CS726, Fall 2006
Final Examination
Tuesday, December 19, 2006, 10:05am-12:05pm

Answer all FOUR questions below. One handwritten sheet of notes (written front and back) is allowed. Explain all your answers.

1. The DFP quasi-Newton updating formula for the approximate Hessian B_k can be written as follows:

$$B_{k+1} = (I - \rho_k y_k s_k^T) B_k (I - \rho_k s_k y_k^T) + \rho_k y_k y_k^T,$$

where

$$\rho_k = \frac{1}{y_k^T s_k}.$$

Show that if B_k is positive definite and the curvature condition $y_k^T s_k > 0$ holds, then B_{k+1} is also positive definite.

Solution: Defining $E_k = I - \rho_k y_k s_k^T$, the update formula can be rewritten as

$$B_{k+1} = E_k B_k E_k^T + \rho_k y_k y_k^T.$$

For any $z \in \mathbf{R}^n$ we have

$$z^T B_{k+1} z = z^T E_k B_k E_k^T z + \rho_k z^T y_k y_k^T z = (E_k^T z)^T B_k (E_k^T z) + \rho \| y_k^T z \|_2^2.$$

The first term is nonnegative (by postive definiteness of B_k) while the second term is also nonnegative, since $\rho_k > 0$ by the curvature condition. Hence, B_{k+1} is at least positive semidefinite.
If $z^T B_{k+1} z = 0$ for some nonzero z, we must have $y_k^T z = 0$ (from the second term). But then $E_k^T z = z$ so we have $z^T B_{k+1} z = z^T B_k z > 0$, a contradiction. Hence $z^T B_{k+1} z > 0$ for all nonzero z and so B_{k+1} is positive definite.

---

2. (a) Consider the problem

$$\min_x f(x) \text{ subject to } l \le x \le u,$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, and the lower- and upper-bound vectors l and u are also in $\mathbb{R}^n$. Write down the first-order necessary (KKT) conditions for this problem, eliminating Lagrange multipliers to obtain a simplified form.

Solution: Introducing Lagrange multipliers λ and μ for $x - l \ge 0$ and $u - x \ge 0$, resp., we can write the Lagrangian as

$$\mathcal{L}(x, \lambda, \mu) = f(x) - \lambda^T (x - l) - \mu^T (u - x).$$

The KKT conditions are then:

$$\begin{aligned}
\nabla_x \mathcal{L}(x, \lambda, \mu) = \nabla f(x) - \lambda + \mu = 0, \\
x - l \ge 0, \quad u - x \ge 0, \\
\lambda \ge 0, \quad \mu \ge 0, \\
\lambda^T (x - l) = 0, \quad \mu^T (u - x) = 0.
\end{aligned}$$

Assuming that $l < u$, we have from the complementarity conditions that at most one of λ_i and μ_i is nonzero for each i. We consider three cases:

(i) $l_i < x_i < u_i$: Then $\lambda_i = \mu_i = 0$ and $\frac{\partial f}{\partial x_i} = 0$;

(ii) $x_i = l_i$: Then $\mu_i = 0$ and $\lambda_i \ge 0$, and we have by substituting into the first condition that $\frac{\partial f}{\partial x_i} \ge 0$;

(iii) $x_i = u_i$: Then $\lambda_i = 0$ and $\mu_i \ge 0$, and we have by substituting into the first condition that $\frac{\partial f}{\partial x_i} \le 0$.

(b) Consider the problem

$$\begin{aligned}
\min_{(x_1, x_2) \in \mathbb{R}^2} \quad & \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 - x_1 x_2 - 2x_2 \\
\text{subject to} \quad & 1 - x_1 - x_2 \ge 0, \\
& 1 + x_1 - x_2 \ge 0.
\end{aligned}$$

Show that the KKT conditions are satisfied at $x^* = (0, 1)$, and determine the optimal values of the Lagrange multipliers for the two constraints.

