\|$ be a norm with dual norm $\|\cdot\|_*$. Then

$$R^*(\boldsymbol{u}) = \begin{cases} 0 & \|\boldsymbol{u}\|_* \le 1, \\ +\infty & o.w. \end{cases}$$

Properties

Convexity of conjugate R^* is closed and convex.

Properties

Convexity of conjugate R^* is closed and convex.

Fenchel–Young inequality Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper. Then

$$R(\boldsymbol{x}) + R^*(\boldsymbol{u}) \geq \langle \boldsymbol{x} \mid \boldsymbol{u} \rangle.$$

Properties

Convexity of conjugate R^* is closed and convex.

Fenchel–Young inequality Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper. Then

$$R(\boldsymbol{x}) + R^*(\boldsymbol{u}) \geq \langle \boldsymbol{x} \mid \boldsymbol{u} \rangle.$$

Let F, R be functions from $\mathbb{R}^n$ to $[-\infty, +\infty]$. Then

- $R^{**} \leq R$.
- $F \leq R \implies \left[F^* \geq R^* \quad \text{and} \quad F^{**} \leq R^{**}\right]$.

Properties

Convexity of conjugate R^* is closed and convex.

Fenchel–Young inequality Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper. Then

$$R(\boldsymbol{x}) + R^*(\boldsymbol{u}) \geq \langle \boldsymbol{x} \mid \boldsymbol{u} \rangle.$$

Let F, R be functions from $\mathbb{R}^n$ to $[-\infty, +\infty]$. Then

- $R^{**} \leq R$.
- $F \leq R \implies \left[F^* \geq R^* \quad \text{and} \quad F^{**} \leq R^{**}\right]$.

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$. Then

- $\forall \alpha > 0$,

$$(\alpha R)^* = \alpha R^*(\cdot/\alpha) \quad \text{and} \quad \big(\alpha R(\cdot/\alpha)\big)^* = \alpha R^*.$$

- Let $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^n$ be bijective. Then

$$(R \circ \boldsymbol{K})^* = R^* \circ \boldsymbol{K}^{*-1}.$$

The Fenchel–Moreau theorem

Theorem - Fenchel–Moreau

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper. Then R is closed and convex if and only if $R = R^{**}$. In this case, R^* is proper as well.

The Fenchel–Moreau theorem

Theorem - Fenchel–Moreau

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ be proper. Then R is closed and convex if and only if $R = R^{**}$. In this case, R^* is proper as well.

Corollary

Let $R \in \Gamma_0(\mathbb{R}^n)$, then $R^* \in \Gamma_0(\mathbb{R}^n)$ and $R^{**} = R$.

Moreau's decomposition

Theorem - Moreau's decomposition

Let $R \in \Gamma_0(\mathbb{R}^n)$ be continuous and convex, let $\gamma > 0$. Then the following hold

- **Moreau's identify** given any $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{x} = \operatorname{prox}_{\gamma R}(\boldsymbol{x}) + \gamma \operatorname{prox}_{R^*/\gamma}\left(\frac{\boldsymbol{x}}{\gamma}\right).$$

Moreau's decomposition

Theorem - Moreau's decomposition

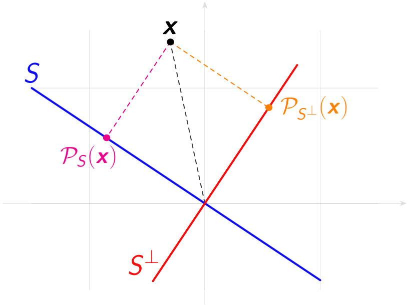
Let $R \in \Gamma_0(\mathbb{R}^n)$ be continuous and convex, let $\gamma > 0$. Then the following hold

- **Moreau's identify** given any $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{x} = \mathrm{prox}_{\gamma R}(\boldsymbol{x}) + \gamma\, \mathrm{prox}_{R^*/\gamma}\left(\frac{\boldsymbol{x}}{\gamma}\right).$$

- For any $\boldsymbol{x} \in \mathbb{R}^n$,

$$R\big(\mathrm{prox}_{\gamma R}(\boldsymbol{x})\big) + R^*\big(\mathrm{prox}_{R^*/\gamma}(\boldsymbol{x}/\gamma)\big) = \langle \mathrm{prox}_{\gamma R}(\boldsymbol{x}) \mid \mathrm{prox}_{R^*/\gamma}(\boldsymbol{x}/\gamma) \rangle.$$



Theorem - Subdifferential and conjugation

Let $R \in \Gamma_0(\mathbb{R}^n)$, let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{v} \in \mathbb{R}^n$. Then the following are equivalent

- $(\boldsymbol{x}, \boldsymbol{u}) \in \mathrm{gra}(\partial R)$.
- $R(\boldsymbol{x}) + R^*(\boldsymbol{u}) = \langle \boldsymbol{x} \mid \boldsymbol{u} \rangle$.
- $(\boldsymbol{u}, \boldsymbol{x}) \in \mathrm{gra}(\partial R^*)$.

- When $R \in \Gamma_0(\mathbb{R}^n)$, $\partial R^* = (\partial R)^{-1}$.

Theorem - Strong convexity and conjugation

Let $F : \mathbb{R}^n \to \mathbb{R}$ be continuous and convex, let $\beta > 0$. Then the following are equivalent

- ∇F is $1/\beta$-Lipschitz continuous.
- F^* is β-strongly convex.

