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 8

Room: 02.09.023
Wednesday, 15.01.2020, 12:15-14:00
Submission deadline: Monday, 13.01.2020, 16:15, Room 02.09.023

Exact Line Search (14+6 Points)

Exercise 1 (6 Points)**.** Let $Q \in \mathbb{R}^{n\times n}$ be a positive definite symmetric matrix. Prove the following inequality for any vector $x \in \mathbb{R}^n$

$$\frac{(x^\top x)^2}{(x^\top Q x)(x^\top Q^{-1} x)} \geq \frac{4\lambda_n\lambda_1}{(\lambda_n+\lambda_1)^2},$$

where λ_n and λ_1 are, respectively, the largest and smallest eigenvalues of Q.

Solution. Since $Q \in \mathbb{R}^{n\times n}$ is symmetric positive definite, we can write it as $Q = U\Lambda U^\top$, where Λ is a $n \times n$ diagonal matrix containing the eigenvalues of Q.

$$\begin{aligned}\frac{(x^\top x)^2}{(x^\top Q x)(x^\top Q^{-1} x)} &= \frac{\|x\|^4}{\langle x, U\Lambda U^\top x\rangle\langle x, U\Lambda^{-1}U^\top x\rangle} \\ &= \frac{\|y\|^4}{\langle y, \Lambda y\rangle\langle y, \Lambda^{-1} y\rangle} \\ &= \frac{\|y\|^2\|y\|^2}{\left(\sum_{i=1}^n y_i^2\lambda_i\right)\left(\sum_{i=1}^n y_i^2\frac{1}{\lambda_i}\right)} = \dots\end{aligned} \tag{1}$$

where $y = U^\top x \in \mathbb{R}^n$. ($y$ can attain any value in $\mathbb{R}^n$ since $U^\top$ is full rank.)

We used the fact that $\langle x, U\Lambda U^\top x\rangle = \langle U^\top x, \Lambda U^\top x\rangle = \langle y, \Lambda y\rangle$ and $\|x\|^2 = \langle x, UU^\top x\rangle = \langle U^\top x, U^\top x\rangle = \|y\|^2$.

Now let $\xi_i = y_i^2/\|y\|^2$, then we have

$$\dots = \frac{1}{\left(\sum_{i=1}^n \frac{y_i^2}{\|y\|^2}\lambda_i\right)\left(\sum_{i=1}^n \frac{y_i^2}{\|y\|^2}\frac{1}{\lambda_i}\right)} = \frac{1/\sum_{i=1}^n \xi_i\lambda_i}{\sum_{i=1}^n(\xi_i\frac{1}{\lambda_i})} =: \frac{\phi(\xi)}{\Psi(\xi)}. \tag{2}$$

Since $\xi_i \geq 0$ and $\sum_{i=1}^n \xi_i = 1$ we have a ratio of two functions involving convex combinations.

Let $f(x) = 1/x$, and $\bar\lambda := \sum_i \xi_i\lambda_i$. Then $\phi(\xi) = f(\bar\lambda)$. Furthermore, take the affine function

$$g(\lambda) = \frac{1}{\lambda_n} + \frac{\frac{1}{\lambda_1}-\frac{1}{\lambda_n}}{\lambda_n-\lambda_1}(\lambda_n-\lambda).$$

Since f is convex (on $\mathbb{R}^+$) we have that $f(\lambda) \leq g(\lambda), \forall \lambda > 0$. Then

$$\Psi(\xi) = \sum_i \xi_i f(\lambda_i) \leq \sum_i \xi_i g(\lambda_i) = g(\sum_i \xi_i \lambda_i) = g(\bar{\lambda})$$

Then we have

$$\begin{aligned}
\frac{(x^\top x)^2}{(x^\top Q x)(x^\top Q^{-1} x)} &= \frac{\phi(\xi)}{\Psi(\xi)} = \frac{f(\bar{\lambda})}{\Psi(\xi)} \\
&\geq \frac{f(\bar{\lambda})}{g(\bar{\lambda})} \geq \min_{\lambda \in [\lambda_1, \lambda_n]} \frac{f(\lambda)}{g(\lambda)} = \min_{\lambda \in [\lambda_1, \lambda_n]} \frac{\lambda^{-1}}{\frac{1}{\lambda_n} + \frac{\frac{1}{\lambda_1} - \frac{1}{\lambda_n}}{\lambda_n - \lambda_1}(\lambda_n - \lambda)} \\
&= \min_{\lambda \in [\lambda_1, \lambda_n]} \frac{\lambda^{-1}}{\frac{1}{\lambda_n} + \frac{\frac{\lambda_n - \lambda_1}{\lambda_1 \lambda_n}}{\lambda_n - \lambda_1}(\lambda_n - \lambda)} \\
&= \min_{\lambda \in [\lambda_1, \lambda_n]} \frac{\lambda^{-1}}{\frac{1}{\lambda_n} + \frac{1}{\lambda_1 \lambda_n}(\lambda_n - \lambda)} \qquad (3) \\
&= \lambda_1 \lambda_n \min_{\lambda \in [\lambda_1, \lambda_n]} \frac{1}{\lambda(\lambda_1 + \lambda_n - \lambda)} \\
&\overset{\hat{\lambda} = \frac{\lambda_1 + \lambda_n}{2}}{=} \frac{\lambda_1 \lambda_n}{\frac{\lambda_1 + \lambda_n}{2}(\lambda_1 + \lambda_n - \frac{\lambda_1 + \lambda_n}{2})} \\
&= \frac{\lambda_1 \lambda_n}{\frac{\lambda_1^2 + 2\lambda_1 \lambda_n + \lambda_n^2}{2} - \frac{(\lambda_1 + \lambda_n)^2}{4}} = \frac{4 \lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}.
\end{aligned}$$

