Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Tao Wu
Exercises: Emanuel Laude, Zhenzhang Ye
Summer Semester 2018

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 6

Room: 02.09.023

Wednesday, 06.06.2018, 12:15-14:00

Submission deadline: Monday, 04.06.2018, 16:15, Room 02.09.023

Porximal operator (8+4 Points)

Exercise 1 (4 Points)**.** Assume function $J : \mathbb{R}^n \to \mathbb{R}$ is convex and subdifferentiable on its domain. Show that u^* minimizes J if and only if $u^* = \text{prox}_J(u^*)$.

Solution. If u^* minimizes J, we have $J(u) \geq J(u^*)$ for any $u \in \text{domJ}$. Therefore,

$$J(u) + \frac{1}{2}\|u - u^*\|_2^2 \geq J(u^*) = J(u^*) + \frac{1}{2}||u^* - u^*||_2^2$$

for any u, which means $u^* = \text{prox}_J(u^*)$.
Conversely, as J is a convex function, a point $u = \text{prox}_J(u^*)$ if and only if

$$0 \in \partial J(u) + (u - u^*)$$

Replace u with u^*, we can get the optimality condition. Since J is convex, we know that u^* is the minimizer.

Exercise 2 (4 Points)**.** Prove following properties of proximal operator:

- If $J(u) = \alpha f(u) + b$, with $\alpha > 0$, then $\text{prox}_{\lambda J}(v) = \text{prox}_{\alpha\lambda f}(v)$.
- If $J(u) = f(Qu)$, where Q is an orthogonal matrix, then $\text{prox}_{\lambda J}(v) = Q^\top \text{prox}_{\lambda f}(Qv)$

Solution.

- $$\begin{aligned}\text{prox}_{\lambda J}(v) &= \text{argmin}_u\, J(u) + \frac{1}{2\lambda}||u - v||^2 \\ &= \text{argmin}_u\, \alpha f(u) + b + \frac{1}{2\lambda}||u - v||^2 \\ &= \text{argmin}_u\, \alpha(f(u) + \frac{1}{2\lambda\alpha}||u - v||^2) \\ &= \text{argmin}_u\, f(u) + \frac{1}{2\lambda\alpha}||u - v||^2 \\ &= \text{prox}_{\alpha\lambda f}(v)\end{aligned}$$

- $$\begin{aligned}
\operatorname{prox}_{\lambda J}(v) &= \operatorname{argmin}_u J(u) + \frac{1}{2\lambda}||u-v||^2 \\
&= \operatorname{argmin}_u f(Qu) + \frac{1}{2\lambda}||u-v||^2 \\
&= \operatorname{argmin}_u f(Qu) + \frac{1}{2\lambda}||Qu-Qv||^2 \\
&\overset{t=Qu}{=} Q^\top \operatorname{argmin}_t f(t) + \frac{1}{2\lambda}||t-Qv||^2 \\
&= Q^\top \operatorname{prox}_{\lambda f}(Qv)
\end{aligned}$$

Exercise 3 (4 Points)**.** Show that the ℓ_1-norm proximity operator of $v \in \mathbb{R}^n$ is given as

$$\operatorname{prox}_{\lambda\|\cdot\|_1}(v) = u \in \mathbb{R}^n, \quad u_i := \begin{cases} v_i + \lambda & \text{if } v_i < -\lambda \\ 0 & \text{if } v_i \in [-\lambda, \lambda] \\ v_i - \lambda & \text{if } v_i > \lambda. \end{cases}$$

Solution. We begin reformulating the optimality condition

$$0 \in \partial\left(\frac{1}{2\lambda}(u_i - v_i)^2 + |u_i|\right)$$

of the optimal u_i

$$0 = \frac{1}{\lambda}(u_i - v_i) + p, \quad p \in \partial|u_i| := \begin{cases} -1 & \text{if } u_i < 0 \\ [-1, 1] & \text{if } u_i = 0 \\ 1 & \text{if } u_i > 0 \end{cases}$$

$$v_i \in u_i + \begin{cases} -\lambda & \text{if } u_i < 0 \\ [-\lambda, \lambda] & \text{if } u_i = 0 \\ \lambda & \text{if } u_i > 0. \end{cases}$$

Recall that we are looking for a u_i that satisfies the condition above given a fixed v_i. We distinguish the following cases:

1. Assume $v_i \in [-\lambda, \lambda]$. Choosing $u_i := 0$ satisfies the condition above.
2. Assume $v_i > \lambda$. Choosing $u_i := v_i - \lambda$ again satisfies the condition.
3. Assume $v_i < -\lambda$. Choosing $u_i := v_i + \lambda$ is the right choice.

