

Tutorial 7  
Xiaojian Xu  
04/12/2019

# ESE415 OPTIMIZATION

# Guideline



HW4 RECAP



HW5 RECITATION

# HW4 Recap(P5)

## Problem 5

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

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

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

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{f(\mathbf{x})\} \quad \text{with} \quad f(\mathbf{x}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_{\ell_2}^2 + g(\mathbf{x}). \quad (3)$$

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

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

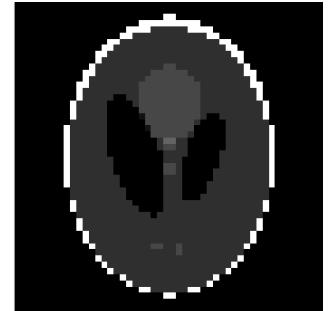
Note that while images are two-dimensional, for mathematical convenience we will use the vectorized notation for images.

$$g(\mathbf{x}) = \frac{\rho}{2} \|\mathbf{D}\mathbf{x}\|_{\ell_2}^2, \quad \rho > 0.$$

- **Q1:** Why could we use  $\ell_2$  norm as the data-fidelity term?
- **Q2:** What does this regularizer  $g(\mathbf{x})$  mean?

- Statistic interpretation of the optimization objective function

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|y - x\|_2^2 + \lambda \text{TV}(x) \right\}$$



True image

$$\begin{aligned} y &= x + e \\ e &\sim N(0, \sigma^2) \end{aligned}$$

observation:  $y$ .  $\Rightarrow$  goal: get  $x$  from  $y$ .

true:  $x$

method  $x = \arg \min_x \{f(x)\}$ ,  $f(x) = \frac{1}{2} \|x - y\|^2 + \lambda r(x)$ .

we show below why we use this objective function from a statistic perspective:

MAP (maximum a posteriori probability):

$$\hat{x}_{\text{MAP}} = \arg \max_x P(x|y) = \arg \max_x \frac{P(y|x)P(x)}{P(y)} = \arg \max_x P(y|x)P(x).$$

$$= \arg \min_x -\log \{P(y|x)\} - \log \{P(x)\}.$$

①  $-\log \{P(y|x)\}$  : log-likelihood.  
 ②  $-\log \{P(x)\}$  : prior

In our denoising model:  $y = e + x$ ,  $e \sim N(0, \sigma^2)$ .

$$\Rightarrow y \sim N(x, \sigma^2). \Rightarrow P(y|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right).$$

PDF of Normal distribution:  

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\begin{cases} \textcircled{1} -\log \{P(y|x)\} = C + \frac{1}{2\sigma^2} (y-x)^2. \\ \textcircled{2} r(x) \triangleq -\log \{P(x)\} \end{cases}$$

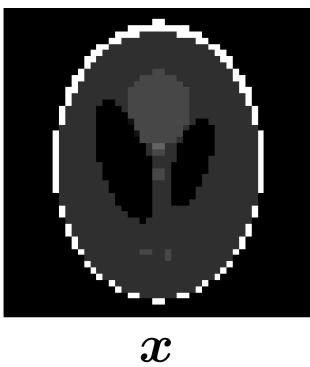
$$\hat{x}_{\text{map}} = \arg \min_x \frac{1}{2\sigma^2} \|y - x\|^2 + \lambda r(x) = \arg \min_x \underbrace{\frac{1}{2} \|y - x\|^2}_{\textcircled{1}} + \lambda r(x) = \arg \min_x f(x),$$

Our objective function.

- Optimization objective function

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \boxed{\text{TV}(\mathbf{x})} \right\}$$

- 1: Anisotropic (our topic today)  
 $\text{TV}_1(\mathbf{x}) = \|D(\mathbf{x})\|_1 = \||D_x(\mathbf{x})| + |D_y(\mathbf{x})|\|_1$   
 2: Isotropic (HW4 is based on this one)  
 $\text{TV}_2(\mathbf{x}) = \|D(\mathbf{x})\|_2 = \|\sqrt{|D_x(\mathbf{x})|^2 + |D_y(\mathbf{x})|^2}\|_1$



$\mathbf{x}$

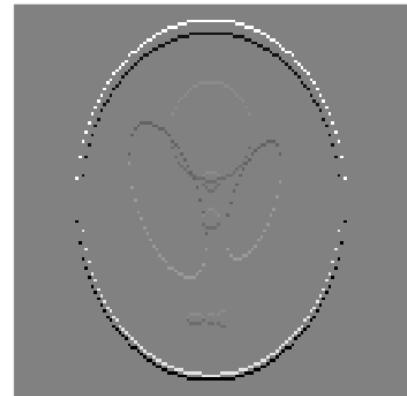
$D(\mathbf{x})$

$D_x(\mathbf{x})$

$D_y(\mathbf{x})$

Horizontal gradient

Vertical gradient



# 1: Anisotropic (our topic today)

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \text{TV}(\mathbf{x}) \right\} \quad \text{with} \quad \text{TV}_1(\mathbf{x}) = \|D(\mathbf{x})\|_1 = \||D_x(\mathbf{x})| + |D_y(\mathbf{x})|\|_1$$

