Introductory Course on Non-smooth Optimisation

Lecture 06 - Primal–Dual splitting

Jingwei Liang

Department of Applied Mathematics and Theoretical Physics

Table of contents

Composed monotone inclusion

Problem

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

Assumptions

- $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone.
- $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping.
- $C : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is maximal monotone.
- $\mathrm{zer}(A + L^* \circ C \circ L) \neq \emptyset$.

Let $x^\star \in \mathrm{zer}(A + L^* \circ C \circ L)$, then $\exists v^\star \in C \circ Lx^\star$ such that

$$0 \in A(x^\star) + L^* v^\star$$

and $Lx^\star \in C^{-1}(v^\star)$.

Saddle-point problem

$$\text{Find } \; v \in \mathbb{R}^m \; \text{ such that } \; \exists x \in \mathbb{R}^n \begin{cases} 0 \in A(x) + L^* v, \\ 0 \in C^{-1}(v) - Lx. \end{cases}$$

Denote $\mathcal{X}$ and $\mathcal{V}$ the set of primal and dual solutions.

Outline

Primal–Dual splitting

Let $x_0 \in \mathbb{R}^n$, $v_0 \in \mathbb{R}^n$ and $\gamma_A, \gamma_C > 0$, $\theta \in [-1, 1]$:

$$\begin{cases} x_{k+1} = \mathcal{J}_{\gamma_A A}(x_k - \gamma_A L^* v_k), \\ \bar{x}_{k+1} = x_{k+1} + \theta(x_{k+1} - x_k), \\ v_{k+1} = \mathcal{J}_{\gamma_C C^{-1}}(v_k + \gamma_C L\bar{x}_{k+1}). \end{cases}$$

- Known as Chambolle-Pock Primal–Dual method in optimisation.
- Douglas–Rachford is the limiting case of Primal–Dual.
- Moreau's identity

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

PPA structure of Primal–Dual splitting

- definition of resolvent

$$\frac{1}{\gamma_A}(x_k - x_{k+1}) - L^* v_k \in A(x_{k+1}),$$
$$\frac{1}{\gamma_C}(v_k - v_{k+1}) + L\big(x_{k+1} + \theta(x_{k+1} - x_k)\big) \in C^{-1}(v_{k+1}).$$

- arrange terms

$$\frac{1}{\gamma_A}(x_k - x_{k+1}) - L^*(v_k - v_{k+1}) \in A(x_{k+1}) + L^* v_{k+1},$$
$$\frac{1}{\gamma_C}(v_k - v_{k+1}) + \theta L(x_{k+1} - x_k) \in C^{-1}(v_{k+1}) - Lx_{k+1}.$$

- inclusion

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{bmatrix} A & L^* \\ -L & C^{-1} \end{bmatrix} \begin{pmatrix} x_{k+1} \\ v_{k+1} \end{pmatrix} + \begin{bmatrix} \mathrm{Id}_n/\gamma_A & -L^* \\ -\theta L & \mathrm{Id}_m/\gamma_C \end{bmatrix} \begin{pmatrix} x_{k+1} - x_k \\ v_{k+1} - v_k \end{pmatrix}.$$

- inclusion

$$\boldsymbol{z} = \begin{pmatrix} x \\ v \end{pmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} A & L^* \\ -L & C^{-1} \end{bmatrix} \quad \text{and} \quad \boldsymbol{V} = \begin{bmatrix} \mathrm{Id}_n/\gamma_A & -L^* \\ -\theta L & \mathrm{Id}_m/\gamma_C \end{bmatrix}.$$

- $\boldsymbol{A}$ is skew symmetric, hence maximal monotone.
- $\boldsymbol{V}$ is symmetric if $\theta = 1$ and moreover positive definite if $\gamma_A \gamma_C \|L\|^2 < 1$.
- $\boldsymbol{V}\boldsymbol{z}_k \in \boldsymbol{A}(\boldsymbol{z}_{k+1}) + \boldsymbol{V}\boldsymbol{z}_{k+1}$, hence

$$\begin{aligned} \boldsymbol{z}_{k+1} &= (\boldsymbol{V} + \boldsymbol{A})^{-1}(\boldsymbol{V}\boldsymbol{z}_k) \\ &= (\mathbf{Id} + \boldsymbol{V}^{-1}\boldsymbol{A})^{-1}(\boldsymbol{z}_k). \end{aligned}$$

- which is PPA under metrix $\boldsymbol{V}$.

