Tutorial 8
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 $\boldsymbol{y} \in \mathbb{R}^n$, 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

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

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.

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

- Q1: Why could we use l2 norm as the data-fidelity term?
- Q2: What does this regularizer g(x) mean?

- Statistic interpretation of the optimization objective function

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

True image

$\boxed{\begin{array}{l} y = x+e. \\ e \sim N(0, \sigma^2). \end{array}}$

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

true : x

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

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

MAP(maximum a posteriori probability):

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

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

① $-\log\{P(y|x)\}$: log-likelihood.

② $-\log\{P(x)\}$: prior

PDF of Normal distribution: $\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$

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

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

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

$$\hat{x}_{MAP} = \underset{x}{\arg\min}\ \frac{1}{2\sigma^2}\|y-x\|^2 + \lambda r(x). = \underset{x}{\arg\min}\ \underbrace{\frac{1}{2}\|y-x\|^2 + \lambda' r(x)}_{\text{Our objective function.}} = \underset{x}{\arg\min}\ f(x).,$$

- Optimization objective function

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

1: Anisotropic (our topic today)

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

2: Isotropic (HW4 is based on this one)

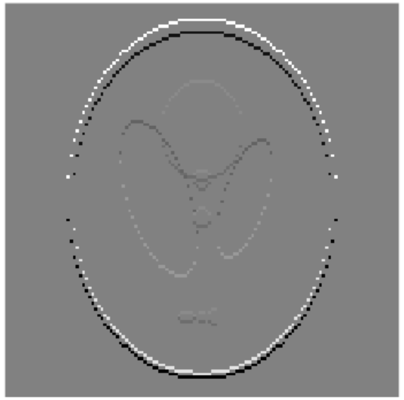
$$\mathrm{TV}_2(\boldsymbol{x}) = \|D(\boldsymbol{x})\|_2 = \|\sqrt{|D_x(\boldsymbol{x})|^2 + |D_y(\boldsymbol{x})|^2}\|_1$$

$\boldsymbol{x}$

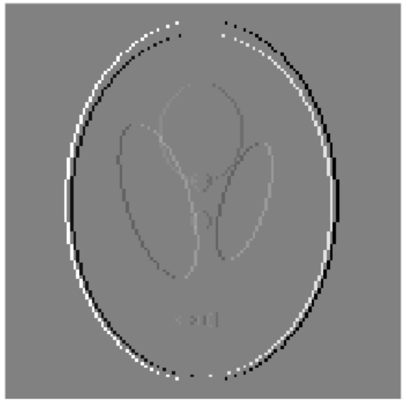
$D(\boldsymbol{x})$

$D_x(\boldsymbol{x})$ $D_y(\boldsymbol{x})$

Horizontal gradient



Vertical gradient



hints

1: Anisotropic (our topic today)

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

Solution for TV-reg:

$P: \min\{\frac{1}{2}\|x-y\|_2^2 + \tau\|Dx\|_1\}.$

(ani) $TV_1 = \|Dx\|_1 = \|D_x x\|_1 + \|D_y x\|_1 = \sum_n \left(|[D_x x]_n| + |[D_y x]_n|\right) \Rightarrow \sum |\boxed{D_x x}| + |\boxed{D_y x}|$

(iso) $TV_2 = \|[Dx]_n\|_2 = \sum_n \sqrt{|[D_x x]_n|^2 + |[D_y x]_n|^2} \Longrightarrow \sum\sqrt{|\boxed{D_x x}|^2 + |\boxed{D_y x}|^2}$

↳各向同性

Property 1:

$$\frac{1}{p} + \frac{1}{q} = 1$$

$$\|x\|_p = \sup_{\|d\|_q \le 1} \{x^T d\} \qquad \text{(dual problem).}$$

i.e. ① $\|x\|_2 = \sup_{\|d\|_2 \le 1} \{x^T d\} \Rightarrow \begin{cases} d^* = \frac{x}{\|x\|_2} \\ \|x\|_2 = \sup\{x^T d^*\} = \|x\| \end{cases}$

② $\|x\|_1 = \sup_{\|d\|_\infty \le 1} \{x^T d\}. \Rightarrow \begin{cases} d^* = \mathrm{sign}\{x^T\}. \\ \|x\|_1 = \sum |x_i| \end{cases}$

We use this property to transfer ous problem to its dual problem ⇒.

1: Anisotropic (our topic today)

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

hints

$$P: \min_x \left\{ \frac{1}{2}\|x-y\|^2 + \tau\|Dx\|_1 \right\} = \min_x \left\{ \frac{1}{2}\|x-y\|^2 + \tau \max_{\|d\|_\infty\le 1} \{[Dx]^T d\} \right\}$$

$$= \min_x \max_{\|d\|_\infty \le 1} \left\{ \frac{1}{2}\|x-y\|^2 + \tau[Dx]^T d \right\} = \max_{\|d\|_\infty\le 1} \min_x \left\{ \frac{1}{2}\|x-y\|^2 + \tau x^T D^T d \right\}$$

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

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

$$P: \min_x \left\{ \frac{1}{2}\|x-y\|^2 + \tau\|Dx\|_1 \right\} = \max_{\|d\|_\infty\le 1} \min_x \{ f(x,d) \}$$

$$= \max_{\|d\|_\infty\le 1} \min_x \left\{ \frac{1}{2}\|x-(y-\tau D^T d)\|^2 - \frac{1}{2}\|y-\tau D^T d\|^2 + \frac{1}{2}\|y\|^2 \right\}$$

$$= \max_{\|d\|_\infty\le 1} \min_x \left\{ \underbrace{\frac{1}{2}\|x-(y-\tau D^T d)\|^2}_{\text{quadratic}} - \underbrace{\frac{1}{2}\|y-\tau D^T d\|^2}_{\text{has nothin to do with } x.} \right\}.$$

We solve these two optimization problem separately $\Rightarrow$

hints

1: Anisotropic (our topic today)

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

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

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

$d^+ = \mathrm{proj}_{\|\cdot\|_\infty \le 1}\left(d - \gamma\tau D(\tau D^T d - y)\right)$, where $\mathrm{proj}_{\|\cdot\|_\infty}(z) = \begin{cases} z_i & \text{if } |z_i| \le 1 \\ \frac{z_i}{|z_i|} & \text{if } |z_i| > 1 \end{cases}$

γ: step size in the gradient descent step

To conclude:

problem:

$$\begin{cases} P: \min\limits_{x}\left\{\frac{1}{2}\|x-y\|^2 + \tau\|Dx\|_1\right\} \\ = \max\limits_{\|d\|_\infty} \min\limits_{x}\left\{\frac{1}{2}\|x-(y-\tau D^T d)\|^2 - \frac{1}{2}\|y-\tau D^T d\|^2\right\} \end{cases}$$

Solution:

$$\begin{cases} \text{For } t = 1, 2, \ldots \\ \quad d^t = \mathrm{proj}_{\|\cdot\|_\infty \le 1}\left(d^{t-1} - \gamma\tau D(\tau D^T d^{t-1} - y)\right) \\ \quad x^t = y - \tau D^T d^t \end{cases}$$

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 $\boldsymbol{v} = (1, -1, 1, -1)$.

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

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

The Gram matrix and its inverse are given by

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

The minimum norm solution is given by

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

c) We would like to solve the following problem

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

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

$$\boldsymbol{A}\boldsymbol{z} = \boldsymbol{c} \quad \text{with} \quad \boldsymbol{c} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 12 \\ 2 \end{bmatrix} = \begin{bmatrix} -11 \\ 0 \end{bmatrix}.$$

We can then obtain

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

Which leads to

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

Thanks