Matrix Completion (Due: 11.06.2018) (12 Points)

Exercise 4 (12 Points)**.** In this exercise, you need to solve a matrix completion problem and use duality gap as stop criterion. The problem can be descirbed as following: assume there exists one true matrix $A \in \mathbb{R}^{m\times n}$ which is unknown. Only a observation matrix $F \in \mathbb{R}^{m\times n}$ is given. In this matrix F, some entries $ij \in \Omega$ with a small noise is known. Now, we want to recover the original true matrix A as accurate as possible.
If $|\Omega| < mn$, this problem is undetermined. However, if we assume A has low rank, it is possible to recover it by formulating following problem:

$$\operatorname{argmin}_{X\in\mathbb{R}^{m\times n}} \|X\|_{\text{nuc}} + \frac{\alpha}{2}\|P_\Omega(X-F)\|^2$$

where $P_\Omega(X)$ can be viewed as a projection operation on X. It extracts the corresponding entry $x_{ij}, ij \in \Omega$ from X and turn them into a vector $\mathbb{R}^{|\Omega|}$. Its transpose $P_\Omega^\top$ is putting the entry back to the original position of a matrix. You are asked to do the following steps:

- Performe a proximal gradient to get the optimal solution X^*. In detail, apply the gradient step to the latter quadratic term and proximal step on nuclear norm.
- Compute the dual problem of original and use the optimality condition on the saddle point problem to recover the dual variable.(You don't have to write down this. Only need to compute this on your own).
- Now at each iteration, besides update X, you need to compute the dual variable as well. And compute the primal energy and dual energy. The stop criterion now becomes either it exceeds the maximal iteration number or the duality gap is less than ϵ.

Hint: To compute the eigenvalues of a matrix X, use the function $svd(X)$ in matlab. Use *doc svd* for more details. Besides, you might need to consider $||X||_{\text{spec}} > 1$ when you compute the dual problem. In the template, there is a variable *tol*. Because of the accuracy, instead of using $\sigma_1 > 1$ directly, use $(\sigma_1 - 1) > tol$, where σ_1 is the maximum eigenvalue of X.

Solution. Here, we denote $G(X)$ and $F(KX)$ as following:

$$\operatorname{argmin}_{X\in\mathbb{R}^{m\times n}} \underbrace{\|X\|_{\text{nuc}}}_{F(KX)} + \underbrace{\frac{\alpha}{2}\|P_\Omega(X-F)\|^2}_{G(X)}$$

where K is an identity matrix in this case. Now we need to compute $G^*(Y)$ and $F^*(Y)$.

$$G^*(Y) = \sup_X \langle X, Y\rangle - \frac{\alpha}{2}\|P_\Omega(X-F)\|^2$$

which leads to

$$Y - \alpha P_\Omega^\top P_\Omega(X^* - F) = 0$$

$$\Rightarrow X^* = X_{ij}^* = \begin{cases} \frac{1}{\alpha}Y_{ij} + F_{ij}, & \text{if} ij \in \Omega \\ 0, & \text{otherwise} \end{cases}$$

where X^* is the optimal solution for $G^*(Y)$. Replacing X in $G^*(Y)$ with X^*, we get

$$G^*(Y) = \langle \frac{Y}{\alpha} + F, Y\rangle|_\Omega + \frac{1}{2\alpha}\|Y|_\Omega\|^2 = \frac{1}{2\alpha}\|Y|_\Omega\|^2 + \langle F, Y\rangle|_\Omega$$

where $Y|_\Omega$ represents extracting elements from Y with index in Ω.

$$F^*(Y) = \sup_X \langle X, Y\rangle - \|X\|_{\text{nuc}}$$

which leads to

$$Y \in \partial \|X\|_{\text{nuc}}$$

Recall that the subdifferential of nuclear norm is

$$\partial \|X\|_{\text{nuc}} = \{Z \in \mathbb{R}^{m\times n} : \langle Z, X\rangle = \|X\|_{\text{nuc}}, \|Z\|_{\text{spec}} \leq 1\}.$$

Therefore, we can get $F^*(Y)$:

$$F^*(Y) = \begin{cases} 0, & \text{if } \|Y\|_{\text{spec}} \leq 1 \\ \infty, & \text{otherwise} \end{cases}$$

Therefore, the dual problem is

$$\text{argmin}_Y\, G^*(-KY) + F^*(Y) = \frac{1}{2\alpha}\|Y|_\Omega\|^2 - \langle F, Y\rangle + \delta_{\|Y\|_{\text{spec}}\leq 1}(Y)$$

To recover dual variable from primal variable, we use $-Y = \partial G(X)$, which leads to

$$Y|_\Omega = \alpha(F - X)|_\Omega$$