Introductory Course on Non-smooth Optimisation

Lecture 04 - Backward–Backward splitting

Jingwei Liang

Department of Applied Mathematics and Theoretical Physics

Table of contents

Problem

Let $B : \mathbb{R}^n \to \mathbb{R}^n$ be β-cocoercive for some $\beta > 0$, $s > 1$ be a positive integer, such that for each $i \in \{1, ..., s\}$: $A_i : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone. Consider the problem

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

- A_i can be composed with linear mapping, *e.g.* $L^* \circ A \circ L$.
- Even if the resolvents of B and each A_i are simple, the resolvent of $B + \sum_i A_i$ in most cases is not solvable.
- Use the properties of operators and structure of problem to derive operator splitting schemes.

Outline

Monotone inclusion

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

Assumptions

- $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone.
- $B : \mathbb{R}^n \to \mathbb{R}^n$ is β-cocoersive.
- $\mathrm{zer}(A+B) \neq \emptyset$.

Characterisation of minimiser: $\gamma > 0$

$$x^\star - \gamma B(x^\star) \in x^\star + \gamma A(x^\star) \quad \Leftrightarrow \quad x^\star = \mathcal{J}_{\gamma A} \circ (\mathrm{Id} - \gamma B)(x^\star).$$

Example Let $R \in \Gamma_0$ and $F \in C_L^1$,

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

Fixed-point operator: $\gamma \in]0, 2\beta[$

$$\mathcal{T}_{\mathrm{FB}} = \mathcal{J}_{\gamma A} \circ (\mathrm{Id} - \gamma B).$$

- $\mathcal{J}_{\gamma A}$ is firmly non-expansive.
- $\mathrm{Id} - \gamma B$ is $\frac{\gamma}{2\beta}$-averaged non-expansive.
- $\mathcal{T}_{\mathrm{FB}}$ is $\frac{2\beta}{4\beta-\gamma}$-averaged non-expansive.
- $\mathrm{fix}(\mathcal{T}_{\mathrm{FB}}) = \mathrm{zer}(A+B)$.

Forward–Backward splitting

Let $\gamma \in]0, 2\beta[,\ \lambda_k \in [0, \frac{4\beta-\gamma}{2\beta}]$:

$$x_{k+1} = (1-\lambda_k)x_k + \lambda_k \mathcal{T}_{\mathrm{FB}}(x_k).$$

- Special case of Krasnosel'skiĭ-Mann iteration.
- Recovers proximal point algorithm when $B = 0$.

Outline

Let $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n$ be closed convex and non-empty, such that $\mathcal{X} \cap \mathcal{Y} \neq \emptyset$

$$\min_{x \in \mathbb{R}^n} \ \iota_{\mathcal{X}}(x) + \iota_{\mathcal{Y}}(x).$$

Method of alternating projection (MAP)

Let $x_0 \in \mathcal{X}$:

$$y_{k+1} = \mathcal{P}_{\mathcal{Y}}(x_k),$$
$$x_{k+1} = \mathcal{P}_{\mathcal{X}}(y_{k+1}).$$

Fixed-point operator: $x_{k+1} = \mathcal{T}_{\mathrm{MAP}}(x_k)$,

$$\mathcal{T}_{\mathrm{MAP}} \stackrel{\mathrm{def}}{=} \mathcal{P}_{\mathcal{X}} \circ \mathcal{P}_{\mathcal{Y}}.$$

- $\mathcal{P}_{\mathcal{X}}, \mathcal{P}_{\mathcal{Y}}$ are firmly non-expansive.
- $\mathcal{T}_{\mathrm{MAP}}$ is $\frac{2}{3}$-averaged non-expansive.
- $\mathrm{fix}(\mathcal{T}_{\mathrm{MAP}}) = \mathcal{X} \cap \mathcal{Y}$.

Derive MAP

- Feasibility problem is equivalent to

$$\min_{x,y\in\mathbb{R}^n} \iota_{\mathcal{X}}(x) + \frac{1}{2}\|x-y\|^2 + \iota_{\mathcal{Y}}(y).$$

- Optimality condition

$$\begin{aligned} 0 &\in \mathcal{N}_{\mathcal{Y}}(y^\star) + y^\star - x^\star, \\ 0 &\in \mathcal{N}_{\mathcal{X}}(x^\star) + x^\star - y^\star. \end{aligned}$$

- Fixed-point characterisation

$$\begin{aligned} y^\star &= \mathcal{P}_{\mathcal{Y}}(x^\star), \\ x^\star &= \mathcal{P}_{\mathcal{X}}(y^\star). \end{aligned}$$

