# ESE415 OPTIMIZATION

Tutorial 7

Xiaojian Xu

04/05/2019

# Guideline

PROXIMAL ALGORITHM RECAP

HW4 RECITATION

# Proximal algorithm recap

1. Why do we want to use proximal method?
2. What are the advantages of the algorithm?

# Proximal algorithm recap

True signal $\boldsymbol{x}$ → Forward model $\boxed{\boldsymbol{H}}$ → measurement $\boldsymbol{y}$

$$\boldsymbol{x} \longrightarrow \boxed{\boldsymbol{H}} \longrightarrow \boldsymbol{y}$$

- A feasible optimization framework to solve the problem

$\boldsymbol{H}$ Forward model, $\boldsymbol{y}$ measurement ⇒ $\widehat{\boldsymbol{x}}$ Estimation

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \{ f(\boldsymbol{x}) \}$$

$$f(\boldsymbol{x}) = \boxed{d(\boldsymbol{x})} + \boxed{r(\boldsymbol{x})}$$

Data-fidelity term

$$d(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2$$

$$r(x) = \|\boldsymbol{x}\|_1 + \|\boldsymbol{x}\|_2 \quad \cdots$$

Regularizer term

Proximal algorithm recap

- Modeled optimization problem

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \{f(\boldsymbol{x})\} \quad \text{with} \quad f(\boldsymbol{x}) = d(\boldsymbol{x}) + \boxed{r(\boldsymbol{x})}$$

Could be nondifferentiable

$$\boxed{\|\boldsymbol{x}\|_1} \quad \cdots \quad \|\boldsymbol{x}\|_2$$

- Gradient descent method

$$\boldsymbol{x}^t \leftarrow \boldsymbol{x}^{t-1} - \gamma \boxed{\nabla f(\boldsymbol{x}^{t-1})}$$

- Proximal gradient method

$$\boldsymbol{x}^t \leftarrow \boxed{\mathsf{P}(\boldsymbol{x}^{t-1})} \quad \begin{cases} \boldsymbol{y}^t = \boldsymbol{x}^{t-1} - \gamma \nabla d(\boldsymbol{x}^{t-1}) \\ \boldsymbol{x}^t = \boxed{\mathsf{prox}_{\gamma r}(\boldsymbol{y}^t)} \end{cases}$$

Proximal gradient mapping

$$\triangleq \mathsf{prox}_{\gamma r}(\boldsymbol{x} - \gamma \nabla d(\boldsymbol{x}))$$

Proximal operation

$$\mathsf{prox}_{\gamma r}(\boldsymbol{y}) \triangleq \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \gamma r(\boldsymbol{x}) \right\}$$

- Getting rid of the forward model H
- Usually have a closed form solution
- Or could be solved using dual property

# HW4 Recitation(P3)

**Problem 3**

(a) Consider the scalar function

$$\varphi(x) = \begin{cases} x & \text{when } x \geq 0 \\ +\infty & \text{when } x < 0. \end{cases}$$

Find the expression for the following proximal operator

$$S_\lambda(y) \triangleq \mathsf{prox}_{\lambda\varphi}(y) = \mathop{\mathsf{arg\,min}}_{x \in \mathbb{R}} \left\{ \frac{1}{2}(x-y)^2 + \lambda\varphi(x) \right\}, \quad y \in \mathbb{R}.$$

(b) Show the following closed-form expression for the proximal operator

$$\mathsf{prox}_{\lambda\|\cdot\|}(\boldsymbol{y}) = \mathop{\mathsf{arg\,min}}_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2 + \lambda\|\boldsymbol{x}\| \right\} = \begin{cases} (\|\boldsymbol{y}\| - \lambda)_+ \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} & \text{when } \boldsymbol{y} \neq 0 \\ \boldsymbol{0} & \text{when } \boldsymbol{y} = 0. \end{cases},$$

where $(x)_+ \triangleq \mathsf{max}(x, 0)$ extracts the positive part of its input.

