Numerical Optimization with applications: Homework 03

104021601 林俊傑
104021602 吳彥儒
104021615 黃翊軒

November 2, 2016

Exercise 6. *The Cauchy-Schwarz inequality states that for any vectors u and v, we have*

$$|u^T v|^2 \leq (u^T u)(v^T v),$$

with equality only when u and v are parallel. When B is positive definite, use this inequality to show that

$$\gamma := \frac{\|g\|^4}{(g^T B g)(g^T B^{-1} g)} \leq 1,$$

with equality only if g and Bg (and $B^{-1}g$) are parallel.

Proof. B is a positive definite matrix, so there exists an orthonormal matrix Q and a diagonal matrix $\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$ s.t. $B = Q\Lambda Q^T$.

Define the matrix $\sqrt{B} = Q\sqrt{\Lambda}Q^T$ where $\sqrt{\Lambda} = \begin{pmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_n} \end{pmatrix}$

Obviously, $\sqrt{B}$ is also symmetric.
Claim:$(\sqrt{B})^{-1} = \sqrt{B^{-1}}$
proof of claim:
$(\sqrt{B})^{-1} = Q(\sqrt{\Lambda})^{-1}Q^T = Q\sqrt{\Lambda^{-1}}Q^T = \sqrt{B^{-1}}$
We proved the claim.
Now we use the claim above, the symmetricity of $\sqrt{B}$ and Cauchy-Schwarz inequality. We have the following statement:

$$\begin{aligned}
\|g\|^4 &= (g^T g)^2 = (g^T \sqrt{B}(\sqrt{B})^{-1} g)^2 = ((\sqrt{B}g)^T(\sqrt{B^{-1}}g))^2 \\
&\leqslant (\sqrt{B}g)^T(\sqrt{B}g)(\sqrt{B^{-1}}g)^T(\sqrt{B^{-1}}g) \\
&= (g^T\sqrt{B}\sqrt{B}g)(g^T\sqrt{B^{-1}}\sqrt{B^{-1}}g) \\
&= (g^T B g)(g^T B^{-1} g)
\end{aligned}$$

When the equality holds only if $\sqrt{B}g$ and $\sqrt{B^{-1}}g$ are parallel.
i.e.$\sqrt{B}g = k\sqrt{B^{-1}}g$ for some constant k.
1. Multiplying both side by $\sqrt{B}$.
 $Bg = kg \Longrightarrow Bg$ and g are parallel.

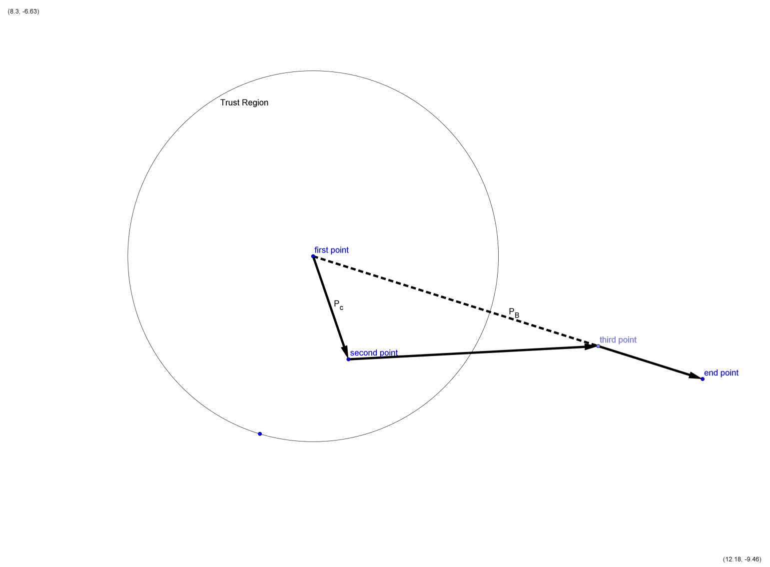
2. Multiplying both side by $\sqrt{B^{-1}}$.
 $g = kB^{-1}g \Longrightarrow B^{-1}g$ and g are parallel.

□

Exercise 7. *When B is positive definite, the double-dog leg method constructs a path with three line segments from the origin to the full step. The four points that define the path are*

- *the origin;*
- *the unconstrained Cauchy step $p^c = -(g^T g)/(g^T Bg)g$;*
- *a fraction of the full step $\bar{\gamma} p^B = -\bar{\gamma} B^{-1} g$, for some $\bar{\gamma} \in (\gamma, 1]$, where γ is defined in the previous question; and*
- *the full step $p^B = -B^{-1} g$*

Show that $\|p\|$ increases monotonically along this path.
(Note: The double-dogleg method, as discussed in Dennis and Schnabel [92, Section 6.4.2], was for some time thought to be superior to the standard dogleg method, but later testing has not shown much difference in performance.)