- Fixed-point iteration

$$\begin{aligned} y_{k+1} &= \mathcal{P}_{\mathcal{Y}}(x_k), \\ x_{k+1} &= \mathcal{P}_{\mathcal{X}}(y_{k+1}). \end{aligned}$$

Example: SDP feasibility

SDP feasibility

Find $X \in \mathcal{S}^n$ such that

$$X \succeq 0 \quad \text{and} \quad \mathrm{Tr}(A_i X) = b_i,\ i = 1, ..., m.$$

Two sets and projection:

- $\mathcal{X} = \mathcal{S}^n_+$ is the positive semidefinite cone. Let $Y_k = \sum_{i=1}^n \sigma_i u_i u_i^T$ be the eigenvalue decomposition of Y_k, then

$$\mathcal{P}_{\mathcal{X}}(Y_k) = \textstyle\sum_{i=1}^n \max\{0, \sigma_i\} u_i u_i^T.$$

- $\mathcal{Y}$ is the affine set in $\mathcal{S}^n$ define by the linear inequalities,

$$\mathcal{P}_{\mathcal{Y}}(X_k) = X_k - \textstyle\sum_{i=1}^m u_i A_i,$$

  where u_i are found from the normal equations

$$Gu = \big(\mathrm{Tr}(A_i X_k) - b_i, \cdots, \mathrm{Tr}(A_i X_k) - b_m\big),\ \ G_{i,j} = \mathrm{Tr}(A_i A_j).$$

Let $\mathcal{X}, \mathcal{Y}$ be two subspaces, and assume

$$1 \leq p \stackrel{\text{def}}{=} \dim(\mathcal{X}) \leq q \stackrel{\text{def}}{=} \dim(\mathcal{Y}) \leq n-1.$$

Principal angles The principal angles $\theta_k \in [0, \frac{\pi}{2}], k = 1, \ldots, p$ between $\mathcal{X}$ and $\mathcal{Y}$ are defined by, with $u_0 = v_0 \stackrel{\text{def}}{=} 0$, and

$$\begin{aligned} \cos(\theta_k) \stackrel{\text{def}}{=} \langle u_k, v_k \rangle = \max \langle u, v \rangle \quad \text{s.t.} \quad & u \in \mathcal{X}, v \in \mathcal{Y}, \|u\| = 1, \|v\| = 1, \\ & \langle u, u_i \rangle = \langle v, v_i \rangle = 0, i = 0, \cdots, k-1. \end{aligned}$$

Friedrichs angle The Friedrichs angle $\theta_F \in]0, \frac{\pi}{2}]$ between $\mathcal{X}$ and $\mathcal{Y}$ is

$$\begin{aligned} \cos\big(\theta_F(\mathcal{X}, \mathcal{Y})\big) \stackrel{\text{def}}{=} \max \langle u, v \rangle \quad \text{s.t.} \quad & u \in \mathcal{X} \cap (\mathcal{X} \cap \mathcal{Y})^\perp, \|u\| = 1, \\ & v \in \mathcal{Y} \cap (\mathcal{X} \cap \mathcal{Y})^\perp, \|v\| = 1. \end{aligned}$$

Lemma

The Friedrichs angle is θ_{d+1} where $d \stackrel{\text{def}}{=} \dim(\mathcal{X} \cap \mathcal{Y})$. Moreover,

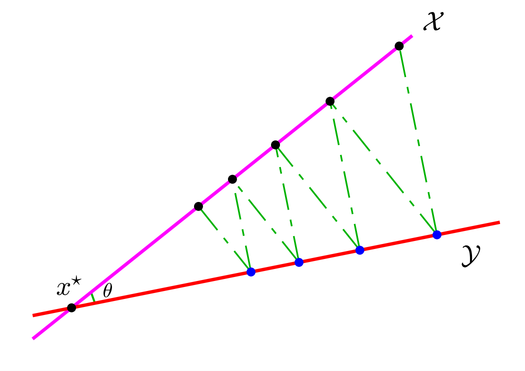
$$\theta_F(\mathcal{X}, \mathcal{Y}) > 0.$$

