Introductory Course on Non-smooth Optimisation

Lecture 07 - Other operator splitting methods

Jingwei Liang

Department of Applied Mathematics and Theoretical Physics

Table of contents

Problem

$$\text{Find } x \in \mathbb{R}^n \text{ such that } 0 \in A(x) + B(x) + C(x).$$

Assumptions

- $A, B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are maximal monotone.
- $C : \mathbb{R}^n \to \mathbb{R}^n$ is β-cocoercive.
- $\text{zer}(A + B + C) \neq \emptyset$.

Solution characterisation

- given $x^\star \in \mathrm{zer}(A + B + C)$, there exists $z^\star \in \mathbb{R}^n$ such that

$$\begin{cases} x^\star - z^\star \in \gamma A(x^\star) + \gamma C(x^\star) \\ z^\star - x^\star \in \gamma B(x^\star) \end{cases} \implies \begin{cases} 2x^\star - z^\star - \gamma C(x^\star) \in x^\star + \gamma A(x^\star), \\ z^\star \in x^\star + \gamma B(x^\star). \end{cases}$$

- apply the resolvent

$$\begin{cases} x^\star = \mathcal{J}_{\gamma A}(2x^\star - z^\star - \gamma C(x^\star)), \\ x^\star = \mathcal{J}_{\gamma B}(z^\star). \end{cases}$$

- equivalent formulation

$$\begin{cases} z^\star = z^\star + \mathcal{J}_{\gamma A}(2x^\star - z^\star - \gamma C(x^\star)) - x^\star, \\ x^\star = \mathcal{J}_{\gamma B}(z^\star). \end{cases}$$

- fixed-point iteration

$$\begin{cases} z_{k+1} = z_k + (\mathcal{J}_{\gamma A}(2x_k - z_k - \gamma C(x_k)) - x_k), \\ x_{k+1} = \mathcal{J}_{\gamma B}(z_{k+1}). \end{cases}$$

Three-operator splitting

Let $z_0 \in \mathbb{R}^n$, $\gamma \in]0, 2\beta[$ and $x_0 = \mathcal{J}_{\gamma B}(z_0)$, $\lambda \in]0, \frac{4\beta - \gamma}{2\beta}[$:

$$u_{k+1} = \mathcal{J}_{\gamma A}\big(2x_k - z_k - \gamma C(x_k)\big),$$
$$z_{k+1} = (1 - \lambda)z_k + \lambda(z_k + u_{k+1} - x_k),$$
$$x_{k+1} = \mathcal{J}_{\gamma B}(z_{k+1}).$$

- Recovers Douglas–Rachford when $C = 0$.
- Recovers Forward–Backward when $B = 0$.

Fixed-point formulation

- $u_{k+1} = \mathcal{J}_{\gamma A}\big(2x_k - z_k - \gamma C(x_k)\big) = \mathcal{J}_{\gamma A} \circ (2\mathcal{J}_{\gamma B} - \mathsf{Id} - \gamma C \circ \mathcal{J}_{\gamma B})(z_k)$.

- For z_k,
$$z_{k+1} = (1 - \lambda)z_k + \lambda(z_k + u_{k+1} - x_k)$$
$$= (1 - \lambda)z_k + \lambda\big(\mathsf{Id} - \mathcal{J}_{\gamma B} + \mathcal{J}_{\gamma A} \circ (2\mathcal{J}_{\gamma B} - \mathsf{Id} - \gamma C \circ \mathcal{J}_{\gamma B})\big)(z_k).$$

Property

- $\mathcal{T}_{\mathsf{TOS}} \stackrel{\text{def}}{=} \mathsf{Id} - \mathcal{J}_{\gamma B} + \mathcal{J}_{\gamma A} \circ (2\mathcal{J}_{\gamma B} - \mathsf{Id} - \gamma C \circ \mathcal{J}_{\gamma B})$ is $\frac{2\beta}{4\beta - \gamma}$-averaged.

Problem

$$\text{Find } x \in \mathbb{R}^n \text{ such that } 0 \in A(x) + \mathcal{N}_V(x) + C(x).$$

Assumptions

- $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are maximal monotone.
- $V \subseteq \mathbb{R}^n$ is a closed subspace.
- $C : \mathbb{R}^n \to \mathbb{R}^n$ is β-cocoercive.
- $\mathrm{zer}(A + \mathcal{N}_V + C) \neq \emptyset$.