Proof. It is obviously that $\|p\|$ increases monotonically along the first segment and the last segment because $\alpha\|v\|$ increases as α increases, where $\alpha \in (0, 1)$. Now we consider the second segment. Let $P^A = -\bar{\gamma} B^{-1} g$, and $P^U = -(g^T g)/(g^T Bg)g$, then the parametrization of the second segment is

$$P(\alpha) = \alpha(P^A - P^U) + P^U.$$

Define

$$\begin{aligned} h(\alpha) &= (1/2)\|P(\alpha)\|^2 \\ &= (1/2)\|\alpha(P^A - P^U) + P^U\|^2 \\ &= (1/2)\|P^U\|^2 + \alpha(P^U)^T(P^A - P^U) + (1/2)\alpha^2\|P^A - P^U\|^2 \end{aligned}$$

Then we have

$$\begin{aligned} h'(\alpha) &= -(P^U)^T(P^U - P^B) + \alpha\|P^U - P^B\|^2 \\ &\geq -(P^U)^T(P^U - P^A) \\ &= \frac{g^T g}{g^T Bg} g^T \left(-\frac{g^T g}{g^T Bg} g + \bar{\gamma} B^{-1} g \right) \\ &= g^T g \frac{gB^{-1}g}{gBg} \left(\bar{\gamma} - \frac{(g^T g)^2}{(g^T Bg)(g^T B^{-1} g)} \right) \\ &> 0 \end{aligned}$$

Since $h'(\alpha) > 0$ for all $\alpha \in (0, 1)$, $h(\alpha)$ is increasing monotonically on $(0, 1)$, that is, $\|p(\alpha)\|$ is increasing monotonically on $(0, 1)$.Therefore, the $\|p\|$ increases monotonically along this segament.

□

Exercise 8. *Show that*

$$\lambda^{(l+1)} = \lambda^{(l)} - \frac{\phi_2(\lambda^{(l)})}{\phi_2'(\lambda^{(l)})}, \quad \textit{and} \quad \lambda^{(l+1)} = \lambda^{(l)} + \left(\frac{\|p_l\|}{\|q_l\|}\right)^2 \left(\frac{\|p_l\| - \Delta}{\Delta}\right)$$

are equivalents.

Proof. First,we caculate

$$\phi_2'(\lambda) = \frac{d}{d\lambda}\left(\frac{1}{\|p(\lambda)\|}\right) = \frac{d}{d\lambda}\left(\|p(\lambda)\|^2\right)^{-1/2} = -\frac{1}{2}\left(\|p(\lambda)\|^2\right)^{-3/2}\frac{d}{d\lambda}\|p(\lambda)\|^2$$

Since B is symmetric, there is an orthonormal matrix U and a diagonal matrix Λ such that $B = U\Lambda U^T$, where

$$\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

Then, $B + \lambda I = U(\Lambda + \lambda I)U^T$. We have,

$$p(\lambda) = -U(\Lambda + \lambda I)U^T g = -\sum_{j=1}^{n} -\frac{u_j^T g}{\lambda_j + \lambda} u_j$$

where u_j denotes the jth column of U. Therefore, by orthonormality of $u_1, u_2, \cdots, u_n$, we have

$$\|p(\lambda)\|^2 = \sum_{j=1}^{n}\frac{(u_j^T g)^2}{(\lambda_j + \lambda)^2}$$

Hence, we can caculate

$$\frac{d}{d\lambda}\|p(\lambda)\|^2 = -2\sum_{j=1}^{n}\frac{(u_j^T g)^2}{(\lambda_j + \lambda)^3}$$

On the other hand, we have

$$\begin{aligned}\|q_l\|^2 = \left\|R^{-T}p_l\right\|^2 = p_l^T R^{-1}R^{-T}p_l = [-(R^TR)^{-1}g]^T(R^TR)^{-1}[-(R^TR)^{-1}g] = g^T[(R^TR)^{-1}]^3 g \\ = g^T[(B + \lambda^{(l)}I)^{-1}]^3 g = g^T U(\Lambda + \lambda^{(l)}I)^{-3}U^T g = \sum_{j=1}^{n}\frac{(u_j^T g)^2}{(\lambda_j + \lambda^{(l)})^3}\end{aligned}$$

We conclude that

$$\phi_2'(\lambda^{(l)}) = -\frac{1}{2}\left\|p(\lambda^{(l)})\right\|^{-3}\left(-2\sum_{j=1}^{n}\frac{(u_j^T g)^2}{(\lambda_j + \lambda^{(l)})^3}\right) = \|p_l\|^{-3}\|q_l\|^2$$

Finally, we get

$$-\frac{\phi_2(\lambda^{(l)})}{\phi_2'(\lambda^{(l)})} = \left(\frac{1}{\Delta} - \frac{1}{\|p(\lambda^{(l)})\|}\right)\left(\frac{\|p_l\|^3}{\|q_l\|^2}\right) = \left(\frac{\|p_l - \Delta\|}{\Delta\|p_l\|}\right)\left(\frac{\|p_l\|^3}{\|q_l\|^2}\right) = \left(\frac{\|p_l\|}{\|q_l\|}\right)^2\left(\frac{\|p_l\| - \Delta}{\Delta}\right)$$

Therefore, the two equations above are equivalent. □