Numerical Optimization with applications: Homework 05

104021601 林俊傑
104021602 吳彥儒
104021615 黃翊軒

November 23, 2016

Exercise 6. *The square root of a matrix A is a matrix $A^{1/2}$ such that $A^{1/2}A^{1/2} = A$. Show that any symmetric positive definite matrix A has a square root, and that this square root is itself symmetric and positive definite.(Hint: factorization $A = UDU^T$ (A.16), where U is orthogonal and D is diagonal with positive diagonal elements.)*

Proof. First, we show that a real symmetric matrix A is diagonalizable. Prove it by contradiction, which means there is a generalized eigenvector v of order 2, that is $(A - \lambda I)v \neq 0$ and $(A - \lambda I)^2 = 0$, and we have the following statement.

$$\begin{aligned} 0 &= v^T(A - \lambda I)^2 v = v^T(A - \lambda I)^T(A - \lambda I)v \\ &= \|(A - \lambda I)v\|^2 \neq 0 \rightarrow\leftarrow \end{aligned}$$

Thus, every eigenvector of A is of order 1 and A is diagonalizable. We may Assume $A = UDU^T$, where U is orthogonal and by $A > 0$, $D = diag(\lambda_1, \lambda_2, ..., \lambda_n)$ is diagonal with positive diagonal elements. Define the square root of A

$$A^{1/2} := U\sqrt{D}U^T = U diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, ... \sqrt{\lambda_n})U^T$$

Obviously, $A^{1/2}$ is symmetric, and positive number $\sqrt{\lambda_i}$ is the eigenvalue correspond to the ith column vector of U. Hence $A^{1/2}$ is also positive definite.

□

Exercise 10. *(a) Show that $\det(I + xy^T) = 1 + y^T x$, where x and y are n-vectors.*

(b) Using similar technique to prove that

$$\det(I + xy^T + uv^T) = (1 + y^T x)(1 + v^T u) - (x^T v)(y^T u).$$

(c) Use this relation to establish

$$\det(B_{k+1}) = \det(B_k)\frac{y_k^T s_k}{s_k^T B_k s_k}.$$

Proof. (a) Assuming $x \neq 0$, we can find vectors $q_1, q_2, \cdots, q_{n-1}$ such that the matrix Q defined by

$$Q = [x, q_1, q_2, \cdots, q_{n-1}]$$

is nonsingular and $x = Qe_1$.If we define

$$y^T Q = (w_1, w_2, \cdots, w_n)$$

then

$$w_1 = y^T Q e_1 = y^T x$$

and

$$\det(I + xy^T) = \det(Q^{-1}(I + xy^T)Q) = \det(I + Q^{-1}xy^TQ) = \det(I + e_1 y^T Q)$$

$$= \det\left(I + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} (w_1, w_2, \cdots, w_n)\right) = \det\left(\begin{bmatrix} 1 + w_1 & w_2 & \cdots & w_{n-1} & w_n \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}\right)$$

$$= \det\left(\begin{bmatrix} 1+w_1 & w_2 & \cdots & w_{n-1} \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}\right) = \cdots = \det\left(\begin{bmatrix} 1+w_1 & w_2 \\ 0 & 1 \end{bmatrix}\right)$$
$$= 1 + w_1 = 1 + y^T x$$

(b) Assuming $x, u \neq 0$, we can find vectors $q_1, q_2, \cdots, q_{n-2}$ such that the matrix Q defined by

$$Q = [x, u, q_1, q_2, \cdots, q_{n-2}]$$

is nonsingular and $x = Qe_1$, $u = Qe_2$.If we define

$$y^T Q = (w_1, w_2, \cdots, w_n)$$
$$v^T Q = (z_1, z_2, \cdots, z_n)$$

then

$$w_1 = y^T Q e_1 = y^T x \qquad w_2 = y^T Q e_2 = y^T u$$
$$z_1 = v^T Q e_1 = v^T x \qquad z_2 = v^T Q e_2 = v^T u$$

and

$$\det(I + xy^T + uv^T) = \det\left(I + [x \quad u]\begin{bmatrix} y^T \\ v^T \end{bmatrix}\right) = \det\left(Q^{-1}(I + [x\ u]\begin{bmatrix} y^T \\ v^T \end{bmatrix})Q\right)$$
$$= \det\left(I + [Q^{-1}x \quad Q^{-1}u]\begin{bmatrix} y^T Q \\ v^T Q \end{bmatrix}\right) = \det\left(I + [e_1 \quad e_2]\begin{bmatrix} w_1 & w_2 & \cdots & w_n \\ z_1 & z_2 & \cdots & z_n \end{bmatrix}\right)$$
$$= \det\left(I + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}\begin{bmatrix} w_1 & w_2 & \cdots & w_n \\ z_1 & z_2 & \cdots & z_n \end{bmatrix}\right) = \det\left(\begin{bmatrix} 1+w_1 & w_2 & \cdots & w_{n-1} & w_n \\ z_1 & 1+z_2 & \cdots & z_{n-1} & z_n \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}\right)$$
$$= \det\left(\begin{bmatrix} 1+w_1 & w_2 & \cdots & w_{n-1} \\ z_1 & 1+z_2 & \cdots & z_{n-1} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}\right) = \cdots = \det\left(\begin{bmatrix} 1+w_1 & w_2 \\ z_1 & 1+z_2 \end{bmatrix}\right)$$
$$= (1+w_1)(1+z_2) - z_1 w_2 = (1 + y^T x)(1 + v^T u) - (x^T v)(y^T u)$$