Duality

Fenchel-Rockafellar duality

Fenchel-Rockafellar duality

Proposition - Duality inequality

Let $F : \mathbb{R}^n \to]-\infty, +\infty]$ be proper and $R : \mathbb{R}^m \to]-\infty, +\infty]$ be proper, let $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear. Then

$$(\forall \boldsymbol{x} \in \mathbb{R}^n)(\forall \boldsymbol{u} \in \mathbb{R}^m) \qquad F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \geq -F^*(-\boldsymbol{K}^*\boldsymbol{u}) - R^*(\boldsymbol{u}).$$

and

$$\inf(F + R \circ \boldsymbol{K})(\mathbb{R}^n) \geq -\inf(F^* \circ -\boldsymbol{K}^* + R^*)(\mathbb{R}^n).$$

Definition - Fenchel-Rockafellar duality

Let $F : \mathbb{R}^n \to]-\infty, +\infty]$, $R : \mathbb{R}^m \to]-\infty, +\infty]$ and $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear. The *primal problem* associated with the composite function $F + R \circ \boldsymbol{K}$ is

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}), \tag{$\mathscr{P}$}$$

its **dual problem** is

$$\min_{\boldsymbol{u} \in \mathbb{R}^m} F^*(-\boldsymbol{K}^*\boldsymbol{u}) + R^*(\boldsymbol{u}). \tag{$\mathscr{D}$}$$

The *primal and dual optimal values* are

$$\mu = \inf(F + R \circ \boldsymbol{K})(\mathbb{R}^n) \quad \text{and} \quad \mu^* = \inf(F^* \circ -\boldsymbol{K}^* + R^*)(\mathbb{R}^m)$$

and the *duality gap* is

$$\mathcal{G}_{F,R,\boldsymbol{K}} = \begin{cases} 0 : & \text{if } \mu = -\mu^* \in \{-\infty, +\infty\}, \\ \mu + \mu^* : & \text{o.w.} \end{cases}$$

Fenchel-Rockafellar duality

Definition - Duality gap

Let $F : \mathbb{R}^n \to]-\infty, +\infty]$ and $R : \mathbb{R}^m \to]-\infty, +\infty]$ be proper, let $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear. Set

$$\mu = \inf(F + R \circ \boldsymbol{K})(\mathbb{R}^n) \quad \text{and} \quad \mu^* = \inf(F^* \circ -\boldsymbol{K}^* + R^*)(\mathbb{R}^m).$$

Then the following hold

- $\mu \geq -\mu^*$.
- $\mathcal{G}_{F,R,\boldsymbol{K}} \in [0, +\infty]$.
- $\mu = -\mu^* \iff \mathcal{G}_{F,R,\boldsymbol{K}} = 0$.

Proposition - Strong duality

Let $F \in \Gamma_0(\mathbb{R}^n)$ and $R \in \Gamma_0(\mathbb{R}^m)$ such that

$$\mathrm{ri}\big(\boldsymbol{K}\mathrm{dom}(F)\big) \cap \mathrm{ri}\big(\mathrm{dom}(R)\big) \neq \emptyset.$$

Then

$$\inf(F + R \circ \boldsymbol{K})(\mathbb{R}^n) = -\min(F^* \circ -\boldsymbol{K}^* + R^*)(\mathbb{R}^m).$$

Primal-Dual splitting method

Algorithm, and convergence

Proposition - Saddle-point problem

Let $F:\mathbb{R}^n \to]-\infty,+\infty]$ and $R:\mathbb{R}^m \to]-\infty,+\infty]$ be proper, let $\boldsymbol{K}:\mathbb{R}^n \to \mathbb{R}^m$ be bounded linear such that $\mathrm{dom}(R) \cap \boldsymbol{K}\mathrm{dom}(F) \neq \emptyset$. Then the following hold

- The *primal problem* is

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

Saddle-point problem

Proposition - Saddle-point problem

Let $F:\mathbb{R}^n \to]-\infty,+\infty]$ and $R:\mathbb{R}^m \to]-\infty,+\infty]$ be proper, let $\boldsymbol{K}:\mathbb{R}^n \to \mathbb{R}^m$ be bounded linear such that $\mathrm{dom}(R) \cap \boldsymbol{K}\mathrm{dom}(F) \neq \emptyset$. Then the following hold

- The *primal problem* is

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

- The *dual problem* is

$$\min_{\boldsymbol{u}\in\mathbb{R}^m} F^*(-\boldsymbol{K}^*\boldsymbol{u}) + R^*(\boldsymbol{u}). \tag{$\mathscr{D}$}$$

Proposition - Saddle-point problem

Let $F : \mathbb{R}^n \to]-\infty, +\infty]$ and $R : \mathbb{R}^m \to]-\infty, +\infty]$ be proper, let $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear such that $\mathrm{dom}(R) \cap \boldsymbol{K}\mathrm{dom}(F) \neq \emptyset$. Then the following hold

- The *primal problem* is

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

- The *dual problem* is

$$\min_{\boldsymbol{u}\in\mathbb{R}^m} F^*(-\boldsymbol{K}^*\boldsymbol{u}) + R^*(\boldsymbol{u}). \tag{$\mathscr{D}$}$$