Fixed-point formulation

$$z_{k+1} = (\mathbf{Id} + \boldsymbol{V}^{-1}\boldsymbol{A})^{-1}(\mathbf{z}_k).$$

Property space $(\mathbb{R}^n \times \mathbb{R}^n)_{\boldsymbol{V}}$

- $\mathcal{T}_{\mathrm{PD}} = (\mathbf{Id} + \boldsymbol{V}^{-1}\boldsymbol{A})^{-1}$ is firmly non-expansive when $\theta = 1$ and $\gamma_A\gamma_C\|L\|^2 < 1$.
- for $\theta \in [-1, 1[$, a correction step is needed.
- Douglas–Rachford is the limiting case of Primal–Dual when

$$L = \mathrm{Id} \quad \text{and} \quad \gamma_A\gamma_C = 1.$$

Relation with Douglas–Rachford

- let $\theta = 1$ and $L = \mathrm{Id}$, $\gamma_A \gamma_C = 1$.
- change the order of updating variables,

$$\left\lfloor \begin{aligned} v_{k+1} &= \mathcal{J}_{\gamma_C C^{-1}}(v_k + \gamma_C \bar{x}_k), \\ x_{k+1} &= \mathcal{J}_{\gamma_A A}(x_k - \gamma_A v_{k+1}), \\ \bar{x}_{k+1} &= 2x_{k+1} - x_k. \end{aligned} \right.$$

- apply Moreau's identity to $\mathcal{J}_{\gamma_C C^{-1}}$,

$$v_{k+1} = \mathcal{J}_{\gamma_C C^{-1}}(v_k + \gamma_C \bar{x}_k) = v_k + \gamma_C \bar{x}_k - \gamma_C \mathcal{J}_{C/\gamma_C}\left(\frac{v_k + \gamma_C \bar{x}_k}{\gamma_C}\right).$$

- let $\gamma_C = 1/\gamma_A$ and define $z_{k+1} = x_k - \gamma_A v_{k+1}$,

$$\left\lfloor \begin{aligned} u_{k+1} &= \mathcal{J}_{\gamma_A J}(2x_k - z_k), \\ z_{k+1} &= z_k + u_{k+1} - x_k, \\ x_{k+1} &= \mathcal{J}_{\gamma_A R}(z_{k+1}), \end{aligned} \right.$$

Convergence rate

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

$$\mathcal{X} = \{x : ax = 0\}, \ \mathcal{Y} = \{x : bx = 0\}$$

  and assume

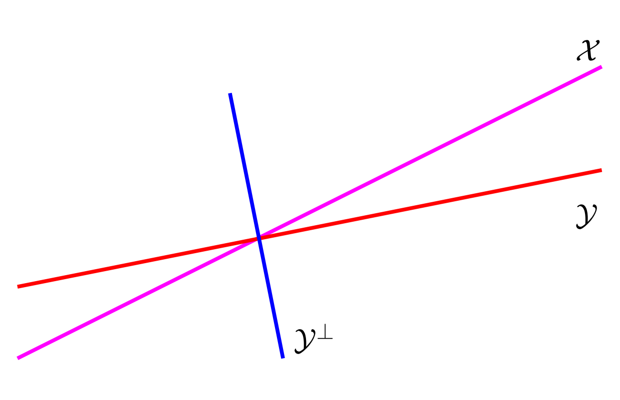
$$1 \leq p \stackrel{\text{def}}{=} \dim(\mathcal{X}) \leq q \stackrel{\text{def}}{=} \dim(\mathcal{Y}) \leq n - 1.$$

