CS5321 Numerical Optimization Midterm

April 19, 15:20-17:20

Note: Since all problems are in sets, to avoid your wrong answers in the previous problems affecting the latter ones, write down your derivations. Also, if you have any questions about the problems, write them down.

1. Consider a function $f(x_1, x_2) = x_1^3 x_2 - 2x_1^2 x_2^2 + x_1 x_2^3$.

   (a) (5pt) Compute the gradient of f.

$$\nabla f(x_1, x_2) = \begin{pmatrix} 3x_1^2 x_2 - 4x_1 x_2^2 + x_2^3 \\ x_1^3 - 4x_1^2 x_2 + 3x_1 x_2^2 \end{pmatrix}$$

   (b) (5pt) Compute the Hessian of f.

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 6x_1 x_2 - 4x_2^2 & 3x_1^2 - 8x_1 x_2 + 3x_2^2 \\ 3x_1^2 - 8x_1 x_2 + 3x_2^2 & -4x_1^2 + 6x_1 x_2 \end{pmatrix}$$

   (c) (5pt) Is $(x_1, x_2) = (1, 1)$ a local minimizer? Justify your answer?

$$\nabla f(1,1) = \begin{pmatrix} 3-4+1 \\ 1-4+3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

   and

$$\nabla^2 f(1,1) = \begin{pmatrix} 6-4 & 3-8+3 \\ 3-8+3 & -4+6 \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

   has eigenvalue $2, 0$, which makes it positive semidefinite, so we cannot make sure if $(1, 1)$ is a local minimizer.

   (d) (5pt) What is the steepest descent direction of f at $(x_1, x_2) = (1, 2)$?

$$\vec{p} = -\nabla f(1,2) = -\begin{pmatrix} 3*2-4*4+2^3 \\ 1^3-4*2+3*2^2 \end{pmatrix} = \begin{pmatrix} 2 \\ -5 \end{pmatrix}$$

(e) (5pt) Compute the LDL decomposition of the Hessian of f at $(x_1, x_2) = (1, 2)$. (No pivoting)

$$\nabla^2 f(1,2) = \begin{pmatrix} 6*2-4*4 & 3-8*2+3*4 \\ 3-8*2+3*4 & -4+6*2 \end{pmatrix} = \begin{pmatrix} -4 & -1 \\ -1 & 8 \end{pmatrix}$$

$$L = \begin{pmatrix} 1 & 0 \\ .25 & 1 \end{pmatrix} D = \begin{pmatrix} -4 & 0 \\ 0 & 8.25 \end{pmatrix}$$

(f) (5pt) What is the Newton's direction of f at $(x_1, x_2) = (1, 2)$?

$$\vec{p} = -(\nabla^2 f)^{-1}\nabla f = -(L^T)^{-1}D^{-1}L^{-1}\nabla f = \begin{pmatrix} -1/3 \\ -2/3 \end{pmatrix}$$

(g) (5pt) Is the Newton's direction of f at $(x_1, x_2) = (1, 2)$ a *descent direction*? Justify your answer.

$$\vec{p}^T \nabla f = -2*(-1/3) + 5*(-2/3) = -8/3 < 0$$

$\vec{p}$ is a descent direction.

(h) (5pt) Modify the LDL decomposition computed in (d) such that all diagonal elements of D is larger than or equal to 1, and use the modified LDL decomposition to compute a modified Newton's direction at $(x_1, x_2) = (1, 2)$.

Let

$$\hat{D} = \begin{pmatrix} 1 & 0 \\ 0 & 8.25 \end{pmatrix}.$$

$$\vec{p} = -(L^T)^{-1}\hat{D}^{-1}L^{-1}\nabla f = \begin{pmatrix} 13/6 \\ -2/3 \end{pmatrix}$$

2. (Line search method) Suppose $\phi(\alpha) = f(\vec{x}_k + \alpha\vec{p}_k) = (\alpha - 1)^2$.