Example $\mathcal{X}, \mathcal{Y}$ are defined by

$$\mathcal{X} = \{x : Ax = 0\},\ \mathcal{Y} = \{x : Bx = 0\}.$$

Projection onto subspace

$$\mathcal{P}_{\mathcal{X}}(x) = x - A^T(AA^T)^{-1}Ax.$$



- Define diagonal matrices

$$c = \mathrm{diag}\big(\cos(\theta_1), \cdots, \cos(\theta_p)\big),$$
$$s = \mathrm{diag}\big(\sin(\theta_1), \cdots, \sin(\theta_p)\big).$$

- Suppose $p + q < n$, then there exists orthogonal matrix U such that

$$\mathcal{P}_{\mathcal{X}} = U \left[\begin{array}{cc|cc} \mathrm{Id}_p & 0 & 0 & 0 \\ 0 & 0_p & 0 & 0 \\ \hline 0 & 0 & 0_{q-p} & 0 \\ 0 & 0 & 0 & 0_{n-p-q} \end{array}\right] U^*,$$

$$\mathcal{P}_{\mathcal{Y}} = U \left[\begin{array}{cc|cc} c^2 & cs & 0 & 0 \\ cs & c^2 & 0 & 0 \\ \hline 0 & 0 & \mathrm{Id}_{q-p} & 0 \\ 0 & 0 & 0 & 0_{n-p-q} \end{array}\right] U^*.$$

- Fixed-point operator

$$\begin{aligned}\mathcal{T}_{\mathrm{MAP}} &= \mathcal{P}_{\mathcal{X}} \circ \mathcal{P}_{\mathcal{Y}} \\ &= U \left[\begin{array}{cc|cc} c^2 & cs & 0 & 0 \\ 0 & 0_p & 0 & 0 \\ \hline 0 & 0 & 0_{q-p} & 0 \\ 0 & 0 & 0 & 0_{n-p-q} \end{array}\right] U^*.\end{aligned}$$

- Consider relaxation

$$\begin{aligned}\mathcal{T}_{\mathrm{MAP}}^{\lambda} &= (1-\lambda)\mathrm{Id} + \lambda \mathcal{T}_{\mathrm{MAP}} \\ &= U \left[\begin{array}{cc|c} (1-\lambda)\mathrm{Id}_p + \lambda c^2 & \lambda cs & 0 \\ 0 & (1-\lambda)\mathrm{Id}_p & 0 \\ \hline 0 & 0 & (1-\lambda)\mathrm{Id}_{n-2p} \end{array}\right] U^*.\end{aligned}$$

Convergence rate

- Eigenvalues

$$\sigma(\mathcal{T}_{\mathrm{MAP}}^{\lambda}) = \big\{1 - \lambda \sin^2(\theta_i) | i = 1, ..., p\big\} \cup \{1 - \lambda\}.$$

- Spectral radius

$$\rho(\mathcal{T}_{\mathrm{MAP}}^{\lambda}) = \max\big\{1 - \lambda \sin^2(\theta_F), |1 - \lambda|\big\}.$$

- No relaxation

$$\rho(\mathcal{T}_{\mathrm{MAP}}) = \cos^2(\theta_F).$$

- Convergence rate, $C > 0$ is some constant

$$\begin{aligned}\|x_k - x^\star\| &= \|\mathcal{T}_{\mathrm{MAP}} x_{k-1} - \mathcal{T}_{\mathrm{MAP}} x^\star\| \\ &= ... \\ &= \|\mathcal{T}_{\mathrm{MAP}}^k (x_0 - x^\star)\| \\ &\leq C \|\mathcal{T}_{\mathrm{MAP}}\|^k \|x_0 - x^\star\|.\end{aligned}$$

Outline

When $\mathcal{X} \cap \mathcal{Y} \neq \emptyset$, MAP returns $x_k, y_k \to x^\star \in \mathcal{X} \cap \mathcal{Y}$.

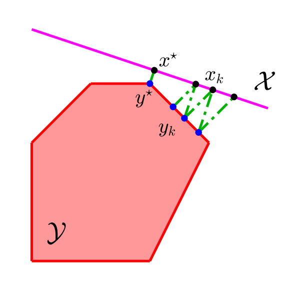
Best pair problem

Let $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n$ be closed and convex, such that

$$\mathcal{X} \cap \mathcal{Y} = \emptyset.$$

Consider finding two points in $\mathcal{X}$ and $\mathcal{Y}$ such that they are the closest, that is

$$\min_{x \in \mathcal{X}, y \in \mathcal{Y}} \|x - y\|.$$



- MAP can be applied and

$$(x_k, y_k) \to (x^\star, y^\star)$$

where $(x^\star, y^\star)$ is a best pair.

Consider

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

- $A, B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are maximal monotone.
- The set of solition is non-empty.

