Numerical Optimization with applications: Homework 02

104021601 林俊傑
104021602 吳彥儒
104021615 黃翊軒

October 13, 2016

Exercise 5. *Prove that* $\|Bx\| \geq \frac{\|x\|}{\|B^{-1}\|}$ *for any nonsingular matrix* B*. Use this fact to establish* (3.19)*.*

Proof. For simplicity, we drop the iteration index k in the proof.
Note that from (3.2) we use the fact $P = -B^{-1}\nabla f$. Thus by multiplying both sides by B and taking transport, we have $BP = -\nabla f$ and $P^T B^T = -\nabla f^T$. We are now prepared to estimate $\cos\theta$:

$$\begin{aligned}
\cos\theta &= \frac{-\nabla f^T P}{\|\nabla f\|\|P\|} \\
&= \frac{\left(P^T B^T\right) P}{\|\nabla f\|\|B^{-1}BP\|} \\
&= \frac{P^T BP}{\|BP\|\|B^{-1}BP\|} \\
&\geq \frac{P^T BP}{\|B\|\|P\|\|B^{-1}\|\|BP\|} \\
&= \left(\frac{P^T BP}{\|P\|\|BP\|}\right)\frac{1}{\|B\|\|B^{-1}\|} \\
&\geq \frac{1}{\|B\|\|B^{-1}\|} \\
&\geq \frac{1}{M},
\end{aligned}$$

where the last two inequality hold by the assumption that B is positive definite and has a uniformly bounded condition number. Therefore, (3.19) follows. $\square$

Exercise 7. *Prove the result (3.28) by working through the following steps. First, use (3.26) to show that*

$$\|x_k - x^*\|_Q^2 - \|x_{k+1} - x^*\|_Q^2 = 2\alpha_k \nabla f_k^T Q(x_k - x^*) - \alpha_k^2 \nabla f_k^T Q \nabla f_k$$

where $\|\cdot\|_Q$ *is defined by (3.27). Second, use the fact that* $\nabla f_k = Q(x_k - x^*)$ *to obtain*

$$\|x_k - x^*\|_Q^2 - \|x_{k+1} - x^*\|_Q^2 = \frac{2(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)} - \frac{(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)}$$

and

$$\|x_k - x^*\|_Q^2 = \nabla f_k^T Q^{-1} \nabla f_k.$$

Proof. (1)

$$\begin{aligned}
&\|x_k - x^*\|_Q^2 - \|x_{k+1} - x^*\|_Q^2 \\
=\ & 2f(x_k) - 2f(x_{k+1}) \\
=\ & x_k^T Q x_k - 2b^T x_k - (x_k - \alpha_k \nabla f_k)^T Q (x_k - \alpha_k \nabla f_k) + 2b^T(x_k - \alpha_k \nabla f_k) \\
=\ & x_k^T Q(\alpha_k \nabla f_k) + (\alpha_k \nabla f_k)^T Q x_k - \alpha_k^2 \nabla f_k^T Q \nabla f_k - 2\alpha_k b^T \nabla f_k \\
=\ & 2\alpha_k \nabla f_k^T Q x_k - \alpha_k^2 \nabla f_k^T Q \nabla f_k - 2\alpha_k (Qx^*)^T \nabla f_k \\
=\ & 2\alpha_k \nabla f_k^T Q(x_k - x^*) - \alpha_k^2 \nabla f_k^T Q \nabla f_k
\end{aligned}$$

(2) Combining the result of (1) and the fact that $\nabla f_k = Q(x_k - x^*)$, we have

$$\begin{aligned}
& \|x_k - x^*\|_Q^2 - \|x_{k+1} - x^*\|_Q^2 \\
= \;& 2\alpha_k \nabla f_k^T \nabla f_k - \alpha_k^2 \nabla f_k^T Q \nabla f_k \\
= \;& 2(\frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k}) \nabla f_k^T \nabla f_k - (\frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k})^2 \nabla f_k^T Q \nabla f_k \\
= \;& \frac{2(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)} - \frac{(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)}
\end{aligned}$$

(3) The definition of the weight norm: $\|x\|_Q^2 = x^T Q x$, therefore

$$\begin{aligned}
\|x_k - x^*\|_Q^2 &= (x_k - x^*)^T Q (x_k - x^*) \\
&= (x_k - x^*) T (Q Q^{-1}) Q (x_k - x^*) \\
&= (Q(x_k - x^*))^T Q^{-1} Q (x_k - x^*) \\
&= \nabla f_k^T Q^{-1} \nabla f_k
\end{aligned}$$

Since Q is symmetric and nonsingular.