- The *saddle-point problem* is

$$\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) = \min_{\boldsymbol{x}\in\mathrm{dom}(F)} \max_{\boldsymbol{u}\in\mathrm{dom}(R^*)} F(\boldsymbol{x}) + \langle \boldsymbol{K}\boldsymbol{x} \mid \boldsymbol{u} \rangle - R^*(\boldsymbol{u}).$$

Proposition - Saddle-point problem

Let $F : \mathbb{R}^n \to]-\infty, +\infty]$ and $R : \mathbb{R}^m \to]-\infty, +\infty]$ be proper, let $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear such that $\mathrm{dom}(R) \cap \boldsymbol{K}\mathrm{dom}(F) \neq \emptyset$. Then the following hold

- The *primal problem* is

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

- The *dual problem* is

$$\min_{\boldsymbol{u}\in\mathbb{R}^m} F^*(-\boldsymbol{K}^*\boldsymbol{u}) + R^*(\boldsymbol{u}). \tag{$\mathscr{D}$}$$

- The *saddle-point problem* is

$$\mathscr{L}(\boldsymbol{x};\boldsymbol{u}) = \min_{\boldsymbol{x}\in\mathrm{dom}(F)} \max_{\boldsymbol{u}\in\mathrm{dom}(R^*)} F(\boldsymbol{x}) + \langle \boldsymbol{K}\boldsymbol{x} \mid \boldsymbol{u} \rangle - R^*(\boldsymbol{u}).$$

- Suppose the optimal values μ of ($\mathscr{P}$) and μ^* of ($\mathscr{D}$) satisfy $\mu = -\mu^* \in \mathbb{R}$, let $(\boldsymbol{x}^\star, \boldsymbol{u}^\star) \in \mathrm{dom}(F) \times \mathrm{dom}(R^*)$. Then $(\boldsymbol{x}^\star, \boldsymbol{u}^\star)$ is a saddle point of $\mathscr{L}(\boldsymbol{x};\boldsymbol{u})$ if and only if

$$-\boldsymbol{K}^*\boldsymbol{u}^\star \in \partial F(\boldsymbol{x}^\star) \quad \text{and} \quad \boldsymbol{K}\boldsymbol{x}^\star \in \partial R^*(\boldsymbol{u}^\star).$$

A Primal–Dual splitting method

Let

$$L = \|K\| \overset{\text{def}}{=} \max\left\{\|\boldsymbol{K}\boldsymbol{x}\| \ : \ \boldsymbol{x} \in \mathbb{R}^n \quad \text{such that} \quad \|\boldsymbol{x}\| \le 1\right\}.$$

Algorithm - Primal–Dual splitting method [Chambolle & Pock '11]

initial: $(\boldsymbol{x}^{(0)}, \boldsymbol{u}^{(0)}) \in \mathrm{dom}(F) \times \mathrm{dom}(R)$ and $\bar{\boldsymbol{x}}^{(0)} = \boldsymbol{x}^{(0)}$; $\theta \in [0,1]$ and $\sigma, \tau > 0$ such that

$$\sigma\tau L^2 \le 1.$$

repeat:

1. Dual update: $\boldsymbol{u}^{(k+1)} = \mathrm{prox}_{\sigma R^*}(\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K}\bar{\boldsymbol{x}}^{(k)})$
2. Primal update: $\boldsymbol{x}^{(k+1)} = \mathrm{prox}_{\tau F}(\boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k+1)})$
3. Extrapolation: $\bar{\boldsymbol{x}}^{(k+1)} = \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$

until: stopping criterion is satisfied.

Fixed-point characterization

Consider the following order of iteration

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= \operatorname{prox}_{\tau F}(\boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k)}) \\ \bar{\boldsymbol{x}}^{(k+1)} &= \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \\ \boldsymbol{u}^{(k+1)} &= \operatorname{prox}_{\sigma R^*}(\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K} \bar{\boldsymbol{x}}^{(k+1)}) \end{aligned}$$

Fixed-point characterization

Consider the following order of iteration

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= \operatorname{prox}_{\tau F}(\boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k)}) \\ \bar{\boldsymbol{x}}^{(k+1)} &= \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \\ \boldsymbol{u}^{(k+1)} &= \operatorname{prox}_{\sigma R^*}(\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K} \bar{\boldsymbol{x}}^{(k+1)}) \end{aligned}$$

The property of proximity operator yields

$$\begin{aligned} \boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k)} - \boldsymbol{x}^{(k+1)} &\in \tau \partial F(\boldsymbol{x}^{(k+1)}) \\ \bar{\boldsymbol{x}}^{(k+1)} &= \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \\ \boldsymbol{u}^{(k)} + \sigma \boldsymbol{K} \bar{\boldsymbol{x}}^{(k+1)} - \boldsymbol{u}^{(k+1)} &\in \sigma \partial R^*(\boldsymbol{u}^{(k+1)}) \end{aligned}$$

Fixed-point characterization

Consider the following order of iteration

$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \operatorname{prox}_{\tau F}(\boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k)}) \\
\bar{\boldsymbol{x}}^{(k+1)} &= \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \\
\boldsymbol{u}^{(k+1)} &= \operatorname{prox}_{\sigma R^*}(\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K} \bar{\boldsymbol{x}}^{(k+1)})
\end{aligned}
$$

