Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Tao Wu
Exercises: Zhenzhang Ye
Winter Semester 2019/20

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 9

Room: 02.09.023

Wednesday, 22.01.2020, 12:15-14:00

Submission deadline: Monday, 20.01.2020, 16:15, Room 02.09.023

Majorize Minimization (12+6 Points)

Exercise 1 (4 Points)**.** Given following convex minimization problem:

$$\min_{x\in\mathbb{R}} f(x) + g(x), \tag{1}$$

where $f(x)$ is non-differentiable and $g(x)$ is L-Lipschitz differentiable, we can apply proximal gradient to solve it.
State that proximal gradient is an example of majorize minimization.
Hint: Write down the explicit updating step of proximal gradient. Then figure out the majorant.

Solution. Firstly, we write down the proximal gradient step at k-th iteration:

$$\begin{aligned} x^{k+1/2} &= x^k - \tau\nabla g(x^k) \\ x^{k+1} &= \text{argmin}_x \frac{1}{2\tau}\left\|x - x^{k+1/2}\right\|^2 + f(x) \end{aligned} \tag{2}$$

where τ is the step size in $(0, \frac{1}{L}]$. We combine above two equations into one:

$$\begin{aligned} x^{k+1} &= \text{argmin}_x \frac{1}{2\tau}\left\|x - (x^k - \tau\nabla g(x^k))\right\|^2 + f(x) \\ &= \text{argmin}_x \frac{1}{2\tau}\left\|x - x^k\right\|^2 + \left\langle \nabla g(x^k), x - x^k\right\rangle + f(x) \end{aligned} \tag{3}$$

We want to create a majorant for original function. We are allowed to add arbitrary constant in the last energy function to make it satisfy the definition of majorant and not change argmin. Therefore, we can construct following majorant function:

$$q_\tau(x, y) = \frac{1}{2\tau}\|x - y\|^2 + \langle\nabla g(y), x - y\rangle + g(y) + f(x) \tag{4}$$

The first condition is easily shown by replacing y with x. The second condition is satisfied due to the L-lipschitz differentiable and τ in $(0, \frac{1}{L}]$.

Now consider following theorem:

Theorem. Suppose $J(u)$ is an even, differentiable function on $\mathbb{R}$ such that thte ratio $J'(u)/u$ is decreasing on $(0, \infty)$. Then the quadratic:

$$\hat{J}(v; u) = \frac{J'(u)}{2u}(v^2 - u^2) + J(u)$$

is a majorant of $J(\cdot)$ at the point u.

Exercise 2 (4 Points)**.** Prove above theory.
Hint: 1. Since J is an even function, prove the case $0 \leq v \leq u$ and then generate to other cases.
2. You might use this equation: $J(u) - J(v) = \int_v^u J'(z)dz$.

Solution. It is obvious that $\hat{J}(u; u) = J(u)$, we need to prove the other condition $\hat{J}(v; u) \geq J(v), \forall v$. Considering the case $0 \leq v \leq u$, we have:

$$\begin{aligned} J(u) - J(v) &= \int_v^u J'(z)dz \\ &= \int_v^u \frac{J'(z)}{z} z dz \\ &\geq \frac{J'(u)}{u} \int_v^u z dz \\ &= \frac{J'(u)}{u} \frac{1}{2}(u^2 - v^2) \\ &= J(u) - \hat{J}(v; u), \end{aligned}$$

where the inequality comes from the assumption that $J'(u)/u$ is decreasing. It follows that $\hat{J}(v; u) \geq J(v), \forall v$. For the case that $0 \leq u \leq v$, we use the same idea starting with $J(v) - J(u)$. Because J and $\hat{J}$ are both even, all othere cases reduce to these two cases.

Exercise 3 (4 Points)**.** The Huber function $h_\varepsilon : \mathbb{R} \to \mathbb{R}$ is given as

$$h_\varepsilon(u) = \begin{cases} \frac{x^2}{2\varepsilon} & \text{if } |x| \leq \varepsilon, \\ |x| - \frac{\varepsilon}{2} & \text{otherwise.} \end{cases}$$

Given the energy function:

$$\text{argmin}_{u \in \mathbb{R}^n} \frac{1}{2} \|u - f\|^2 + H_\varepsilon(Ku), \tag{5}$$

where $f \in \mathbb{R}^n$ is a known variable, the Huber function $H_\varepsilon : \mathbb{R}^m \to \mathbb{R}$ defined as $H_\varepsilon(x) = \sum_{i=1}^m h_\varepsilon(x_i)$ and $K \in \mathbb{R}^{m \times n}$ is a linear operator.