- Projection onto subspace

$$\mathcal{P}_{\mathcal{X}}(x) = x - a^T(aa^T)^{-1}ax.$$

- Moreau's identity

$$x = \mathcal{P}_{\mathcal{X}}(x) + \mathcal{P}_{\mathcal{X}^\perp}(x).$$


$\mathcal{X}$
$\mathcal{Y}$
$\mathcal{Y}^\perp$

Convergence rate

- linearisation of PD

$$M_{\text{PD}} = \begin{bmatrix} \mathrm{Id}_n & -\gamma_A \mathcal{P}_{\mathcal{X}} \mathcal{P}_{\mathcal{Y}^\perp} \\ \gamma_C \mathcal{P}_{\mathcal{Y}^\perp} \mathcal{P}_{\mathcal{X}} & \mathrm{Id}_n - 2\gamma_C \gamma_A \mathcal{P}_{\mathcal{Y}^\perp} \mathcal{P}_{\mathcal{X}} \mathcal{P}_{\mathcal{Y}^\perp} \end{bmatrix}.$$

- linearisation of DR

$$M_{\text{DR}} = \begin{bmatrix} \mathrm{Id}_n & -\gamma_A \mathcal{P}_{\mathcal{X}} \mathcal{P}_{\mathcal{Y}^\perp} \\ \frac{1}{\gamma_A} \mathcal{P}_{\mathcal{Y}^\perp} \mathcal{P}_{\mathcal{X}} & \mathrm{Id}_n - 2\mathcal{P}_{\mathcal{Y}^\perp} \mathcal{P}_{\mathcal{X}} \mathcal{P}_{\mathcal{Y}^\perp} \end{bmatrix}.$$

- both M_{PD} and M_{DR} are convergent.
- let ω be the largest principal angle (yet smaller than $\pi/2$) between $\mathcal{X}$ and $\mathcal{Y}^\perp$.
- spectral radius

$$\begin{aligned} \rho(M_{\text{PD}} - M_{\text{PD}}^\infty) &= \sqrt{1 - \gamma_C \gamma_A \cos^2(\omega)} \\ &\geq \sqrt{1 - \cos^2(\omega)} = \sin(\omega) = \cos(\pi/2 - \omega) = \rho(M_{\text{DR}} - M_{\text{DR}}^\infty). \end{aligned}$$

Outline

Parallel sum

Let $C, D : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be two set-valued operators, the parallel sum of C and D is defined by

$$C \Box D \stackrel{\text{def}}{=} (C^{-1} + D^{-1})^{-1}.$$

- $(C \Box D)x = \cup_{y \in \mathbb{R}^n} (A(x) \cap B(x - y))$.
- if C and D are monotone, then $C \Box D$ is monotone.

Primal problem

$$\text{find } x \in \mathbb{R}^n \text{ such that } 0 \in (A+B)(x) + L^*\big((C \square D)(Lx)\big).$$

Assumptions

- $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone, $B : \mathbb{R}^n \to \mathbb{R}^n$ is β_B-cocoercive for some $\beta_B > 0$.
- $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator.
- $C, D : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ are maximal monotone, D is β_D-strongly monotone for some $\beta_D > 0$.
- $0 \in \mathrm{ran}(A + B + L^*(C \square D)L)$.

Saddle-point problem

$$\text{find } v \in \mathbb{R}^m \text{ such that } (\exists x \in \mathbb{R}^n) \begin{cases} 0 \in (A+B)(x) + L^* v, \\ 0 \in (C^{-1} + D^{-1})(v) - Lx. \end{cases}$$

Primal–Dual splitting

Let $x_0 \in \mathbb{R}^n$, $v_0 \in \mathbb{R}^n$ and $\gamma_A, \gamma_C > 0$, $\theta \in [-1, 1]$:

$$\begin{cases} x_{k+1} = \mathcal{J}_{\gamma_A A}(x_k - \gamma_A B(x_k) - \gamma_A L^* v_k), \\ \bar{x}_{k+1} = x_{k+1} + \theta(x_{k+1} - x_k), \\ v_{k+1} = \mathcal{J}_{\gamma_C C^{-1}}(v_k - \gamma_C D^{-1}(v_k) + \gamma_C L\bar{x}_{k+1}). \end{cases}$$

- can be cast as Forward–Backward splitting.