The property of proximity operator yields

$$
\begin{aligned}
\boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k)} - \boldsymbol{x}^{(k+1)} &\in \tau \partial F(\boldsymbol{x}^{(k+1)}) \\
\bar{\boldsymbol{x}}^{(k+1)} &= \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \\
\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K} \bar{\boldsymbol{x}}^{(k+1)} - \boldsymbol{u}^{(k+1)} &\in \sigma \partial R^*(\boldsymbol{u}^{(k+1)})
\end{aligned}
$$

which further leads to

$$
\begin{aligned}
\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k+1)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k)} + \tau \boldsymbol{K}^* \boldsymbol{u}^{(k+1)} &\in \tau \partial F(\boldsymbol{x}^{(k+1)}) + \tau \boldsymbol{K}^* \boldsymbol{u}^{(k+1)} \\
\boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k+1)} + \sigma \theta \boldsymbol{K}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) &\in \sigma \partial R^*(\boldsymbol{u}^{(k+1)}) - \sigma \boldsymbol{K} \boldsymbol{x}^{(k+1)}
\end{aligned}
$$

Fixed-point characterization

Consider the following order of iteration

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathrm{prox}_{\tau F}(\mathbf{x}^{(k)} - \tau \mathbf{K}^* \mathbf{u}^{(k)}) \\ \bar{\mathbf{x}}^{(k+1)} &= \mathbf{x}^{(k+1)} + \theta(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) \\ \mathbf{u}^{(k+1)} &= \mathrm{prox}_{\sigma R^*}(\mathbf{u}^{(k)} + \sigma \mathbf{K} \bar{\mathbf{x}}^{(k+1)}) \end{aligned}$$

The property of proximity operator yields

$$\begin{aligned} \mathbf{x}^{(k)} - \tau \mathbf{K}^* \mathbf{u}^{(k)} - \mathbf{x}^{(k+1)} &\in \tau \partial F(\mathbf{x}^{(k+1)}) \\ \bar{\mathbf{x}}^{(k+1)} &= \mathbf{x}^{(k+1)} + \theta(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) \\ \mathbf{u}^{(k)} + \sigma \mathbf{K} \bar{\mathbf{x}}^{(k+1)} - \mathbf{u}^{(k+1)} &\in \sigma \partial R^*(\mathbf{u}^{(k+1)}) \end{aligned}$$

which further leads to

$$\begin{aligned} \frac{\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}}{\tau} - \mathbf{K}^*(\mathbf{u}^{(k)} - \mathbf{u}^{(k+1)}) &\in \partial F(\mathbf{x}^{(k+1)}) + \mathbf{K}^* \mathbf{u}^{(k+1)} \\ \frac{\mathbf{u}^{(k)} - \mathbf{u}^{(k+1)}}{\sigma} - \theta \mathbf{K}(\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}) &\in \partial R^*(\mathbf{u}^{(k+1)}) - \mathbf{K}\mathbf{x}^{(k+1)} \end{aligned}$$

Fixed-point characterization

Consider the following order of iteration

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \mathrm{prox}_{\tau F}(\boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k)}) \\
\bar{\boldsymbol{x}}^{(k+1)} &= \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \\
\boldsymbol{u}^{(k+1)} &= \mathrm{prox}_{\sigma R^*}(\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K} \bar{\boldsymbol{x}}^{(k+1)})
\end{aligned}$$

which further leads to

$$\begin{aligned}
\frac{\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k+1)}}{\tau} - \boldsymbol{K}^*(\boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k+1)}) &\in \partial F(\boldsymbol{x}^{(k+1)}) + \boldsymbol{K}^* \boldsymbol{u}^{(k+1)} \\
\frac{\boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k+1)}}{\sigma} - \theta \boldsymbol{K}(\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k+1)}) &\in \partial R^*(\boldsymbol{u}^{(k+1)}) - \boldsymbol{K} \boldsymbol{x}^{(k+1)}
\end{aligned}$$

Rearrange terms

$$\begin{bmatrix} \boldsymbol{Id}_n/\tau & -\boldsymbol{K}^* \\ -\theta \boldsymbol{K} & \boldsymbol{Id}_m/\sigma \end{bmatrix} \begin{pmatrix} \boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k+1)} \\ \boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k+1)} \end{pmatrix} \in \begin{bmatrix} \partial F & \boldsymbol{K}^* \\ -\boldsymbol{K} & \partial R^* \end{bmatrix} \begin{pmatrix} \boldsymbol{x}^{(k+1)} \\ \boldsymbol{u}^{(k+1)} \end{pmatrix}$$

Fixed-point characterization

Consider the following order of iteration

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= \mathrm{prox}_{\tau F}(\boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k)}) \\ \bar{\boldsymbol{x}}^{(k+1)} &= \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \\ \boldsymbol{u}^{(k+1)} &= \mathrm{prox}_{\sigma R^*}(\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K} \bar{\boldsymbol{x}}^{(k+1)}) \end{aligned}$$

Lastly

$$\begin{pmatrix} \boldsymbol{x}^{(k+1)} \\ \boldsymbol{u}^{(k+1)} \end{pmatrix} = \left(\begin{bmatrix} \boldsymbol{Id}_n/\tau & -\boldsymbol{K}^* \\ -\theta \boldsymbol{K} & \boldsymbol{Id}_m/\sigma \end{bmatrix} + \begin{bmatrix} \partial F & \boldsymbol{K}^* \\ -\boldsymbol{K} & \partial R^* \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{Id}_n/\tau & -\boldsymbol{K}^* \\ -\theta \boldsymbol{K} & \boldsymbol{Id}_m/\sigma \end{bmatrix} \begin{pmatrix} \boldsymbol{x}^{(k)} \\ \boldsymbol{u}^{(k)} \end{pmatrix}$$

