

# Introductory Course on Non-smooth Optimisation

## Lecture 04 - Backward-Backward splitting

---

Jingwei Liang

Department of Applied Mathematics and Theoretical Physics

- 1 Problem
- 2 Forward-Backward splitting revisit
- 3 MAP continue
- 4 Backward-Backward splitting
- 5 Numerical experiments

## Problem

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

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

- $A_i$  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.

- 1 Problem
- 2 Forward-Backward splitting revisit**
- 3 MAP continue
- 4 Backward-Backward splitting
- 5 Numerical experiments

## Monotone inclusion

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

### Assumptions

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

Characterisation of minimiser:  $\gamma > 0$

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

**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:  $\gamma \in ]0, 2\beta[$

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

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

## Forward-Backward splitting

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

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

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

- 1 Problem
- 2 Forward-Backward splitting revisit
- 3 MAP continue**
- 4 Backward-Backward splitting
- 5 Numerical experiments

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  $x_0 \in \mathcal{X}$ :

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

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

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

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



- 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

$$0 \in \mathcal{N}_{\mathcal{Y}}(y^*) + y^* - x^*,$$

$$0 \in \mathcal{N}_{\mathcal{X}}(x^*) + x^* - y^*.$$

- Fixed-point characterisation

$$y^* = \mathcal{P}_{\mathcal{Y}}(x^*),$$

$$x^* = \mathcal{P}_{\mathcal{X}}(y^*).$$

- Fixed-point iteration

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

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

### SDP feasibility

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

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

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

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

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

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

where  $u_i$  are found from the normal equations

$$Gu = (\text{Tr}(A_1 X_k) - b_1, \dots, \text{Tr}(A_m X_k) - b_m), \quad G_{i,j} = \text{Tr}(A_i A_j).$$

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

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

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

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

## Lemma

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

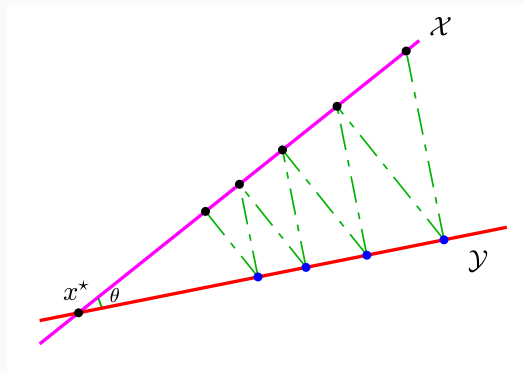
$$\theta_F(\mathcal{X}, \mathcal{Y}) > 0.$$

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

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

Projection onto subspace

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



- Define diagonal matrices

$$c = \text{diag}(\cos(\theta_1), \dots, \cos(\theta_p)),$$

$$s = \text{diag}(\sin(\theta_1), \dots, \sin(\theta_p)).$$

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

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

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

- Fixed-point operator

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

- Consider relaxation

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

- Eigenvalues

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

- Spectral radius

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

- No relaxation

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

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

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

- 1 Problem
- 2 Forward-Backward splitting revisit
- 3 MAP continue
- 4 Backward-Backward splitting**
- 5 Numerical experiments



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

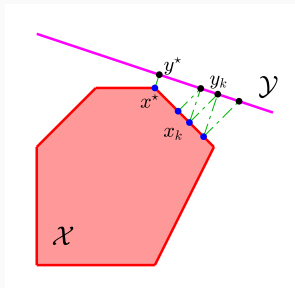
## Best pair problem

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

$$(x_k, y_k) \rightarrow (x^*, y^*)$$

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

Consider

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

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

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

$$y^* - x^* \in \gamma A(x^*),$$

$$x^* - y^* \in \gamma B(y^*).$$

## 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

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

which is  $\gamma$ -cocoercive.

- Regularised monotone inclusion

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

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

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

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

$$\begin{aligned} x_{k+1} &= \mathcal{J}_{\gamma A} \circ (\text{Id} - \gamma {}^\gamma B)(x_k) \\ &= \mathcal{J}_{\gamma A} \circ \left( \text{Id} - \gamma \frac{1}{\gamma} (\text{Id} - \mathcal{J}_{\gamma B}) \right) (x_k) \\ &= \mathcal{J}_{\gamma A} \circ \mathcal{J}_{\gamma B}(x_k). \end{aligned}$$

## An inertial Backward-Backward splitting

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

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

## An inertial BB splitting based on Yosida approximation

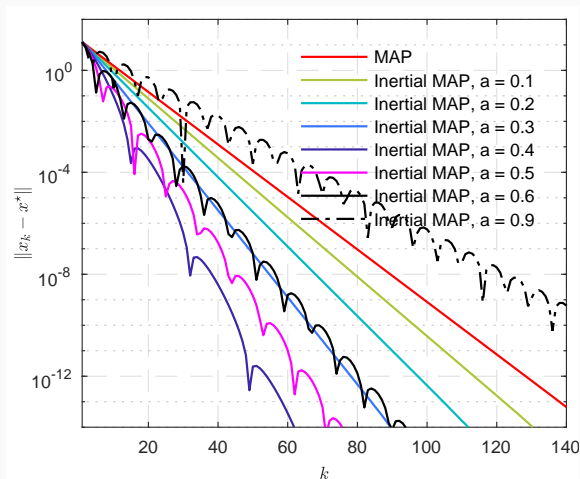
**Initial :**  $x_0 \in \mathbb{R}^n$ ,  $x_{-1} = x_0$  and  $\gamma > 0$ ;

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

- 1 Problem
- 2 Forward-Backward splitting revisit
- 3 MAP continue
- 4 Backward-Backward splitting
- 5 Numerical experiments**

Feasibility problem for two subspaces:

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



- 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.