Let $\theta = 1$,

- definition of resolvent

$$\frac{1}{\gamma_A}(x_k - x_{k+1}) - B(x_k) - L^* v_k \in A(x_{k+1}),$$
$$\frac{1}{\gamma_C}(v_k - v_{k+1}) - D^{-1}(v_k) + L\big(x_{k+1} + (x_{k+1} - x_k)\big) \in C^{-1}(v_{k+1}).$$

- arrange terms

$$\frac{1}{\gamma_A}(x_k - x_{k+1}) - B(x_k) - L^*(v_k - v_{k+1}) \in A(x_{k+1}) + L^* v_{k+1},$$
$$\frac{1}{\gamma_C}(v_k - v_{k+1}) - D^{-1}(v_k) + L(x_{k+1} - x_k) \in C^{-1}(v_{k+1}) - Lx_{k+1}.$$

- inclusion

$$-\begin{bmatrix} B & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{pmatrix} x_k \\ v_k \end{pmatrix} \in \begin{bmatrix} A & L^* \\ -L & C^{-1} \end{bmatrix} \begin{pmatrix} x_{k+1} \\ v_{k+1} \end{pmatrix} + \begin{bmatrix} \mathrm{Id}_n/\gamma_A & -L^* \\ -L & \mathrm{Id}_m/\gamma_C \end{bmatrix} \begin{pmatrix} x_{k+1} - x_k \\ v_{k+1} - v_k \end{pmatrix}.$$

FB structure of Primal–Dual splitting

- inclusion

$$\boldsymbol{z} = \begin{pmatrix} x \\ v \end{pmatrix}, \ \ \boldsymbol{A} = \begin{bmatrix} A & L^* \\ -L & C^{-1} \end{bmatrix}, \ \ \boldsymbol{B} = \begin{bmatrix} B & 0 \\ 0 & D^{-1} \end{bmatrix} \ \text{ and } \ \boldsymbol{V} = \begin{bmatrix} \mathrm{Id}_n/\gamma_A & -L^* \\ -L & \mathrm{Id}_m/\gamma_C \end{bmatrix}.$$

- $\boldsymbol{A}$ is skew symmetric, hence maximal monotone.
- $\boldsymbol{B}$ is $\min\{\beta_B, \beta_D\}$-cocoercive.
- $\boldsymbol{V}$ is symmetric positive definite for $\gamma_A \gamma_C \|L\|^2 < 1$.
- $\boldsymbol{V}\boldsymbol{z}_k - \boldsymbol{B}(z_k) \in \boldsymbol{A}(\boldsymbol{z}_{k+1}) + \boldsymbol{V}\boldsymbol{z}_{k+1}$, hence

$$\begin{aligned} \boldsymbol{z}_{k+1} &= (\boldsymbol{V} + \boldsymbol{A})^{-1}(\boldsymbol{V} - \boldsymbol{B})(\boldsymbol{z}_k) \\ &= (\mathbf{Id} + \boldsymbol{V}^{-1}\boldsymbol{A})^{-1}(\mathbf{Id} - \boldsymbol{V}^{-1}\boldsymbol{B})(\boldsymbol{z}_k). \end{aligned}$$

- which is Forward–Backward splitting under metric $\boldsymbol{V}$.