Fixed-point characterization

Consider the following order of iteration

$$\begin{aligned}\mathbf{x}^{(k+1)} &= \mathrm{prox}_{\tau F}(\mathbf{x}^{(k)} - \tau \mathbf{K}^* \mathbf{u}^{(k)})\\ \bar{\mathbf{x}}^{(k+1)} &= \mathbf{x}^{(k+1)} + \theta(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})\\ \mathbf{u}^{(k+1)} &= \mathrm{prox}_{\sigma R^*}(\mathbf{u}^{(k)} + \sigma \mathbf{K}\bar{\mathbf{x}}^{(k+1)})\end{aligned}$$

Lastly

$$\begin{pmatrix}\mathbf{x}^{(k+1)}\\ \mathbf{u}^{(k+1)}\end{pmatrix} = \left(\begin{bmatrix}\mathbf{Id}_n/\tau & -\mathbf{K}^*\\ -\theta\mathbf{K} & \mathbf{Id}_m/\sigma\end{bmatrix} + \begin{bmatrix}\partial F & \mathbf{K}^*\\ -\mathbf{K} & \partial R^*\end{bmatrix}\right)^{-1} \begin{bmatrix}\mathbf{Id}_n/\tau & -\mathbf{K}^*\\ -\theta\mathbf{K} & \mathbf{Id}_m/\sigma\end{bmatrix}\begin{pmatrix}\mathbf{x}^{(k)}\\ \mathbf{u}^{(k)}\end{pmatrix}$$

Define the following operator: let $\theta = 1$,

$$\mathbf{z}^{(k)} = \begin{pmatrix}\mathbf{x}^{(k)}\\ \mathbf{u}^{(k)}\end{pmatrix} \in \mathbb{R}^{m+n}, \quad \mathcal{A} = \begin{bmatrix}\partial F & \mathbf{K}^*\\ -\mathbf{K} & \partial R^*\end{bmatrix} \quad \text{and} \quad \mathcal{V} = \begin{bmatrix}\mathbf{Id}_n/\tau & -\mathbf{K}^*\\ -\mathbf{K} & \mathbf{Id}_m/\sigma\end{bmatrix}.$$

Proposition - Properties of $\mathcal{A}$ and $\mathcal{V}$

The following hold

- $\mathcal{A}$ is maximal monotone.
- Let $\sigma\tau\|\boldsymbol{K}\|^2 < 1$, then $\mathcal{V}$ is positive definite.

The simplified characterization

$$\begin{aligned}\boldsymbol{z}^{(k+1)} = (\mathcal{V}+\mathcal{A})^{-1}\mathcal{V}\boldsymbol{z}^{(k)} &= \left(\mathcal{V}(\boldsymbol{Id}+\mathcal{V}^{-1}\mathcal{A})\right)^{-1}\mathcal{V}\boldsymbol{z}^{(k)} \\ &= (\boldsymbol{Id}+\mathcal{V}^{-1}\mathcal{A})^{-1}\boldsymbol{z}^{(k)}\end{aligned}$$

Proposition - Properties of $\mathcal{A}$ and $\mathcal{V}$

The following hold

- $\mathcal{A}$ is maximal monotone.
- Let $\sigma\tau\|\boldsymbol{K}\|^2 < 1$, then $\mathcal{V}$ is positive definite.

The simplified characterization

$$\begin{aligned} \boldsymbol{z}^{(k+1)} = (\mathcal{V} + \mathcal{A})^{-1}\mathcal{V}\boldsymbol{z}^{(k)} &= \left(\mathcal{V}(\boldsymbol{Id} + \mathcal{V}^{-1}\mathcal{A})\right)^{-1}\mathcal{V}\boldsymbol{z}^{(k)} \\ &= (\boldsymbol{Id} + \mathcal{V}^{-1}\mathcal{A})^{-1}\boldsymbol{z}^{(k)} \end{aligned}$$

Proposition - Monotonicity of $\mathcal{V}^{-1}\mathcal{A}$

Let $\mathcal{A} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximally monotone, let $\mathcal{V} \in \mathbb{R}^{n\times n}$ be symmetric and positive definite. Let $\mathbb{R}^n_{\mathcal{V}}$ be the space obtained by endowing $\mathbb{R}^n$ with the scalar product

$$(\boldsymbol{x}, \boldsymbol{y}) \mapsto \langle \boldsymbol{x} \mid \boldsymbol{y}\rangle_{\mathbb{R}^n_{\mathcal{V}}} = \langle \mathcal{V}\boldsymbol{x} \mid \boldsymbol{y}\rangle.$$

Then $\mathcal{V}^{-1}\mathcal{A} : \mathbb{R}^n_{\mathcal{V}} \rightrightarrows \mathbb{R}^n_{\mathcal{V}}$ is maximally monotone.

Primal-Dual splitting as PPA

Proposition - Properties of $\mathcal{A}$ and $\mathcal{V}$

The following hold

- $\mathcal{A}$ is maximal monotone.
- Let $\sigma\tau\|\boldsymbol{K}\|^2 < 1$, then $\mathcal{V}$ is positive definite.