hints

Solution for TV-reg:

$$P: \min \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{D}\mathbf{x}\|_1 \right\}$$

$$(\text{ani}) \text{ TV}_1: \|\mathbf{D}\mathbf{x}\|_1 = \|D_x\mathbf{x}\|_1 + \|D_y\mathbf{x}\|_1 = \sum_n (|D_x\mathbf{x}|_n + |D_y\mathbf{x}|_n) \Rightarrow \boxed{|D_x\mathbf{x}|} + \boxed{|D_y\mathbf{x}|}$$

$$(\text{iso}) \text{ TV}_2: \|\mathbf{D}\mathbf{x}\|_2 = \sqrt{\sum_n (|D_x\mathbf{x}|_n)^2 + (|D_y\mathbf{x}|_n)^2} \Rightarrow \sum_n \boxed{|D_x\mathbf{x}|^2} + \boxed{|D_y\mathbf{x}|^2}$$

↳各向同性

Property 1:

$$\|\mathbf{x}\|_p = \sup_{\|\mathbf{d}\|_q \leq 1} \{ \mathbf{x}^\top \mathbf{d} \} \quad (\text{dual problem}).$$

$$\text{i.e. } \textcircled{1} \quad \|\mathbf{x}\|_2 = \sup_{\|\mathbf{d}\|_2 \leq 1} \{ \mathbf{x}^\top \mathbf{d} \} \Rightarrow \begin{cases} \mathbf{d}^* = \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \\ \|\mathbf{x}\|_2 = \sup \{ \mathbf{x}^\top \mathbf{d}^* \} = \|\mathbf{x}\| \end{cases}$$

$$\textcircled{2} \quad \|\mathbf{x}\|_1 = \sup_{\|\mathbf{d}\|_\infty \leq 1} \{ \mathbf{x}^\top \mathbf{d} \} \Rightarrow \begin{cases} \mathbf{d}^* = \text{sign} \{ \mathbf{x}^\top \} \\ \|\mathbf{x}\|_1 = \sum_i |x_i| \end{cases}$$

We use this property to transfer our problem to its dual problem  $\Rightarrow$ .

# 1: Anisotropic (our topic today)

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \text{TV}(\mathbf{x}) \right\} \quad \text{with} \quad \text{TV}_1(\mathbf{x}) = \|D(\mathbf{x})\|_1 = \||D_x(\mathbf{x})| + |D_y(\mathbf{x})|\|_1$$

hints

$$\begin{aligned} P &= \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \|D\mathbf{x}\|_1 \right\} = \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \max_{\|d\|_\infty \leq 1} \{ [D\mathbf{x}]^\top d \} \right\} \\ &= \min_{\mathbf{x}} \max_{\|d\|_\infty \leq 1} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \lambda [D\mathbf{x}]^\top d \right\} = \max_{\|d\|_\infty \leq 1} \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \mathbf{x}^\top D^\top d \right\} \end{aligned}$$

Define  $f(\mathbf{x}, d) \triangleq \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \mathbf{x}^\top D^\top d$ .

$$\begin{aligned} f(\mathbf{x}, d) &= \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \mathbf{x}^\top D^\top d \\ &= \frac{1}{2} \|\mathbf{x}\|^2 - \mathbf{x}^\top \mathbf{y} + \frac{1}{2} \|\mathbf{y}\|^2 + \mathbf{x}^\top (\lambda D^\top d) \\ &= \frac{1}{2} \|\mathbf{x}\|^2 - \mathbf{x}^\top (\mathbf{y} - \lambda D^\top d) + \frac{1}{2} \|\mathbf{y} - \lambda D^\top d\|^2 - \frac{1}{2} \|\mathbf{y} - \lambda D^\top d\|^2 + \frac{1}{2} \|\mathbf{y}\|^2 \\ &= \frac{1}{2} \|\mathbf{x} - (\mathbf{y} - \lambda D^\top d)\|^2 - \frac{1}{2} \|\mathbf{y} - \lambda D^\top d\|^2 + \frac{1}{2} \|\mathbf{y}\|^2 \end{aligned}$$

$$\begin{aligned} P &= \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \lambda \|D\mathbf{x}\|_1 \right\} = \max_{\|d\|_\infty \leq 1} \min_{\mathbf{x}} \{ f(\mathbf{x}, d) \} \\ &= \max_{\|d\|_\infty \leq 1} \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - (\mathbf{y} - \lambda D^\top d)\|^2 - \frac{1}{2} \|\mathbf{y} - \lambda D^\top d\|^2 + \frac{1}{2} \|\mathbf{y}\|^2 \right\} \\ &= \max_{\|d\|_\infty \leq 1} \min_{\mathbf{x}} \left\{ \underbrace{\frac{1}{2} \|\mathbf{x} - (\mathbf{y} - \lambda D^\top d)\|^2}_{\text{quadratic}} - \underbrace{\frac{1}{2} \|\mathbf{y} - \lambda D^\top d\|^2}_{\text{has nothing to do with } \mathbf{x}} \right\}. \end{aligned}$$

