

Lecture: T. Wu

Exercises: E. Laude, T. Möllenhoff

Summer Semester 2017

Computer Vision Group

Institut für Informatik

Technische Universität München

Weekly Exercises 6

Room: 02.09.023

Monday, 19.06.2017, 12:15-14:00

Submission deadline: Wednesday, 14.06.2017, Room 02.09.023

Gradient descent

(4 Points + 4 Bonus)

Exercise 1 (4 Points). Let $Q \in \mathbb{R}^{n \times n}$ be a positive definite symmetric matrix. Prove the following inequality for any vector $x \in \mathbb{R}^n$

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

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 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 .

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

We used the fact that $\langle x, U\Lambda U^\top x \rangle = \langle U^\top x, \Lambda U^\top x \rangle = \text{tr}(U^\top x x^\top \Lambda U) = \text{tr}(x x^\top \Lambda) = \langle x, \Lambda x \rangle$. Now let $\xi_i = x_i^2 / \|x\|^2$, then we have

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

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

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

Then we have

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

Exercise 2 (4 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 $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Consider the quadratic function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto \frac{1}{2}x^\top Qx - b^\top x$ and show gradient descent with exact line search has the following convergence property:

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

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

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

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

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

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

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

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