The simplified characterization

$$\begin{aligned} \boldsymbol{z}^{(k+1)} = (\mathcal{V} + \mathcal{A})^{-1}\mathcal{V}\boldsymbol{z}^{(k)} &= \left(\mathcal{V}(\boldsymbol{Id} + \mathcal{V}^{-1}\mathcal{A})\right)^{-1}\mathcal{V}\boldsymbol{z}^{(k)} \\ &= (\boldsymbol{Id} + \mathcal{V}^{-1}\mathcal{A})^{-1}\boldsymbol{z}^{(k)} \end{aligned}$$

Proposition - Primal-Dual splitting as PPA

Under the previous setting, $\mathcal{V}^{-1}\mathcal{A} : \mathbb{R}^{m+n}_{\mathcal{V}} \rightrightarrows \mathbb{R}^{m+n}_{\mathcal{V}}$ is maximally monotone, and the Primal-Dual splitting iteration is equivalent to the proximal point algorithm for solving

$$\text{find} \quad \boldsymbol{z} \in \mathbb{R}^{m+n} \quad \text{such that} \quad \mathbf{0} \in \mathcal{V}^{-1}\mathcal{A}(\boldsymbol{z}).$$

Convergence

Theorem - Convergence with constant step-size

For Primal-Dual splitting method, let $R \in \Gamma_0(\mathbb{R}^m)$, $F \in \Gamma_0(\mathbb{R}^n)$ and $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear. Let $\theta = 1$, and $\sigma, \tau > 0$ be such that

$$\sigma\tau\|\boldsymbol{K}\|^2 < 1.$$

Then $\{(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)})\}_{k\in\mathbb{N}}$ converges to a saddle point $(\boldsymbol{x}^\star, \boldsymbol{u}^\star)$ of $\mathscr{L}(\boldsymbol{x}; \boldsymbol{u})$.

Theorem - Convergence with constant step-size

For Primal-Dual splitting method, let $R \in \Gamma_0(\mathbb{R}^m)$, $F \in \Gamma_0(\mathbb{R}^n)$ and $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear. Let $\theta = 1$, and $\sigma, \tau > 0$ be such that

$$\sigma\tau\|\boldsymbol{K}\|^2 < 1.$$

Then $\{(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)})\}_{k\in\mathbb{N}}$ converges to a saddle point $(\boldsymbol{x}^\star, \boldsymbol{u}^\star)$ of $\mathscr{L}(\boldsymbol{x}; \boldsymbol{u})$.

Theorem - Convergence speed

With the above convergence result,

- Sequence
$$\|\boldsymbol{z}^{(k)} - \boldsymbol{z}^{(k-1)}\| = o\big(1/\sqrt{k}\big).$$
- Duality gap: let
$$\tilde{\boldsymbol{x}}^{(K)} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{x}^{(k)} \quad \text{and} \quad \tilde{\boldsymbol{u}}^{(K)} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{u}^{(k)}.$$
Then $(F + R \circ \boldsymbol{K})(\tilde{\boldsymbol{x}}^{(K)}) - (F + R \circ \boldsymbol{K})(\boldsymbol{x}^\star) = O(1/K)$.

Variants

Strong convexity, smoothness and three terms

Problem - Non-smooth optimization problem

Consider

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \right\},$$

with

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

Extra assumption

- One function is strongly convex, *e.g.* F is α-strongly convex.

Algorithm - Variant 1

initial: $(\boldsymbol{x}^{(0)}, \boldsymbol{u}^{(0)}) \in \mathrm{dom}(F) \times \mathrm{dom}(R)$ and $\bar{\boldsymbol{x}}^{(0)} = \boldsymbol{x}^{(0)}$; $\sigma_0, \tau_0 > 0$ such that

$$\sigma_0 \tau_0 \|\boldsymbol{K}\|^2 \leq 1.$$

repeat:

1. Dual update: $\boldsymbol{u}^{(k+1)} = \mathrm{prox}_{\sigma_k R^*}(\boldsymbol{u}^{(k)} + \sigma_k \boldsymbol{K} \bar{\boldsymbol{x}}^{(k)})$
2. Primal update: $\boldsymbol{x}^{(k+1)} = \mathrm{prox}_{\tau_k F}(\boldsymbol{x}^{(k)} - \tau_k \boldsymbol{K}^* \boldsymbol{u}^{(k+1)})$
3. Paras update: $\theta_k = \frac{1}{\sqrt{1+2\alpha\tau_k}}, \;\; \tau_{k+1} = \theta_k \tau_k, \;\; \sigma_{k+1} = \sigma_k / \theta_k$
4. Extrapolation: $\bar{\boldsymbol{x}}^{(k+1)} = \boldsymbol{x}^{(k+1)} + \theta_k(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$

until: stopping criterion is satisfied.

Extra assumption

- F is α-strongly convex and R^* is δ-strongly convex.