(c) We have $B_{k+1} = B_k - \dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \dfrac{y_k y_k^T}{y_k^T s_k}$. So,

$$\det(B_{k+1}) = \det(B_k)\det\left(I + \left(\frac{-s_k}{s_k^T B_k s_k}\right)(s_k^T B_k)\left(\frac{B_k^{-1} y_k}{y_k^T s_k}\right)(y_k^T)\right)$$

Let

$$x = \left(\frac{-s_k}{s_k^T B_k s_k}\right) \quad y^T = (s_k^T B_k) \quad u = \left(\frac{B_k^{-1} y_k}{y_k^T s_k}\right) \quad v^T = (y_k^T)$$

Using (b), we can caculate

$$\det\left(I + \left(\frac{-s_k}{s_k^T B_k s_k}\right)(s_k^T B_k) + \left(\frac{B_k^{-1} y_k}{y_k^T s_k}\right)(y_k^T)\right)$$
$$= \left[1 + (s_k^T B_k)\left(\frac{-s_k}{s_k^T B_k s_k}\right)\right]\left[1 + (y_k^T)\left(\frac{B_k^{-1} y_k}{y_k^T s_k}\right)\right] - \left[(y_k^T)\left(\frac{-s_k}{s_k^T B_k s_k}\right)\right]\left[(s_k^T B_k)\left(\frac{B_k^{-1} y_k}{y_k^T s_k}\right)\right]$$
$$= 0 \times \left[1 + (y_k^T)\left(\frac{B_k^{-1} y_k}{y_k^T s_k}\right)\right] - \left[\frac{-y_k^T s_k}{s_k^T B_k s_k}\right] \times 1 = \frac{y_k^T s_k}{s_k^T B_k s_k}$$

We conclude that

$$\det(B_{k+1}) = \det(B_k)\frac{y_k^T s_k}{s_k^T B_k s_k}$$

□

Exercise 12. *Show that if f satisfies Assumption 6.1 and if the sequence of gradients satisfies* $\liminf \|\nabla f_k\| = 0$, *, then the whole sequence of iterates x converges to the solution x^*.*

Proof. Since $f(x_k)$ deceases at each step and by Assumption 6.1(ii) the convexity of the set $\mathcal{L} = \{x|f(x) \leq f(x_0)\}$, the fact $\liminf \|\nabla f_k\| = 0$ implies there exists a subsequence $\{x_{n_j}\}$ converges to the unique minimizer x^*. We are now proving the whole sequence $\{x_k\}$ converges to x^*. By Taylor's thm, for all $x \in \mathbb{R}^n$ we have

$$f(x) = f(x^* + (x - x^*)) = f(x^*) + \nabla f(x^*)^T(x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(\xi)(x - x^*)$$

If x belongs to $\mathcal{L} = \{x|f(x) \leq f(x_0)\}$ and satisfies

$$f(x) \leq f(x^*) + \varepsilon$$

for some given $\varepsilon > 0$, we obtain following by using the fact $\nabla f(x^*) = 0$

$$\frac{1}{2}(x - x^*)^T \nabla^2 f(\xi)(x - x^*) \leq \varepsilon.$$

By Assumption 6.1(ii) again, we conclude that

$$m\|x - x^*\|_2^2 \leq (x - x^*)^T \nabla^2 f(\xi)(x - x^*) \leq 2\varepsilon.$$

So,

$$\|x - x^*\|_2^2 \leq (2\varepsilon/m)$$

On the other hand, the whole sequence $\{f(x_k)\}$ is nonincreasing by any descent direction Algorithm, and we already know that there exists a subsequence $\{f(x_{n_j})\}$ converges to the $f(x^*)$. So given $\varepsilon > 0$, we can find a $N \in \mathbb{N}$ such that

$$f(x_k) \leq f(x_{n_j}) \leq f(x*) + \varepsilon$$

for all $k \geq n_j \geq N$. Hence, combining the two inequality gives

$$\|x_k - x^*\|_2^2 \leq (2\varepsilon/m)$$

for for all $k \geq N$. So the whole sequence $\{x_k\}$ converges to x^*. □