

Convex Optimization for Machine Learning and Computer Vision

Tutorial

09.01.2019

Contents

- 1 ADMM and PDHG
- 2 DRS and PRS
- 3 Customized proximal iteration
- 4 Choice of method

Problem settings

$$\text{minimize } F(Ku) + G(u) \quad \text{over } u \in \mathbb{E}.$$

Assume:

$F, G : \mathbb{E} \rightarrow \mathbb{R}$ is proper, convex, lsc functions and K a matrix.

Alternating direction method of multipliers (ADMM)

Splitting

Let $v = Ku$, we have:

$$\min_u F(Ku) + G(u) = \min_{u,v} J(u, v), J(u, v) := F(v) + G(u) + \delta\{Ku - v = 0\}$$

Alternating direction method of multipliers (ADMM)

Splitting

Let $v = Ku$, we have:

$$\min_u F(Ku) + G(u) = \min_{u,v} J(u, v), J(u, v) := F(v) + G(u) + \delta\{Ku - v = 0\}$$

Augmented Lagrangian

- Idea of Lagrangian: $\delta\{Ku - v = 0\} = \sup_p \langle p, Ku - v \rangle$
- Augmented item $\frac{\tau}{2}\|Ku - v\|^2$ introduces “prox-like” regularity for u, v .
- $\min_{u,v} J(u, v) = \sup_p \inf_{u,v} \mathcal{L}_\tau(u, v; p)$.

$$\mathcal{L}_\tau(u, v; p) := F(v) + G(u) + \langle p, Ku - v \rangle + \frac{\tau}{2}\|Ku - v\|^2$$

Alternating direction method of multipliers (ADMM)

Splitting

Let $v = Ku$, we have:

$$\min_u F(Ku) + G(u) = \min_{u,v} J(u, v), J(u, v) := F(v) + G(u) + \delta\{Ku - v = 0\}$$

Augmented Lagrangian

- Idea of Lagrangian: $\delta\{Ku - v = 0\} = \sup_p \langle p, Ku - v \rangle$
- Augmented item $\frac{\tau}{2}\|Ku - v\|^2$ introduces “prox-like” regularity for u, v .
- $\min_{u,v} J(u, v) = \sup_p \inf_{u,v} \mathcal{L}_\tau(u, v; p).$

$$\mathcal{L}_\tau(u, v; p) := F(v) + G(u) + \langle p, Ku - v \rangle + \frac{\tau}{2}\|Ku - v\|^2$$

$$u^{k+1} = \arg \min_u \mathcal{L}_\tau(u, v^k; p^k)$$

$$v^{k+1} = \arg \min_v \mathcal{L}_\tau(u^{k+1}, v; p^k)$$

$$p^{k+1} = p^k + \tau \cdot \Delta_p \mathcal{L}_\tau(u^{k+1}, v^{k+1}; p) \text{ (gradient ascent for sup of } p\text{)}$$

Primal-dual hybrid gradient (PDHG)

Strong duality

$$\inf_u F(Ku) + G(u) = \sup_p \inf_u \mathcal{L}(u; p), \quad \mathcal{L}(u; p) := \langle p, Ku \rangle + G(u) - F^*(p)$$

→ Saddle-point problem!

Primal-dual hybrid gradient (PDHG)

Strong duality

$$\inf_u F(Ku) + G(u) = \sup_p \inf_u \mathcal{L}(u; p), \quad \mathcal{L}(u; p) := \langle p, Ku \rangle + G(u) - F^*(p)$$

→ Saddle-point problem!

Saddle-point problem updates

$$u^{k+1} = \text{prox}_{\mathcal{L}(\cdot, p^k)/s}(u^k)$$

$$p^{k+1} = \text{prox}_{\mathcal{L}(K(2u^{k+1} - u^k), \cdot)/t}(p^k)$$

→ Note the “over relaxation” term $2u^{k+1} - u^k$ in the update of p .

Primal-dual hybrid gradient (PDHG)

Strong duality

$$\inf_u F(Ku) + G(u) = \sup_p \inf_u \mathcal{L}(u; p), \quad \mathcal{L}(u; p) := \langle p, Ku \rangle + G(u) - F^*(p)$$

→ Saddle-point problem!

Saddle-point problem updates

$$u^{k+1} = \text{prox}_{\mathcal{L}(\cdot, p^k)/s}(u^k)$$

$$p^{k+1} = \text{prox}_{\mathcal{L}(K(2u^{k+1} - u^k), \cdot)/t}(p^k)$$

→ Note the “over relaxation” term $2u^{k+1} - u^k$ in the update of p .

Necessary condition on s, t for convergence

$$s \times t > \|K\|_{\text{spec}}^2$$

DRS, PRS & Reflection operator

Definition

Given a proper, convex, lsc function $J : \mathbb{E} \rightarrow \bar{\mathbb{R}}$ and $\tau > 0$, we call

$$\text{refl}_{\tau J} := 2\text{prox}_{\tau J} - I = 2(I + \tau \partial J)^{-1} - I$$

the **reflection operator** on ∂J .