*Hint: Consider the proximal operator of $g(\boldsymbol{x}) = \varphi(\|\boldsymbol{x}\|)$, $\boldsymbol{x} \in \mathbb{R}^n$, where $\varphi$ is given in (a).*

# Hints (P3)

From part (a) we know

$$S_\lambda(y) = (y-\lambda)_+ = \begin{cases} y-\lambda & \text{if } y \geq \lambda \\ 0 & \text{if } y < \lambda. \end{cases}$$

(b) Define $g(\vec{x}) = \varphi(\|\vec{x}\|)$

$\because \|\vec{x}\| > 0 \Rightarrow \varphi(\|\vec{x}\|) = \|\vec{x}\| \Rightarrow g(\vec{x}) = \varphi(\|\vec{x}\|) = \|\vec{x}\|.$

$\therefore \mathrm{Prox}_{\lambda\|\cdot\|}(\vec{y}) = \mathrm{Prox}_{\lambda \cdot g}(\vec{y}) = \underset{\vec{x} \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\vec{x}-\vec{y}\|^2 + \lambda g(\vec{x}) \right\}.$

Let's consider the following two cases:

① if $\vec{y} = \vec{0}$

$$\mathrm{Prox}_{\lambda\cdot\|\cdot\|}(\vec{y}) = \mathrm{Prox}_{\lambda\cdot\|\cdot\|}(\vec{0}) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\vec{x}-\vec{0}\|^2 + \lambda\|x\| \right\}.$$

$$= \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\vec{x}\|^2 + \lambda\varphi(\|\vec{x}\|) \right\} = x^*.$$

$(t = \|\vec{x}\|)$ $\Rightarrow \underset{x \in \mathbb{R}^n}{\min} \left\{ \frac{1}{2}\|\vec{x}\|^2 + \lambda\varphi(\|\vec{x}\|) \right\} = \underset{t \in \mathbb{R}}{\min} \left\{ \frac{1}{2}t^2 + \lambda\varphi(t) \right\}$

$\Rightarrow t^* = \|x^*\|$

From part (a).

We know $t^* = \underset{t \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}t^2 + \lambda\varphi(t) \right\} = \mathrm{Prox}_{\lambda\varphi}(0) = 0 \Rightarrow x^* = 0.$

② if $\vec{y} \neq 0.$

$$\mathrm{Prox}_{\lambda\|\cdot\|}(\vec{y}) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\vec{x}-\vec{y}\|^2 + \lambda\varphi(\|\vec{x}\|) \right\}.$$

Let's look at this minimization problem:

$$\underset{\vec{x} \in \mathbb{R}^n}{\min} \left\{ \frac{1}{2}\|\vec{x}-\vec{y}\|^2 + \lambda\varphi(\|\vec{x}\|) \right\}$$

$$= \underset{\vec{x} \in \mathbb{R}^n}{\min} \left\{ \frac{1}{2}\|\vec{x}\|^2 - \vec{x}^T\vec{y} + \frac{1}{2}\|\vec{y}\|^2 + \lambda\varphi(\|\vec{x}\|) \right\}$$

$$= \underset{t \in \mathbb{R}}{\min} \left\{ \underset{\vec{x}:\|\vec{x}\|=t}{\min} \left\{ \frac{1}{2}t^2 + \lambda\varphi(t) \underbrace{- \vec{x}^T\vec{y} + \frac{1}{2}\|\vec{y}\|^2} \right\}. \right\}$$

Using Cauchy-Schwarz $\Rightarrow x^* = t^* \frac{\vec{y}}{\|\vec{y}\|}$

with $t^* = \underset{t \in \mathbb{R}}{\operatorname{argmin}} \left\{ \frac{1}{2}t^2 + \lambda\varphi(t) - t\|\vec{y}\| + \frac{1}{2}\|\vec{y}\|^2 \right\}.$

$$= \underset{t \in \mathbb{R}}{\operatorname{argmin}} \left\{ \frac{1}{2}(t - \|\vec{y}\|)^2 + \lambda\varphi(t) \right\} = \mathrm{Prox}_{\lambda\varphi}(\|\vec{y}\|)$$

$$= (\|\vec{y}\| - \lambda)_+$$

$$\Rightarrow x^* = t^* \frac{\vec{y}}{\|\vec{y}\|} = (\|\vec{y}\| - \lambda)_+ \frac{\vec{y}}{\|\vec{y}\|}$$

Combining ① ② $\Rightarrow$ $\mathrm{Prox}_{\lambda\|\cdot\|}(\vec{y}) = \begin{cases} 0 & \text{if } y = 0. \\ (\|\vec{y}\| - \lambda)_+ \frac{\vec{y}}{\|\vec{y}\|} & \text{if } y \neq 0 \end{cases}$

# HW4 Recitation(P4-a)

**Problem 4**

Let $g \in \Gamma^0(\mathbb{R}^n)$, and consider $\theta \neq 0$ and $\boldsymbol{z} \in \mathbb{R}^n$.