Forward–Douglas–Rachford splitting

Forward–Douglas–Rachford splitting

Let $z_0 \in \mathbb{R}^n$, $\gamma \in]0, 2\beta[$ and $x_0 = \mathcal{J}_{\gamma B}(z_0)$, $\lambda \in]0, \frac{4\beta - \gamma}{2\beta}[$:

$$u_{k+1} = \mathcal{J}_{\gamma A}\big(2x_k - z_k - \gamma \mathcal{P}_V \circ C(x_k)\big),$$
$$z_{k+1} = (1 - \lambda)z_k + \lambda(z_k + u_{k+1} - x_k),$$
$$x_{k+1} = \mathcal{P}_V(z_{k+1}).$$

- FDR was proposed before TOS.
- Recovers Douglas–Rachford when $C = 0$.
- Recovers Forward–Backward when $V = \mathbb{R}^n$.

Fixed-point formulation Denote $C_V = \mathcal{P}_V \circ C \circ \mathcal{P}_V$,

- For u_{k+1}: $\mathcal{R}_V \circ C_V = (2\mathcal{P}_V - \mathrm{Id})C_V = C_V$

$$u_{k+1} = \mathcal{J}_{\gamma A} \circ (2\mathcal{P}_V - \mathrm{Id} - \gamma C_V)(z_k)$$
$$= \mathcal{J}_{\gamma A} \circ \mathcal{R}_V \circ (\mathrm{Id} - \gamma C_V)(z_k).$$

- For z_k,

$$z_{k+1} = (1-\lambda)z_k + \lambda(z_k + u_{k+1} - x_k)$$
$$= (1-\lambda)z_k + \lambda\big(\mathcal{J}_{\gamma A} \circ \mathcal{R}_V \circ (\mathrm{Id} - \gamma C_V) + \mathrm{Id} - \mathcal{P}_V\big)(z_k)$$
$$= (1-\lambda)z_k + \lambda\frac{1}{2}(\mathrm{Id} + \mathcal{R}_{\gamma R}\mathcal{R}_V)(\mathrm{Id} - \gamma C_V)(z_k).$$

Identify: $\mathrm{Id} - \mathcal{P}_V = \frac{1}{2}(\mathrm{Id} - \gamma C_V) - \mathcal{P}_V(\mathrm{Id} - \gamma C_V) + \frac{1}{2}(\mathrm{Id} - \gamma C_V)$.

Property

- $\mathcal{T}_{\mathrm{FDR}} \overset{\mathrm{def}}{=} \frac{1}{2}(\mathrm{Id} + \mathcal{R}_{\gamma R}\mathcal{R}_V)(\mathrm{Id} - \gamma C_V)$ is $\frac{2\beta}{4\beta-\gamma}$-averaged.

Outline

Problem $r \geq 2$

$$\text{Find } x \in \mathbb{R}^n \text{ such that } 0 \in \sum_{i=1}^{r} A_i(x) + B(x).$$

Assumptions

- for each i, $A_i : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone.
- $B : \mathbb{R}^n \to \mathbb{R}^n$ is β-cocoercive.
- $\text{zer}(\sum_i A + B) \neq \emptyset$.

Generalised Forward–Backward splitting

Let $(\omega_i)_i \in]0, 1[^r$ s.t. $\sum_i \omega_i = 1$, $\gamma \in]0, 2\beta[$, $\lambda \in]0, \frac{4\beta - \gamma}{2\beta}[$. $z_{i,0} \in \mathbb{R}^n$ and $x_0 = \sum_i \omega_i z_{i,0}$:

$$\text{For } i \in \{1, \cdots, r\}$$

$$\left\lfloor \begin{array}{l} u_{i,k+1} = \mathcal{J}_{\frac{\gamma}{\omega_i} A_i}(2x_k - z_{i,k} - \gamma B(x_k)) \\ z_{i,k+1} = (1 - \lambda)z_{i,k} + \lambda(z_{i,k} + u_{i,k+1} - x_k), \end{array} \right.$$

$$x_{k+1} = \sum_i \omega_i z_{i,k+1}.$$

- Earliest of the three methods.
- Recovers Douglas–Rachford in product space when $B = 0$.
- Recovers Forward–Backward when $r = 1$.