There exists $x^\star, y^\star \in \mathbb{R}^n$ and $\gamma > 0$ such that

$$y^\star - x^\star \in \gamma A(x^\star),$$
$$x^\star - y^\star \in \gamma B(y^\star).$$

Backward–Backward splitting

Let $x_0 \in \mathbb{R}^n$, $\gamma > 0$:

$$y_{k+1} = \mathcal{J}_{\gamma B}(x_k),$$
$$x_{k+1} = \mathcal{J}_{\gamma A}(y_{k+1}).$$

- **Yosida approximation**

$$ {}^{\gamma}A = \frac{1}{\gamma}(\mathrm{Id} - \mathcal{J}_{\gamma A}). $$

which is γ-cocoercive.

- **Regularised monotone inclusion**

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

- **Forward–Backward splitting** $\tau \in]0, 2\gamma]$

$$ x_{k+1} = \mathcal{J}_{\tau A} \circ (\mathrm{Id} - \tau {}^{\gamma}B)(x_k). $$

- **BB as special case of FB** let $\tau = \gamma$

$$ \begin{aligned} x_{k+1} &= \mathcal{J}_{\gamma A} \circ (\mathrm{Id} - \gamma {}^{\gamma}B)(x_k) \\ &= \mathcal{J}_{\gamma A} \circ \big(\mathrm{Id} - \gamma \tfrac{1}{\gamma}(\mathrm{Id} - \mathcal{J}_{\gamma B})\big)(x_k) \\ &= \mathcal{J}_{\gamma A} \circ \mathcal{J}_{\gamma B}(x_k). \end{aligned} $$

An inertial Backward–Backward splitting

Initial : $x_0 \in \mathbb{R}^n, x_{-1} = x_0$ and $\gamma > 0,\ \tau \in]0, 2\gamma]$;

$$y_k = x_k + a_{0,k}(x_k - x_{k-1}) + a_{1,k}(x_{k-1} - x_{k-2}) + \cdots\ ,$$
$$x_{k+1} = \mathcal{J}_{\gamma A} \circ \mathcal{J}_{\gamma B}(y_k),\ \lambda_k \in [0, 1].$$

An inertial BB splitting based on Yosida approximation

Initial : $x_0 \in \mathbb{R}^n, x_{-1} = x_0$ and $\gamma > 0$;

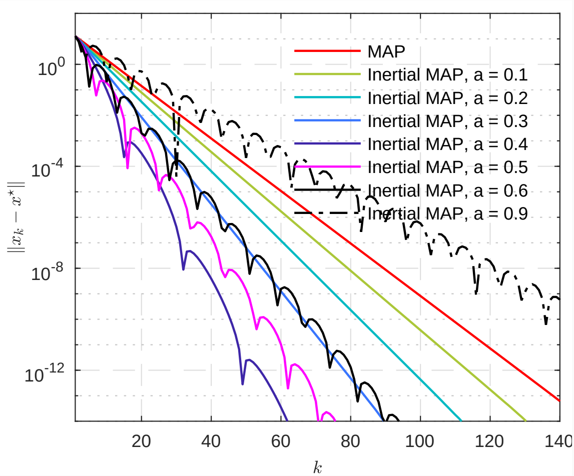
$$y_k = x_k + a_{0,k}(x_k - x_{k-1}) + a_{1,k}(x_{k-1} - x_{k-2}) + \cdots\ ,$$
$$z_k = x_k + b_{0,k}(x_k - x_{k-1}) + b_{1,k}(x_{k-1} - x_{k-2}) + \cdots\ ,$$
$$x_{k+1} = \mathcal{J}_{\tau A} \circ \big(y_k - \tau {}^{\gamma}B(z_k)\big),\ \lambda_k \in [0, 1].$$

Outline

Numerical experiment

Feasibility problem for two subspaces:

$$a = [-4/5, 1] \quad \text{and} \quad b = [-1/5, 1]$$



Reference

- S. Boyd. "Alternating projection", lecture notes.
- H. H. Bauschke, J. Y. Bello Cruz, T. T. A. Nghia, H. M. Pha, and X. Wang. "Optimal rates of linear convergence of relaxed alternating projections and generalized Douglas–Rachford methods for two subspaces". Numerical Algorithms, 73(1):33–76, 2016.
- P. L. Lions and B. Mercier. "Splitting algorithms for the sum of two nonlinear operators". SIAM Journal on Numerical Analysis, 16(6):964–979, 1979.