(a) Show that for $h(\boldsymbol{x}) = g(\theta \boldsymbol{x} + \boldsymbol{z})$, we have

$$\mathsf{prox}_h(\boldsymbol{x}) = \frac{1}{\theta}\left(\mathsf{prox}_{\theta^2 g}(\theta \boldsymbol{x} + \boldsymbol{z}) - \boldsymbol{z}\right).$$

(b) Show that for $r(\boldsymbol{x}) = \theta g(\boldsymbol{x}/\theta)$, we have

$$\mathsf{prox}_r(\boldsymbol{x}) = \theta \mathsf{prox}_{g/\theta}(\boldsymbol{x}/\theta).$$

## Hints (P4-a)

Solution:

(a). $h(x) = g(\theta x + z)$.

$$\text{Prox}_h(x) = \underset{y\in\mathbb{R}^n}{\arg\min}\left\{ \frac{1}{2}\|y-x\|^2 + h(y)\right\}$$

$$= \underset{y\in\mathbb{R}^n}{\arg\min}\left\{ \frac{1}{2}\|y-x\|^2 + g(\theta y + z)\right\}$$

$$= y^*.$$

Set $v = \theta y + z \Rightarrow y = \frac{1}{\theta}(v - z)$

$$= \frac{1}{\theta}\left\{\underbrace{\underset{v\in\mathbb{R}^n}{\arg\min}\left\{\frac{1}{2}\left\|\frac{1}{\theta}(v-z)-x\right\|^2 + g(v)\right\}}_{\triangleq\, v^*} - z\right\}.$$

$$\Rightarrow y^* = \frac{1}{\theta}(v^* - z).$$

$$v^* = \underset{v}{\arg\min}\left\{\frac{1}{2}\left\|\frac{1}{\theta}(v-z)-x\right\|^2 + g(v)\right\}.$$

$$= \underset{v}{\arg\min}\ \frac{1}{\theta^2}\left\{\frac{1}{2}\|(v-z)-\theta x\|^2 + \theta^2 g(v)\right\}.$$

$$= \underset{v}{\arg\min}\left\{\frac{1}{2}\|v-(\theta x+z)\|^2 + \theta^2 g(v)\right\}.$$

$$\Rightarrow v^* = \text{Prox}_{\theta^2 g}(\theta x + z)$$

$$\Rightarrow y^* = \frac{1}{\theta}(v^* - z) = \frac{1}{\theta}\left(\text{Prox}_{\theta^2 g}(\theta x + z) - z\right)$$

# HW4 Recitation(P5)

## Problem 5

We consider the problem of image denoising. Given a noisy image $\boldsymbol{y} \in \mathbb{R}^n$, our goal is to recover a clean image $\boldsymbol{x} \in \mathbb{R}^n$. We consider an additive noise scenario

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e},$$

where $\boldsymbol{e} \in \mathbb{R}^n$ is the unknown noise degrading the image. Image denoising is often formulated as an optimization problem

$$\widehat{\boldsymbol{x}} = \mathop{\mathsf{arg\,min}}_{\boldsymbol{x}} \{f(\boldsymbol{x})\} \quad \text{with} \quad f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_{\ell_2}^2 + g(\boldsymbol{x}). \tag{3}$$

The goal is generally to pick a function $g$ that achieves the best denoising performance. In this assignment, we will evaluate the quality using the *signal-to-noise ratio (SNR)* defined as

$$\text{SNR (dB)} \triangleq 10\log_{10}\left(\frac{\|\boldsymbol{x}\|^2}{\|\boldsymbol{x} - \boldsymbol{y}\|^2}\right).$$

Note that while images are two-dimensional, for mathematical convenience we will use the vectorized notation for images.

Hints (P5)

- Optimization objective function

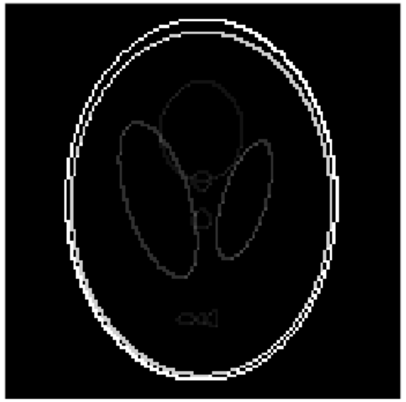
$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \boxed{\mathrm{TV}(\boldsymbol{x})} \right\}$$

$\boldsymbol{x}$

$$\mathrm{TV}(\boldsymbol{x}) = \|\sqrt{|D_x(\boldsymbol{x})|^2+|D_y(\boldsymbol{x})|^2}\|_1$$

$\sqrt{|D_x(\boldsymbol{x})|^2+|D_y(\boldsymbol{x})|^2}$

Edge



$D_x(\boldsymbol{x})$

Horizontal gradient

$D_y(\boldsymbol{x})$

Vertical gradient

Thanks