1. State that we can apply above theorem for the Huber function part *i.e.* $H_\varepsilon(Ku)$ (consider the reparameterization).

2. Compute the majorant of the Huber function part using above theorem.

Solution. 1. Because Huber function is seperable. We can decouple it elementwisely if we set $Ku = v$. Obviously, the Huber function satisfies the assumption in above theorem. While sum will not change the greater relation, the sum of majorant of each function will be the majorant of the sum of functions.

2. We first compute the gradient of $h_\varepsilon(y_i)$:

$$h_\varepsilon(y_i) = \begin{cases} -1, & y_i \leq -\varepsilon \\ \frac{y}{\varepsilon}, & y_i \in (-\varepsilon, \varepsilon) \\ 1, & y_i \geq \varepsilon \end{cases} \tag{6}$$

Plugging it into the formula, we get:

$$g_\varepsilon(x_i, y_i) = \begin{cases} -\frac{1}{2y_i}x_i^2 - \frac{y_i}{2} - \frac{\varepsilon}{2}, & y_i \leq -\varepsilon \\ \frac{1}{2\varepsilon}x^2, & y_i \in (-\varepsilon, \varepsilon) \\ \frac{1}{2y_i}x_i^2 + \frac{y_i}{2} - \frac{\varepsilon}{2}, & y_i \geq \varepsilon \end{cases} \tag{7}$$

In total, we have the majorant for $H_\varepsilon(x)$:

$$\sum_{i=1}^{m} g_\varepsilon(x_i, y_i)$$

Exercise 4 (6 points). 1. Show that the Huber penalty can be expressed as the infimal convolution of the functions $f : \mathbb{R} \to \mathbb{R}$ with $f(x) := \frac{x^2}{2\varepsilon}$ and $g : \mathbb{R} \to \mathbb{R}$ with $g(x) := |x|$:

$$h_\varepsilon(x) = (f \,\square\, g)(x).$$

2. Compute the convex conjugate of the function $H_\varepsilon : \mathbb{R}^N \to \mathbb{R}$.

Solution. 1. For the infimal convolution we have that

$$\left(\frac{(\cdot)^2}{2\varepsilon} \,\square\, |\cdot|\right)(u) = \inf_{v \in \mathbb{R}^n} \frac{1}{2\varepsilon}(u - v)^2 + |v|$$

The minimizer for that is attained at

$$v^* = \begin{cases} u + \varepsilon & \text{if } u < -\varepsilon \\ 0 & \text{if } u \in [-\varepsilon, \varepsilon] \\ u - \varepsilon & \text{if } u > \varepsilon. \end{cases}$$

Plugging this in we obtain for the infimum:

$$\left(\frac{(\cdot)^2}{2\varepsilon} \,\square\, |\cdot|\right)(u) = \begin{cases} \frac{1}{2\varepsilon}\varepsilon^2 + |u + \varepsilon| = \frac{\varepsilon}{2} - u - \varepsilon = -\frac{\varepsilon}{2} - u & \text{if } u < -\varepsilon \\ \frac{1}{2\varepsilon}u^2 & \text{if } u \in [-\varepsilon, \varepsilon] \\ \frac{1}{2\varepsilon}\varepsilon^2 + u - \varepsilon = -\frac{\varepsilon}{2} + u & \text{if } u > \varepsilon. \end{cases}$$

which is obviously what we were supposed to show.

2. Since the elements of the sum are independent the sum decouples. That is one can compute the conjugate of the Huber terms seperately. By some computation from the result above, we obtain:

$$h_\varepsilon^*(y_i) = \frac{\varepsilon}{2} y_i^2 \ + \ \iota_{|\cdot| \leq 1}(y_i).$$

The overall conjugate is then given as:

$$H_\varepsilon^*(y) = \frac{\varepsilon}{2} \|y\|_2^2 \ + \ \iota_{|\cdot|_\infty \leq 1}(y).$$

Image Denoising (Due Date: 27.01) (10 Points)

Given a noisy input image, we want to remove the noises by solving following minimization problem:

$$\text{argmin}_{u \in \mathbb{R}^N} \frac{1}{2} \|u - f\|^2 + \rho H_\varepsilon(Ku).$$

where $f \in \mathbb{R}^N$ is the input image with N pixels, ρ is a scalar weighting the smooth regularizor, H_ε is the Huber function defined as before and K is the gradient operator.

Your tasks are: (1) use the majorize minimization with majorant in exercise 3 to solve above problem.

(2) check the optimality condition by duality gap with the convex conjugate in exerice 4.