(4) Now we turn to prove (3.28) by using the result of (2) and (3).

$$\begin{aligned}
\|x_{k+1} - x^*\|_Q^2 &= \|x_k - x^*\|_Q^2 - \frac{(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)} \\
&= \|x_k - x^*\|_Q^2 - \frac{(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)(\nabla f_k^T Q^{-1} \nabla f_k)} (\nabla f_k^T Q^{-1} \nabla f_k) \\
&= (1 - \frac{(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)(\nabla f_k^T Q^{-1} \nabla f_k)}) \|x_k - x^*\|_Q^2
\end{aligned}$$

Therefore the proof is completed.

□

Exercise 8. *Let Q be a positive definite symmetric matrix. Prove that for any vector x, we have*

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

where λ_n and λ_1 are, respectively the largest and smallest eigenvalues of Q. (This relation, which is known as the Kantorovich inequality, can be used to deduce (3.29) from (3.28).)

Proof. Since Q is positive definite and symmetric, we have eigenvalue decompsition $Q = U \Lambda U^T$. Let $x = Uy$. Then

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} = \frac{(y^T y)^2}{(y^T \Lambda y)(y^T \Lambda^{-1} y)} = \frac{(\sum_{i=1}^n y_i^2)^2}{(\sum_{i=1}^n \lambda_i y_i^2)(\sum_{i=1}^n y_i^2 / \lambda_i)}$$

Let $\eta_i = \frac{y_i^2}{\sum_{j=1}^n y_j^2}$ and $f(\lambda) = \frac{1}{\lambda}$. Then

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} = \frac{1}{(\sum_{i=1}^n \lambda_i \eta_i)(\sum_{i=1}^n f(\lambda_i) \eta_i)}$$

Let $\lambda = \sum_{i=1}^n \lambda_i \eta_i \;, \quad \lambda_f = \sum_{i=1}^n f(\lambda_i) \eta_i$

Since $\eta_i \geq 0 \quad \forall i$ and $\sum_{i=1}^n \eta_i = 1$, $\lambda_1 \leq \lambda \leq \lambda_n$

Write $\lambda_i = \frac{\lambda_n - \lambda_i}{\lambda_n - \lambda_1}\lambda_1 + \frac{\lambda_i - \lambda_1}{\lambda_n - \lambda_1}\lambda_n$. This shows λ_i is a convex combination of λ_1 and λ_n $\quad \forall i$

$\because f$ is convex $\quad \therefore f(\lambda_i) \leq \frac{\lambda_n - \lambda_i}{\lambda_n - \lambda_1} f(\lambda_1) + \frac{\lambda_i - \lambda_1}{\lambda_n - \lambda_1} f(\lambda_n)$

Therefore,

$$\lambda_f \leq \sum_{i=1}^n \left[\frac{\lambda_n - \lambda_i}{\lambda_n - \lambda_1} f(\lambda_1) + \frac{\lambda_i - \lambda_1}{\lambda_n - \lambda_1} f(\lambda_n)\right] \eta_i = \sum_{i=1}^n \left[\frac{\lambda_n - \lambda_i}{\lambda_n - \lambda_1}\frac{1}{\lambda_1} + \frac{\lambda_i - \lambda_1}{\lambda_n - \lambda_1}\frac{1}{\lambda_n}\right]\eta_i$$

$$= \sum_{i=1}^n \frac{\eta_i}{\lambda_n - \lambda_1}\left[\frac{\lambda_n - \lambda_i}{\lambda_1} + \frac{\lambda_i - \lambda_1}{\lambda_n}\right] = \sum_{i=1}^n \frac{\eta_i}{\lambda_n - \lambda_1}\left[\frac{\lambda_n^2 - \lambda_i\lambda_n + \lambda_i\lambda_1 - \lambda_1^2}{\lambda_n\lambda_1}\right]$$