Product space

- Let $\mathcal{H} = \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ be the product space endowed with the scalar product and norm defined by
$$\forall x, x' \in \mathcal{H}, \ \langle x, x' \rangle = \sum_{i=1}^{r} \omega_i \langle x_i, x_i' \rangle, \ \ \|x\| = \sqrt{\sum_{i=1}^{r} \omega_i \|x_i\|^2}.$$

- Let $\mathcal{S} = \{x = (x_i)_i \in \mathcal{H} | x_1 = \cdots = x_r\}$ and $\mathcal{S}^\perp = \left\{ x = (x_i)_i \in \mathcal{H} | \sum_{i=1}^{r} \omega_i x_i = 0 \right\}$. Define the canonical isometry $C : \mathcal{H} \to \mathcal{S}, \ x \mapsto (x, \cdots, x)$, then
$$\mathcal{P}_{\mathcal{S}}(z) \stackrel{\text{def}}{=} C\left(\sum_{i=1}^{r} \omega_i z_i \right), \ \ \forall z \in \mathcal{H}.$$

- Let $\gamma = (\gamma_i)_i \in \,]0, +\infty[^r$. For $A_i, i = 1, ..., r$, define
$$\gamma A : \mathcal{H} \rightrightarrows \mathcal{H}, x = (x_i)_i \mapsto \underset{i=1}{\overset{r}{\times}} \gamma_i A_i(x_i).$$

  For B, define
$$B : \mathcal{H} \to \mathcal{H}, \ x = (x_i)_i \mapsto \big(B(x_i) \big)_i.$$

- Define $B_{\mathcal{S}} = B \circ \mathcal{P}_{\mathcal{S}}$ and $\mathcal{J}_{\gamma A} = (\mathcal{J}_{\gamma_i A_i})_i$.

Fixed-point formulation

- For u_{k+1},
$$u_{k+1} = \mathcal{J}_{\gamma A}\big(2\mathcal{P}_{\mathcal{S}}(z_k) - z_k - \gamma B_{\mathcal{S}}(z_k)\big)$$
$$= \mathcal{J}_{\gamma A} \circ \mathcal{R}_{\mathcal{S}} \circ (\mathsf{Id} - \gamma B_{\mathcal{S}})(z_k).$$

- Identify: $\mathsf{Id} - \mathcal{P}_{\mathcal{S}} = \frac{1}{2}(\mathsf{Id} - \gamma B_{\mathcal{S}}) - \mathcal{P}_{\mathcal{S}} \circ (\mathsf{Id} - \gamma B_{\mathcal{S}}) + \frac{1}{2}(\mathsf{Id} - \gamma B_{\mathcal{S}})$.

- For z_k,
$$z_{k+1} = (1 - \lambda)z_k + \big(z_k + \mathcal{J}_{\gamma A}(2\mathcal{P}_{\mathcal{S}}(z_k) - z_k - \gamma B_{\mathcal{S}}(z_k)) - \mathcal{P}_{\mathcal{S}}(z_k)\big)$$
$$= (1 - \lambda)z_k + \big(\mathcal{J}_{\gamma A} \circ \mathcal{R}_{\mathcal{S}} \circ (\mathsf{Id} - \gamma B_{\mathcal{S}})(z_k) + (\mathsf{Id} - \mathcal{P}_{\mathcal{S}})(z_k)\big)$$
$$= (1 - \lambda)z_k + \lambda\frac{1}{2}(\mathsf{Id} + \mathcal{R}_{\gamma A}\mathcal{R}_{\mathcal{S}}) \circ (\mathsf{Id} - \gamma B_{\mathcal{S}})(z_k).$$

Property

- $\mathcal{T}_{\mathsf{GFB}} \overset{\mathsf{def}}{=} \frac{1}{2}(\mathsf{Id} + \mathcal{R}_{\gamma A}\mathcal{R}_{\mathcal{S}}) \circ (\mathsf{Id} - \gamma B_{\mathcal{S}})$ is $\frac{2\beta}{4\beta - \gamma}$-averaged.

Remarks

- Structure and splitting are the key to design first-order methods.

- Convergence analysis via Krasnosel'skiĭ-Mann iteration.

- Most common structure for Krasnosel'skiĭ-Mann operator: PPA and FB.

- Acceleration in general is difficult.

Reference

- D. Davis and W. Yin." A three-operator splitting scheme and its optimization applications". Set-Valued and Variational Analysis, 2017.

- L. M. Briceño-Arias. "Forward-douglas–rachford splitting and forward-partial inverse method for solving monotone inclusions". Optimization, 64(5):1239–1261, 2015.

- H. Raguet, M. J. Fadili, and G. Peyré. "Generalized forward-backward splitting". SIAM Journal on Imaging Sciences, 6(3):1199–1226, 2013.