Algorithm - Variant 2

initial: $(\boldsymbol{x}^{(0)}, \boldsymbol{u}^{(0)}) \in \mathrm{dom}(F) \times \mathrm{dom}(R)$ and $\bar{\boldsymbol{x}}^{(0)} = \boldsymbol{x}^{(0)}$; Choose $\mu \leq 2\sqrt{\alpha\delta}/\|\boldsymbol{K}\|$ and

$$\sigma = \frac{\mu}{2\alpha}, \quad \tau = \frac{\mu}{2\delta} \quad \text{and} \quad \theta \in \left[\tfrac{1}{1+\mu}, 1\right].$$

repeat:

1. Dual update: $\boldsymbol{u}^{(k+1)} = \mathrm{prox}_{\sigma R^*}(\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K}\bar{\boldsymbol{x}}^{(k)})$
2. Primal update: $\boldsymbol{x}^{(k+1)} = \mathrm{prox}_{\tau F}(\boldsymbol{x}^{(k)} - \tau \boldsymbol{K}^* \boldsymbol{u}^{(k+1)})$
3. Extrapolation: $\bar{\boldsymbol{x}}^{(k+1)} = \boldsymbol{x}^{(k+1)} + \theta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$

until: stopping criterion is satisfied.

Problem - Non-smooth optimization problem

Consider

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \right\},$$

with

- $F \in C_L^1(\mathbb{R}^n)$ and $R \in \Gamma_0(\mathbb{R}^m)$.
- $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear.

Primal-Dual fixed-point algorithm

Algorithm - Primal-Dual fixed-point algorithm

initial: $(\boldsymbol{x}^{(0)}, \boldsymbol{u}^{(0)}) \in \mathrm{dom}(F) \times \mathrm{dom}(R)$, $0 < \lambda < \|\boldsymbol{K}\|^2$ and $0 < \gamma < 2/L$.

repeat:

1. Forward update: $\boldsymbol{x}^{(k+1/2)} = \boldsymbol{x}^{(k)} - \gamma \nabla F(\boldsymbol{x}^{(k)})$
2. Dual update: $\boldsymbol{u}^{(k+1)} = \left(\boldsymbol{Id} - \mathrm{prox}_{\gamma/\lambda R}\right)\left(\boldsymbol{K}\boldsymbol{x}^{(k+1/2)} + (\boldsymbol{Id} - \lambda \boldsymbol{K}\boldsymbol{K}^*)\boldsymbol{u}^{(k)}\right)$
3. Primal update: $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k+1/2)} - \lambda \boldsymbol{K}^* \boldsymbol{u}^{(k+1)}$

until: stopping criterion is satisfied.

Primal-Dual fixed-point algorithm

Algorithm - Primal-Dual fixed-point algorithm

initial: $(\boldsymbol{x}^{(0)}, \boldsymbol{u}^{(0)}) \in \mathrm{dom}(F) \times \mathrm{dom}(R)$, $0 < \lambda < \|\boldsymbol{K}\|^2$ and $0 < \gamma < 2/L$.

repeat:

1. Forward update: $\boldsymbol{x}^{(k+1/2)} = \boldsymbol{x}^{(k)} - \gamma \nabla F(\boldsymbol{x}^{(k)})$
2. Dual update: $\boldsymbol{u}^{(k+1)} = \left(\boldsymbol{Id} - \mathrm{prox}_{\gamma/\lambda R}\right)\left(\boldsymbol{K}\boldsymbol{x}^{(k+1/2)} + (\boldsymbol{Id} - \lambda \boldsymbol{K}\boldsymbol{K}^*)\boldsymbol{u}^{(k)}\right)$
3. Primal update: $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k+1/2)} - \lambda \boldsymbol{K}^* \boldsymbol{u}^{(k+1)}$

until: stopping criterion is satisfied.

Fixed-point characterization

$$\begin{pmatrix} \boldsymbol{u}^{(k+1)} \\ \boldsymbol{x}^{(k+1)} \end{pmatrix} = \begin{bmatrix} \boldsymbol{Id} & 0 \\ -\lambda \boldsymbol{K}^* & \boldsymbol{Id} \end{bmatrix} \begin{bmatrix} \boldsymbol{Id} - \mathrm{prox}_{\gamma/\lambda R} & 0 \\ 0 & \boldsymbol{Id} \end{bmatrix} \begin{bmatrix} \boldsymbol{Id} - \lambda \boldsymbol{K}\boldsymbol{K}^* & \boldsymbol{K} - \gamma \boldsymbol{K} \nabla F \\ 0 & \boldsymbol{Id} - \gamma \nabla F \end{bmatrix} \begin{pmatrix} \boldsymbol{u}^{(k)} \\ \boldsymbol{x}^{(k)} \end{pmatrix}$$

Primal-Dual splitting for three functions

Problem - Non-smooth optimization problem

Consider

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + G(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \right\},$$

with

- $G \in C_L^1(\mathbb{R}^n)$, $F \in \Gamma_0(\mathbb{R}^n)$ and $R \in \Gamma_0(\mathbb{R}^m)$.
- $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear.

Primal-Dual splitting for three functions

Problem - Non-smooth optimization problem

Consider

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + G(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \right\},$$

with

- $G \in C_L^1(\mathbb{R}^n)$, $F \in \Gamma_0(\mathbb{R}^n)$ and $R \in \Gamma_0(\mathbb{R}^m)$.
- $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear.

Suppose $\mathrm{zer}(\partial F + \nabla G + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K})$ is non-empty, and let $\boldsymbol{x}^\star \in \mathrm{zer}(\partial F + \nabla G + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K})$. Then

$$\boldsymbol{0} \in \partial F(\boldsymbol{x}^\star) + \nabla G(\boldsymbol{x}^\star) + \boldsymbol{K}^* \partial R(\boldsymbol{K}\boldsymbol{x}^\star).$$

Problem - Non-smooth optimization problem

Consider

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \Big\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + G(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \Big\},$$

with

- $G \in C_L^1(\mathbb{R}^n)$, $F \in \Gamma_0(\mathbb{R}^n)$ and $R \in \Gamma_0(\mathbb{R}^m)$.
- $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear.