(a) (10pt) The sufficient decrease condition asks $\phi(\alpha) \le \phi(0) + c_1\alpha\phi'(0)$. Suppose $c_1 = 0.1$. What is the feasible interval of α satisfying this condition? Note that $\alpha \in [0, \infty)$.

$$\phi'(\alpha) = 2(\alpha - 1), \text{ and } \phi'(0) = 2(0-1) = -2.$$

The sufficient decrease condition needs

$$(\alpha - 1)^2 \le (0-1)^2 + 0.1 * \alpha * (-2) = 1 - 0.2\alpha.$$

Solving the inequality, $0 \le \alpha \le 1.8$.

(b) (10pt) The curvature condition asks $\phi'(\alpha) \ge c_2\phi'(0)$. Suppose $c_2 = 0.9$. What is the feasible interval of α satisfying this condition?

$$\phi'(\alpha) = 2(\alpha - 1), \text{ and } \phi'(0) = 2(0-1) = -2.$$

The curvature condition needs

$$2(\alpha - 1) \ge 0.9 * (-2) = -1.8.$$

Solving the inequality, $\alpha \ge 0.1$.

3. (Quasi-Newton method) The BSGS update formula for approximating the inverse of Hessian matrix is

$$B_{k+1} = (I - \rho_k \vec{s}_k \vec{y}_k^T) B_k (I - \rho_k \vec{y}_k \vec{s}_k^T) + \rho_k \vec{s}_k \vec{s}_k^T,$$

where $\rho = 1/(\vec{y}_k^T s_k)$.

(a) (10pt) Prove that B_{k+1} satisfies the secant equation $B_{k+1}\vec{y}_k = \vec{s}_k$.

$$\begin{aligned} B_{k+1}\vec{y}_k &= (I - \rho_k \vec{s}_k \vec{y}_k^T) B_k (I - \rho_k \vec{y}_k \vec{s}_k^T)\vec{y}_k + \rho_k \vec{s}_k \vec{s}_k^T \vec{y}_k \\ &= (I - \rho_k \vec{s}_k \vec{y}_k^T) B_k (\vec{y}_k - \rho_k \vec{y}_k \rho_k^{-1}) + \rho_k \vec{s}_k \rho_k^{-1} \\ &= (I - \rho_k \vec{s}_k \vec{y}_k^T) B_k \vec{0} + \vec{s}_k = \vec{s}_k \end{aligned}$$

(b) (10pt) Prove that B_{k+1} is positive definite if B_k is positive definite and $\vec{y}_k^T B_{k+1}\vec{y}_k > 0.$.

For any vector $\vec{x} \in \mathbb{R}^n$,

$$\vec{x}^T B_{k+1}\vec{x} = \vec{x}^T (I - \rho_k \vec{s}_k \vec{y}_k^T) B_k (I - \rho_k \vec{y}_k \vec{s}_k^T)\vec{x} + \rho_k \vec{x}^T \vec{s}_k \vec{s}_k^T \vec{x}$$

Because B_k is positive definite, $\vec{x}^T (I - \rho_k \vec{s}_k \vec{y}_k^T) B_k (I - \rho_k \vec{y}_k \vec{s}_k^T)\vec{x} > 0$. Also $\vec{x}^T \vec{s}_k \vec{s}_k^T \vec{x} = (\vec{x}^T \vec{s}_k)^2 > 0$. By using the result of (a), $\vec{y}_k^T B_{k+1}\vec{y}_k = \vec{y}_k^T \vec{s}_k = \rho_k > 0$, which implies that $\vec{x}^T B_{k+1}\vec{x} > 0$. Therefore, B_{k+1} is positive definite.

4. (CG method) Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}, \vec{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \vec{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

   (a) (10pt) What are the restrictions and advantages of using the conjugate gradient method to solve $A\vec{x} = \vec{b}$?

   Restriction: A need be spd.

   Advantage: Only need one matrix-vector multiplication per iteration.

   (b) (10pt) Find the scalar β that makes two vectors, $\vec{x}$ and $(\vec{x} - \beta\vec{y})$, A-conjugate.

   Want to make $\vec{x}^T A(\vec{x} - \beta\vec{y}) = 0$.

$$\beta = \frac{\vec{x}^T A\vec{x}}{\vec{x}^T A\vec{y}} = 10.$$

5. (10pt) In many proofs, we need the function f or its derivative to be continuous. Give an example to explain why this property is important.