Fixed-point formulation

$$z_{k+1} = (\mathbf{Id} + \boldsymbol{V}^{-1}\boldsymbol{A})^{-1}(\mathbf{Id} - \boldsymbol{V}^{-1}\boldsymbol{B})(\boldsymbol{z}_k).$$

Property space $(\mathbb{R}^n \times \mathbb{R}^n)_{\boldsymbol{V}}$

- $(\mathbf{Id} + \boldsymbol{V}^{-1}\boldsymbol{A})^{-1}$ is firmly non-expansive when $\gamma_A \gamma_C \|L\|^2 < 1$.
- $\mathbf{Id} - \boldsymbol{V}^{-1}\boldsymbol{B}$ is $\frac{1}{2\beta\nu}$-averaged non-expansive with
$$\nu = \left(1 - \sqrt{\gamma_A \gamma_C \|L\|^2}\right) \min\left\{\frac{1}{\gamma_A}, \frac{1}{\gamma_C}\right\}.$$
- $\mathcal{T}_{\mathrm{PD}}$ is $\frac{2\beta\nu}{4\beta\nu - 1}$-averaged non-expansive.

Outline

Infimal convolution $J, G \in \Gamma_0(\mathbb{R}^m)$

$$(J \overset{+}{\vee} G)(\cdot) \stackrel{\text{def}}{=} \inf_{v \in \mathbb{R}^m} J(\cdot) + G(\cdot - v).$$

- $\partial(J \overset{+}{\vee} G)(\cdot) = (\partial J \square \partial G)(\cdot)$.

Example Moreau envelope

$$J \overset{+}{\vee} \frac{1}{2\gamma}\|\cdot\|^2 = \inf_{v \in \mathbb{R}^m} J(v) + \frac{1}{2\gamma}\|\cdot - v\|^2.$$

Conjugate

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

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

- F^* is closed and convex even F is not.

Biconjugate $F^{**} = (F^*)^*$.

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

$$\sigma_{\mathcal{S}}(v) \stackrel{\text{def}}{=} \sup_{x \in \mathcal{S}} \langle x,\, v \rangle = \iota_{\mathcal{S}}^*(y).$$

Example Let $F = \frac{1}{2}\|\cdot\|^2$,

$$F^*(y) = \frac{1}{2}\|y\|^2.$$

Example Let $\|\cdot\|$ be a norm with dual norm $\|\cdot\|$. Let $F = \|x\|$, then

$$F^*(y) = \begin{cases} 0 & \|y\|_* \leq 1, \\ +\infty & o.w. \end{cases}$$

i.e. the indicator function of the dual norm ball.

Informal convolution

$$(J \overset{+}{\vee} G)^* = J^* + G^*.$$

Fenchel–Moreau Let $F : \mathbb{R}^n \to]-\infty, +\infty]$ be a proper function, then F is convex and lower semi-continuous if and only if $F = F^{**}$.

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

Subdifferential If F is closed and convex, then

$$y \in \partial F(x) \iff x \in \partial F^*(y).$$

Moreau's identify Let function $F \in \Gamma_0(\mathbb{R}^n)$ and $\gamma > 0$, then

$$\mathrm{Id} = \mathrm{prox}_{\gamma F}(\cdot) + \gamma\, \mathrm{prox}_{F^*/\gamma}\Big(\frac{\cdot}{\gamma}\Big).$$

Strong convexity Let F be closed and α-strongly convex, then ∇F^* is $\frac{1}{\alpha}$-Lipschitz.

Primal problem

$$\min_{x\in\mathbb{R}^n} R(x) + F(x) + (J \overset{+}{\triangledown} G)(Lx).$$

Assumptions

- $R, F \in \Gamma_0(\mathbb{R}^n)$, and ∇F is $(1/\beta_F)$-Lipschitz continuous for some $\beta_F > 0$.
- $J, G \in \Gamma_0(\mathbb{R}^m)$, G is β_G-strongly convex for $\beta_G > 0$.
- $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping.
- The inclusion $0 \in \mathrm{ran}(\partial R + \nabla F + L^*(\partial J \square \partial G)L)$ holds.

Saddle-point problem

$$\min_{x\in\mathbb{R}^n}\max_{v\in\mathbb{R}^m} R(x)+F(x)+\langle Lx,\, v\rangle-\big(J^*(v)+G^*(v)\big).$$

Dual problem

$$\min_{v\in\mathbb{R}^m} J^*(v)+G^*(v)+(R^*\overset{+}{\vee}F^*)(-L^*v).$$

Denote by $\mathcal{X}$ and $\mathcal{V}$ the sets of solutions of primal and dual problems, respectively.