Suppose $\mathrm{zer}(\partial F + \nabla G + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K})$ is non-empty, and let $\boldsymbol{x}^\star \in \mathrm{zer}(\partial F + \nabla G + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K})$. Then

$$\boldsymbol{0} \in \partial F(\boldsymbol{x}^\star) + \nabla G(\boldsymbol{x}^\star) + \boldsymbol{K}^* \partial R(\boldsymbol{K}\boldsymbol{x}^\star).$$

Since ∂R is set-valued, there exists $\boldsymbol{u}^\star \in \partial R(\boldsymbol{K}\boldsymbol{x}^\star)$ such that

$$\boldsymbol{0} \in \partial F(\boldsymbol{x}^\star) + \nabla G(\boldsymbol{x}^\star) + \boldsymbol{K}^*\boldsymbol{u}^\star \quad \text{and} \quad \boldsymbol{K}\boldsymbol{x}^\star \in (\partial R)^{-1}(\boldsymbol{u}^\star).$$

Problem - Non-smooth optimization problem

Consider

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \Big\{ \Phi(\boldsymbol{x}) \stackrel{\text{def}}{=} F(\boldsymbol{x}) + G(\boldsymbol{x}) + R(\boldsymbol{K}\boldsymbol{x}) \Big\},$$

with

- $G \in C^1_L(\mathbb{R}^n)$, $F \in \Gamma_0(\mathbb{R}^n)$ and $R \in \Gamma_0(\mathbb{R}^m)$.
- $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ be bounded linear.

Suppose $\mathrm{zer}(\partial F + \nabla G + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K})$ is non-empty, and let $\boldsymbol{x}^\star \in \mathrm{zer}(\partial F + \nabla G + \boldsymbol{K}^* \circ \partial R \circ \boldsymbol{K})$. Then

$$\mathbf{0} \in \partial F(\boldsymbol{x}^\star) + \nabla G(\boldsymbol{x}^\star) + \boldsymbol{K}^* \partial R(\boldsymbol{K}\boldsymbol{x}^\star).$$

$$\text{find} \quad \boldsymbol{u}^\star \in \mathbb{R}^m \quad \text{such that} \quad \exists \boldsymbol{x}^\star \in \mathbb{R}^n \begin{cases} \mathbf{0} \in \partial F(\boldsymbol{x}^\star) + \nabla G(\boldsymbol{x}^\star) + \boldsymbol{K}^* \boldsymbol{u}^\star, \\ \mathbf{0} \in \partial R^*(\boldsymbol{u}^\star) - \boldsymbol{K}\boldsymbol{x}^\star. \end{cases}$$

Primal-Dual splitting for three functions

Algorithm - A Primal-Dual splitting method

initial: $(\boldsymbol{x}^{(0)}, \boldsymbol{u}^{(0)}) \in \mathrm{dom}(F+G) \times \mathrm{dom}(R)$; $\sigma, \tau > 0$ and $\delta = \frac{L}{2}(\frac{1}{\tau} - \sigma\|\boldsymbol{K}\|^2)^{-1}$ such that $\rho_k \in [0, \delta]$

$$\frac{1}{\tau} - \sigma\|\boldsymbol{K}\|^2 \geq \frac{L}{2} \quad \text{and} \quad \sum_{k\in\mathbb{N}} \rho_k(\delta - \rho_k) = +\infty.$$

repeat:

1. Dual update: $\bar{\boldsymbol{u}}^{(k+1)} = \mathrm{prox}_{\sigma R^*}(\boldsymbol{u}^{(k)} + \sigma \boldsymbol{K}\boldsymbol{x}^{(k)})$.
2. Primal update: $\bar{\boldsymbol{x}}^{(k+1)} = \mathrm{prox}_{\tau F}\left(\boldsymbol{x}^{(k)} - \tau\nabla G(\boldsymbol{x}^{(k)}) + \tau\boldsymbol{K}^*(2\bar{\boldsymbol{u}}^{(k+1)} - \boldsymbol{u}^{(k)})\right)$
3. Extrapolation: $(\boldsymbol{x}^{(k+1)}, \boldsymbol{u}^{(k+1)}) = \rho_k(\bar{\boldsymbol{x}}^{(k+1)}, \bar{\boldsymbol{u}}^{(k+1)}) + (1-\rho_k)(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)})$

until: stopping criterion is satisfied.

References

- H. Bauschke and P. L. Combettes. "Convex Analysis and Monotone Operator Theory in Hilbert Spaces". Springer, 2011.
- A. Chambolle and T. Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging". Journal of mathematical imaging and vision 40.1 (2011): 120-145.
- P. L. Combettes and J.-C. Pesquet. "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators". Set-Valued and variational analysis 20.2 (2012): 307-330.
- Peijun Chen, Jianguo Huang, and Xiaoqun Zhang. "A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration." Inverse Problems 29.2 (2013): 025011.
- Laurent Condat. "A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms." Journal of optimization theory and applications 158.2 (2013): 460-479.
- Bang Công Vũ. "A splitting algorithm for dual monotone inclusions involving cocoercive operators." Advances in Computational Mathematics 38.3 (2013): 667-681.