Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Tao Wu
Exercises: Yuesong Shen, Zhenzhang Ye
Winter Semester 2018/19

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 8

Room: 01.09.014

Wednesday, 19.12.2018, 12:15-14:00

Submission deadline: Monday, 17.12.2018, 16:15, Room 01.09.014

Prox and Gradient descent (8+4 Points)

Exercise 1 (6 Points). Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and bounded from below. Consider the scaled gradient descent iteration:

$$x^{k+1} = x^k - \tau^k (H^k)^{-1} \nabla f(x^k). \tag{1}$$

For each k, assume that $\tau^k > 0$, $\nabla f(x^k) \neq 0$, and $H^k \in \mathbb{R}^{n\times n}$ is symmetric positive definite.

1. Prove that for given x^k and H^k, there exists some $\bar{\tau}^k > 0$ such that any $\tau^k \in (0, \bar{\tau}^k]$ will fulfill the following Armijo condition:

$$f(x^{k+1}(\tau^k)) \leq f(x^k) + c \left\langle \nabla f(x^k), x^{k+1}(\tau^k) - x^k \right\rangle, \tag{2}$$

   for some constant $0 < c < 1$.

2. Assume that for each k the condition (2) is satisfied with some chosen $\tau^k > 0$. In addition, assume that $\liminf_{k\to\infty} \tau^k = C_1 > 0$ and $\limsup_{k\to\infty} \lambda_{\max}(H^k) = C_2 < \infty$. Prove $\lim_{k\to\infty} \nabla f(x^k) = 0$.

Exercise 2 (6 points). We want to show that the proximal operator of the nuclear norm is the proximal operator of the ℓ_1-norm applied to the singular values of the input argument. Formally, let $Y \in \mathbb{R}^{n\times n}$ and let $Y = U\Sigma V^\top$ be the singular value decomposition of Y. Our goal is to prove that

$$\text{prox}_{\tau\|\cdot\|_{\text{nuc}}}(Y) = U\text{diag}(\{(\sigma_i - \tau)_+\})V^\top,$$

where $\text{diag}(\{\sigma_i - \tau\}_+) := \text{diag}(\{\max\{0, \sigma_i - \tau\}\}) = \text{prox}_{\tau\|\cdot\|_1}(\{\sigma_i\})$ is the shrinkage (or soft thresholding) operator applied to the singular values σ_i of Y.

For this, we will argue in 2 steps:

1. In general, the proximal operator is well-defined and returns a unique minimizer, why? Give your argument. In our case, denote $\hat{X} = \text{prox}_{\tau\|\cdot\|_{\text{nuc}}}(Y)$, what do we have for the optimality condition?

2. Show that $\hat{X} = U\text{diag}(\{(\sigma_i - \tau)_+\})V^\top$ verifies the optimality condition, and argue that this concludes our proof.

Hint: for step 2, recall from sheet 3 that the subdifferential at point $X \in \mathbb{R}^{n\times n}$ with $s \geq 0$ zero singular values is given as

$$\partial \left\|\cdot\right\|_{\text{nuc}}(X) = \left\{U_1 V_1^\top + U_2 M V_2^\top : M \in \mathbb{R}^{s\times s}, \left\|M\right\|_{\text{spec}} \leq 1\right\}, \tag{3}$$

where $\left\|\cdot\right\|_{\text{spec}}$ denotes the spectral norm, i.e., the largest singular value.

Rewriting the expressions of X and Y with an appropriately defined decomposition $V = [V_1\ V_2]$, $U = [U_1\ U_2]$ can be helpful.