Exercise 2 (6 Points)**.** Let $Q \in \mathbb{R}^{n \times n}$ be symmetric positive definite, and $b \in \mathbb{R}^n$. As in the previous exercise, denote the eigenvalues of Q as $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Consider the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto \frac{1}{2} x^\top Q x - b^\top x$ and show gradient descent with exact line search has the following convergence property:

$$\left\| x^{k+1} - x^* \right\|_Q^2 \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 \left\| x^k - x^* \right\|_Q^2,$$

where $x^* \in \mathbb{R}^n$ denotes the global minimizer of f.
Hint: use the inequality from exercise 1.

Solution. From the lecture we know that the line search procedure has the solution

$$\tau^k = \text{argmin}_\tau\, f(x^k - \tau \nabla f(x^k)) = \frac{\left\| \nabla f(x^k) \right\|^2}{\left\| \nabla f(x^k) \right\|_Q^2}$$

Furthermore, note that $\nabla f(x^k) = Q x^k - b = Q(x^k - x^*)$. We have the following equalities:

$$\left\| x^k - x^* \right\|_Q^2 = \langle x^k - x^*, Q(x^k - x^*) \rangle = \langle Q(x^k - x^*), Q^{-1} Q(x^k - x^*) \rangle = \left\| \nabla f(x^k) \right\|_{Q^{-1}}^2.$$

$$
\begin{aligned}
&\left\|x^k - x^*\right\|_Q^2 - \left\|x^{k+1} - x^*\right\|_Q^2 = \left\|x^k - x^*\right\|_Q - \left\|x^k - \tau_k \nabla f(x^k) - x^*\right\|_Q^2 = \\
&\left\|x^k\right\|_Q^2 - 2\langle x^k, x^* \rangle_Q + \|x^*\|_Q^2 - (\left\|x^k - \tau_k \nabla f(x^k)\right\|_Q^2 - 2\langle x^k - \tau_k \nabla f(x^k), x^* \rangle_Q + \|x^*\|_Q^2) = \\
&\left\|x^k\right\|_Q^2 - \left\|x^k - \tau_k \nabla f(x^k)\right\|_Q^2 - 2\tau_k \langle \nabla f(x^k), x^* \rangle_Q = \\
&2\tau_k \langle x^k, \nabla f(x^k) \rangle_Q - 2\tau_k \langle \nabla f(x^k), x^* \rangle_Q - \tau_k^2 \left\|\nabla f(x^k)\right\|_Q^2 = \\
&2\tau_k \langle \nabla f(x^k), x^k - x^* \rangle_Q - \tau_k^2 \left\|\nabla f(x^k)\right\|_Q^2 = \\
&2\frac{\left\|\nabla f(x^k)\right\|^4}{\|\nabla f(x^k)\|_Q^2} - \frac{\left\|\nabla f(x^k)\right\|^4}{\|\nabla f(x^k)\|_Q^2} = \frac{\left\|\nabla f(x^k)\right\|^4}{\|\nabla f(x^k)\|_Q^2}.
\end{aligned}
\tag{4}
$$

Hence, using exercise 1, we arrive at the estimate from the lecture

$$
\begin{aligned}
\left\|x^{k+1} - x^*\right\|_Q^2 &= \left\|x^k - x^*\right\|_Q^2 - \frac{\left\|\nabla f(x^k)\right\|^4}{\|\nabla f(x^k)\|_Q^2} = \left\|\nabla f(x^k)\right\|_{Q^{-1}}^2 - \frac{\left\|\nabla f(x^k)\right\|^4}{\|\nabla f(x^k)\|_Q^2} \\
&= \left\|\nabla f(x^k)\right\|_{Q^{-1}}^2 \left(1 - \frac{\left\|\nabla f(x^k)\right\|^4}{\|\nabla f(x^k)\|_Q^2 \|\nabla f(x^k)\|_{Q^{-1}}^2}\right) \\
&= \left\|x^k - x^*\right\|_Q^2 \left(1 - \frac{\left\|\nabla f(x^k)\right\|^4}{\|\nabla f(x^k)\|_Q^2 \|\nabla f(x^k)\|_{Q^{-1}}^2}\right) \\
&\overset{\text{exercise 1}}{\leq} \left(1 - \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}\right) \left\|x^k - x^*\right\|_Q^2 = \left(\frac{\lambda_1^2 - 2\lambda_1\lambda_n + \lambda_n^2}{(\lambda_1 + \lambda_n)^2}\right) \left\|x^k - x^*\right\|_Q^2 \\
&= \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 \left\|x^k - x^*\right\|_Q^2.
\end{aligned}
\tag{5}
$$

Image Denoising (Due Date: 13.01) (10 Points)

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

$$\text{argmin}_u \frac{1}{2} \|u - f\|^2 + \rho H_\varepsilon(Ku).$$

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 task is using gradient descent with line search to solve it.

For detailed line search, you can refer to Algorithm 3.5 and 3.6 (Page 79) in Numerical Optimization

`http://www.apmath.spbu.ru/cnsa/pdf/monograf/Numerical_Optimization2006.pdf`.

For zoom function, you can directly use bisection.