Primal–Dual splitting

Let $x_0 \in \mathbb{R}^n$, $v_0 \in \mathbb{R}^n$ and $\gamma_{\scriptscriptstyle R}, \gamma_{\scriptscriptstyle J} > 0$, $\theta \in [-1, 1]$:

$$\begin{cases} x_{k+1} = \mathrm{prox}_{\gamma_R R}(x_k - \gamma_{\scriptscriptstyle R} \nabla F(x_k) - \gamma_{\scriptscriptstyle R} L^* v_k), \\ \bar{x}_{k+1} = x_{k+1} + \theta(x_{k+1} - x_k), \\ v_{k+1} = \mathrm{prox}_{\gamma_J J^*}(v_k - \gamma_{\scriptscriptstyle J} \nabla G^*(v_k) + \gamma_{\scriptscriptstyle J} L\bar{x}_{k+1}). \end{cases}$$

- $A = \partial R$, $B = \nabla F$.
- $C^{-1} = \partial J^*$, $D^{-1} = \nabla G^*$.

Primal problem

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

Assumptions

- $R \in \Gamma_0(\mathbb{R}^n)$.
- $J \in \Gamma_0(\mathbb{R}^m)$.
- $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping.
- The inclusion $0 \in \mathrm{ran}(\partial R + L^*\partial J L)$ holds.

Dual problem

$$\min_{v\in\mathbb{R}^m} J^*(v) + R^*(-L^*v)$$

Primal–Dual splitting

Let $x_0 \in \mathbb{R}^n$, $v_0 \in \mathbb{R}^n$ and $\gamma_{R,0}, \gamma_{J,0} > 0$ such that $\gamma_{R,0}\gamma_{J,0}\|L\|^2 \leq 1$:

$$\begin{cases} v_{k+1} = \mathrm{prox}_{\gamma_{J,k}J^*}(v_k + \gamma_{J,k}L\bar{x}_{k+1}), \\ x_{k+1} = \mathrm{prox}_{\gamma_{R,k}R}(x_k - \gamma_{R,k}L^*v_k), \\ \quad \theta_k = \frac{1}{\sqrt{1+2\alpha\gamma_{R,k}}}, \ \gamma_{R,k+1} = \theta_k\gamma_{R,k}, \ \gamma_{J,k+1} = \gamma_{J,k}/\theta_k, \\ \bar{x}_{k+1} = x_{k+1} + \theta_k(x_{k+1} - x_k). \end{cases}$$

- convergence rate

$$\|x_k - x^\star\| = O(1/k^2).$$

R is α-strongly covnex and J is κ-strongly convex.

Primal–Dual splitting

Let $x_0 \in \mathbb{R}^n$, $v_0 \in \mathbb{R}^n$. Choose $\mu = \frac{2\sqrt{\alpha\kappa}}{L}$, $\gamma_R = \frac{\mu}{2\alpha}$, $\gamma_J = \frac{\mu}{2\kappa}$ and $\theta \in [1/(\mu+1), 1]$:

$$\begin{cases} v_{k+1} = \mathrm{prox}_{\gamma_J J^*}(v_k + \gamma_J L\bar{x}_{k+1}), \\ x_{k+1} = \mathrm{prox}_{\gamma_R R}(x_k - \gamma_R L^* v_k), \\ \bar{x}_{k+1} = x_{k+1} + \theta(x_{k+1} - x_k). \end{cases}$$

- convergence rate

$$\|x_k - x^\star\| = O(\eta^k)$$

with $\eta = \frac{1+\theta}{2+\mu}$.

Reference

- 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.
- H. Bauschke and P. L. Combettes. “Convex Analysis and Monotone Operator Theory in Hilbert Spaces”. Springer, 2011.
- 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.