$$= \sum_{i=1}^n \frac{\eta_i}{\lambda_n - \lambda_1}\left[\frac{(\lambda_n + \lambda_1)(\lambda_n - \lambda_1) - \lambda_i(\lambda_n - \lambda_1)}{\lambda_n\lambda_1}\right] = \sum_{i=1}^n \frac{\lambda_n + \lambda_1 1 - \lambda_i}{\lambda_n\lambda_1}\eta_i$$

$$= \frac{\lambda_n \sum_{i=1}^n \eta_i + \lambda_1 \sum_{i=1}^n \eta_i - \sum_{i=1}^n \lambda_i\eta_i}{\lambda_n\lambda_1} = \frac{\lambda_n + \lambda_1 - \lambda}{\lambda_n\lambda_1}$$

We conclude that

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} = \frac{1}{\lambda\lambda_f} \geq \frac{\lambda_n\lambda_1}{\lambda(\lambda_n + \lambda_1 - \lambda)} \geq \frac{\lambda_n\lambda_1}{\max_{\lambda\in[\lambda_1,\lambda_n]} \lambda(\lambda_n + \lambda_1 - \lambda)}$$

Let $g(\lambda) = \lambda(\lambda_n + \lambda_1 - \lambda) = -\lambda^2 + (\lambda_n + \lambda_1)\lambda$. Then $g(\lambda)$ has maxmun at $\bar{\lambda} = \frac{\lambda_n + \lambda_1}{2} \in [\lambda_1, \lambda_n]$.

$$g(\bar{\lambda}) = -\frac{(\lambda_n + \lambda_1)^2}{4} + \frac{(\lambda_n + \lambda_1)^2}{2} = \frac{(\lambda_n + \lambda_1)^2}{4}$$

This implies that

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} \geq \frac{\lambda_n\lambda_1}{g(\bar{\lambda})} = \frac{4\lambda_n\lambda_1}{(\lambda_n + \lambda_1)^2}$$

□

Exercise 13. *Show that the quadratic function that interpolates $\phi(0)$, $\phi'(0)$, and $\phi(\alpha_0)$ is given by (3.57). Then, make use of the fact that the sufficient decrease condition (3.6a) is not satisfied at $\alpha(0)$ to show that this quadratic has positive curvature and that the minimizer satisfies*

$$\alpha_1 < \frac{\alpha_0}{2(1 - c_1)}.$$

Since c_1 is chosen to be quite small in practice, this inequality indicates that α_1 cannot be much greater than $\frac{1}{2}$ (and may be smaller), which gives us an idea of the new step length.

Proof. By assuming a quadratic function

$$\phi_q(\alpha) = a\alpha^2 + b\alpha + c$$

and solving coefficients through standard calculations, we have

$$\phi_q(0) = c = \phi(0).$$

Also, since

$$\phi_q'(\alpha) = 2a\alpha + b,$$

we find

$$\phi_q'(0) = b = \phi'(0).$$

On the other hand,

$$\phi_q(\alpha_0) = \phi(\alpha_0) = a\alpha_0^2 + \phi'(0)\alpha_0 + \phi(0),$$

we obtain

$$a = \frac{\phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0}{\alpha_0^2}.$$

By assumptions, we now discuss a senario that (3.6a) is not satisfied at α_0. Thus we have

$$\phi(\alpha_0) > \phi(0) + c_1\alpha_0\phi'(0).$$

Hence, a minor manipulation of this inequality gives

$$a = \frac{\phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0}{\alpha_0} > 0,$$

which guarantees that the quadratic has positive curvature. Moreover, since the minimizer of a quadratic is $-b/2a$, we obtain

$$\begin{aligned}\alpha_1 = \frac{-b}{2a} &= \frac{-\alpha_0^2\phi'(0)}{2[\phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0]} \\ &= \frac{\alpha_0}{2[1 - \frac{\phi(\alpha_0)-\phi(0)}{\alpha_0\phi'(0)}]} \\ &\leq \frac{\alpha_0}{2(1-c_1)},\end{aligned}$$

where the second equality holds by dividing $-\alpha_0\phi'(0)$ both upper and lower sides, and the last inequality follows from the given senario.

$\square$