We solve these two optimization problem separately  $\Rightarrow$

# 1: Anisotropic (our topic today)

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \text{TV}(\mathbf{x}) \right\} \quad \text{with} \quad \text{TV}_1(\mathbf{x}) = \|D(\mathbf{x})\|_1 = \||D_x(\mathbf{x})| + |D_y(\mathbf{x})|\|_1$$

hints

$$\begin{cases} \text{For "min" part: } \mathbf{x}^* = \mathbf{y} - \tau D^T d^* \\ \text{For "max" part: } d^* = \arg \max_{\|d\|_\infty \leq 1} \left\{ -\frac{1}{2} \|\mathbf{y} - \tau D^T d\|^2 \right\} = \arg \min_{\|d\|_\infty \leq 1} \left\{ \frac{1}{2} \|\mathbf{y} - \tau D^T d\|^2 \right\} \end{cases}$$

This is a projection problem, and can be solved iteratively using proximal gradient method.

$$d^t = \text{proj}_{\|d\|_\infty \leq 1} (d - r \tau D (D^T d - y)), \text{ where } \text{proj}_{\|z\|_\infty \leq 1}(z) = \begin{cases} z_i & \text{if } |z_i| \leq 1 \\ \frac{z_i}{|z_i|} & \text{if } |z_i| > 1 \end{cases}$$

↑ step size in the gradient descent step

To conclude:

problem:

$$\begin{cases} P = \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \tau \|D\mathbf{x}\|_1 \right\} \\ = \max_{\|d\|_\infty} \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - (\mathbf{y} - \tau D^T d)\|^2 - \frac{1}{2} \|\mathbf{y} - \tau D^T d\|^2 \right\} \end{cases}$$

Solution:

$$\begin{cases} \text{For } t=1, 2, \dots \\ d^t = \text{proj}_{\|d\|_\infty \leq 1} (d^{t-1} - r \tau D (D^T d^{t-1} - y)) \\ x^t = y - \tau D^T d^t \end{cases}$$

This is your reconstructed image at iteration  $t$ .

# HW5 Recitation(P1)

---

Consider the following system of linear equations

$$\begin{cases} x_1 + x_2 + 5x_3 - 7x_4 = 1 \\ x_1 - 3x_2 - x_3 + x_4 = 2 \end{cases}.$$

- a) How many solutions does this system have? Justify your answer.
- b) Find the solution that is closest to the origin in  $\mathbb{R}^4$ .
- c) Find the solution that is closest to the vector  $v = (1, -1, 1, -1)$ .

## Hints (p1)

- a) This system has two independent constraints for four unknowns. This means that the system is *underdetermined* and has infinitely many solutions.
- b) This system of equations can be expressed as

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \text{with} \quad \mathbf{A} = \begin{bmatrix} 1 & 1 & 5 & -7 \\ 1 & -3 & -1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

The Gram matrix and its inverse are given by

$$\mathbf{G} = \mathbf{A}\mathbf{A}^T = \begin{bmatrix} 76 & -14 \\ -14 & 12 \end{bmatrix} \quad \Rightarrow \quad \mathbf{G}^{-1} = (\mathbf{A}\mathbf{A}^T)^{-1} = \frac{1}{358} \begin{bmatrix} 6 & 7 \\ 7 & 38 \end{bmatrix}.$$

The minimum norm solution is given by

$$\mathbf{x}^* = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} = \begin{bmatrix} 1 & 1 \\ 1 & -3 \\ 5 & -1 \\ -7 & 1 \end{bmatrix} \times \frac{1}{358} \begin{bmatrix} 6 & 7 \\ 7 & 38 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{358} \begin{bmatrix} 103 \\ -229 \\ 17 \\ -57 \end{bmatrix}.$$

- c) We would like to solve the following problem

$$\min \|\mathbf{x} - \mathbf{v}\|^2 \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \quad \Leftrightarrow \quad \min \|\mathbf{z}\|^2 \text{ subject to } \mathbf{A}(\mathbf{z} + \mathbf{v}) = \mathbf{b}.$$

Hence, we would like to find the minimum norm solution to the following system

$$\begin{array}{ccccccccc} . & . & . & . & . & [1] & [12] & [-11] \\ . & . & . & . & . & . & . & . \end{array}$$

We can then obtain

$$\mathbf{z}^* = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{c} = \begin{bmatrix} 1 & 1 \\ 1 & -3 \\ 5 & -1 \\ -7 & 1 \end{bmatrix} \times \frac{1}{358} \begin{bmatrix} 6 & 7 \\ 7 & 38 \end{bmatrix} \times \begin{bmatrix} -11 \\ 0 \end{bmatrix} = \frac{1}{358} \begin{bmatrix} -143 \\ 165 \\ -253 \\ 385 \end{bmatrix}.$$

Which leads to

$$\mathbf{x}^* = \mathbf{z}^* + \mathbf{v} = \frac{1}{358} \begin{bmatrix} -143 + 358 \\ 165 - 358 \\ -253 + 358 \\ 385 - 358 \end{bmatrix} = \frac{1}{358} \begin{bmatrix} 215 \\ -193 \\ 105 \\ 27 \end{bmatrix}.$$

Thanks