DRS, PRS & Reflection operator

Definition

Given a proper, convex, lsc function $J : \mathbb{E} \rightarrow \bar{\mathbb{R}}$ and $\tau > 0$, we call

$$\text{refl}_{\tau J} := 2\text{prox}_{\tau J} - I = 2(I + \tau \partial J)^{-1} - I$$

the **reflection operator** on ∂J .

Fix point of refl

$$u = \text{refl}_{\tau J}(u) \Leftrightarrow 0 \in \partial J(u)$$

DRS, PRS & Reflection operator

Definition

Given a proper, convex, lsc function $J : \mathbb{E} \rightarrow \bar{\mathbb{R}}$ and $\tau > 0$, we call

$$\text{refl}_{\tau J} := 2\text{prox}_{\tau J} - I = 2(I + \tau \partial J)^{-1} - I$$

the **reflection operator** on ∂J .

Fix point of refl

$$u = \text{refl}_{\tau J}(u) \Leftrightarrow 0 \in \partial J(u)$$

Problem settings for DRS & PRS

$$\text{minimize } F(u) + G(u) \quad \text{over } u \in \mathbb{E}.$$

Assume:

$F, G : \mathbb{E} \rightarrow \mathbb{R}$ is proper, convex, lsc functions.

Douglas-Rachford- & Peaceman-Rachford splitting

Fix point of refl composition

$$v = \text{refl}_{\tau F}(\text{refl}_{\tau G}(v))$$

$$\Leftrightarrow u = \text{prox}_{\tau G}(v) \wedge u - v \in \tau \partial F(u)$$

$$\Leftrightarrow u = \text{prox}_{\tau G}(v) \wedge 0 \in \partial F(u) + \partial G(u).$$

Douglas-Rachford- & Peaceman-Rachford splitting

Fix point of refl composition

$$\begin{aligned} v &= \text{refl}_{\tau F}(\text{refl}_{\tau G}(v)) \\ \Leftrightarrow u &= \text{prox}_{\tau G}(v) \wedge u - v \in \tau \partial F(u) \\ \Leftrightarrow u &= \text{prox}_{\tau G}(v) \wedge 0 \in \partial F(u) + \partial G(u). \end{aligned}$$

Updates

$$\left\{ \begin{array}{l} u^{k+1} = \text{prox}_{\tau G}(v^k), \\ v^{k+1} = \left(\frac{1}{2}I + \frac{1}{2}\text{refl}_{\tau F} \circ \text{refl}_{\tau G} \right)(v^k), \text{ (DRS)} \\ v^{k+1} = (\text{refl}_{\tau F} \circ \text{refl}_{\tau G})(v^k). \text{ (PRS)} \end{array} \right.$$

Remark

PRS is not necessarily convergent! Check out why in later lectures ;)

Customized proximal iteration

Definition

$$0 \in M(\xi^{k+1} - \xi^k) + R(\xi^{k+1}) \Leftrightarrow \xi^{k+1} = (M + R)^{-1}M\xi^k$$

where M is spd matrix and R is maximal monotone operator.

→ General form for convergence analysis.

Customized proximal iteration

Definition

$$0 \in M(\xi^{k+1} - \xi^k) + R(\xi^{k+1}) \Leftrightarrow \xi^{k+1} = (M + R)^{-1}M\xi^k$$

where M is spd matrix and R is maximal monotone operator.

→ General form for convergence analysis.

ADMM, (scaled-)PDHG, DRS as Customized proximal iteration

ADMM $0 \in \begin{bmatrix} \tau K^\top K & K^\top \\ K & \frac{1}{\tau} I \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix} + \begin{bmatrix} \partial G & K^\top \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix}$

(scaled-)PDHG

$$0 \in \begin{bmatrix} S & -K^\top \\ -K & T \end{bmatrix} \left(\begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} - \begin{bmatrix} u^k \\ p^k \end{bmatrix} \right) + \begin{bmatrix} \partial G & K^\top \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix}$$

DRS $0 \in \begin{bmatrix} \frac{1}{\tau} I & -I \\ -I & \tau I \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix} + \begin{bmatrix} \partial G & I \\ -I & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix}$

Choice of method

- cvx diff \rightarrow gradient descent
- cvx subdiff, 1 non-diff part with ok prox \rightarrow prox GD (FBS)
- cvx subdiff, 2 non-diff parts with ok prox each \rightarrow DRS, PDHG, ADMM
- $F(u) + G(Ku)$, F, G with ok prox, but prox of $G(K\cdot)$ hard
 - \rightarrow PDHG: s, t constrained by K , depend on convex conjugate ...
 - \rightarrow ADMM: more flexible (more terms, non-convex), need to solve linear system with K ...