Solution: We have

$$\nabla f(x) = \begin{bmatrix} x_1 - x_2 \\ x_2 - x_1 - 2 \end{bmatrix}, \quad \nabla c_1(x) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla c_1(x) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

At $x^* = (0, 1)^T$, both constraints are active. We need $\lambda_1^* \ge 0$ and $\lambda_2^* \ge 0$ such that

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \lambda_1^* \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \lambda_2^* \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

and it is clear by inspection that $\lambda_1^* = 1$, $\lambda_2^* = 0$ is the unique solution of this system.

---

3. Consider the determination of a quadratic function of two variables using function value information. That is, we seek the values of the scalars a_{11}, a_{12}, a_{22}, b_1, b_2, and c such that for the model function $m(x)$ defined by

$$m(x) = \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + c,$$

we have $m(y^i) = f(y^i)$, for the chosen values y^i, $i = 1, 2, 3, 4, 5, 6$ and the given function f. Show that if the points y^i all lie on an ellipse satisfying

$$(y^i)^T W y^i = \gamma, \quad i = 1, 2, 3, 4, 5, 6,$$

where W is a 2×2 positive definite matrix and γ is a positive scalar, then the quadratic m is not in general determined by the six points y^i, $i = 1, 2, 3, 4, 5, 6$.
Solution:
From the equations $m(y^i) = f(y^i)$ we obtain a 6×6 linear system in which the unknown vector is

$$(a_{11}, a_{12}, a_{22}, b_1, b_2, c)^T$$

and the ith row of the coefficient matrix is

$$\left(\frac{1}{2}(y_1^i)^2, y_1^i y_i^2, \frac{1}{2}(y_2^i)^2, y_1^i, y_2^i, 1) \right).$$

Since the y^i all lie on the ellipse indicated, there are scalars w_{11}, $w12$, w_{22}, and γ such that

$$w_{11}(y_1^i)^2 + 2w_{12} y_1^i y_2^i + w_{22}(y_2^i)^2 - \gamma = 0.$$

Hence, if we multiply the coefficient matrix by the vector

$$(2w_{11}, 2w_{12}, 2w_{22}, 0, 0, -\gamma),$$

we get zero. Hence, this 6×6 matrix is nonsingular, so m is not determined in general.

---

4. (a) Suppose that an algorithm for minimizing the continuously differentiable function f generates a sequence $\{x_k\}$ lying in a bounded set $\mathcal{B}$, such that the sequence of gradient norms $\{\|\nabla f(x_k)\|\}$ has an accumulation point at zero. Show that there exists an accumulation point x_∞ of $\{x_k\}$ such that $\nabla f(x_\infty) = 0$.

Solution:
Let x_{k_j}, $j = 1, 2, \ldots$ be the subsequence for which $\lim_{j\to\infty} \nabla f(x_{k_j}) = 0$. Since x_{k_j} is in the bounded set $\mathcal{B}$ for all j, it has an accumulation point. We have by taking a further subsequence if necessary that $\lim_{j\to\infty} x_{k_j} = x_\infty$ for some x_∞ (which lies in the closure of $\mathcal{B}$, incidentally). By continuity of ∇f we have that $\nabla f(x_\infty) = 0$.

(b) Suppose that an algorithm for minimizing the twice continuously differentiable function f generates a sequence $\{x_k\}$ for which

$$\lim_{k\to\infty} \nabla f(x_k) = 0$$

and

$$\nabla^2 f(x_k) \text{ are positive definite for all } k.$$

Show that all accumulation points of $\{x_k\}$ satisfy second-order necessary conditions to be a minimizer of f.

Solution: If x_∞ is any accumulation point, there is a subsequence $\{x_{k_j}\}_{j=1,2,3,\ldots}$ such that

$$\lim_{j\to\infty} x_{k_j} = x_\infty.$$

By the smoothness assumption we have $\nabla f(x_\infty) = 0$ and $\nabla^2 f(x_\infty)$ positive semidefinite, so the second-order necessary conditions for x_∞ to be a minimizer of f are satisfied.

(c) Can we claim that all accumulation points for the sequence in part (b) satisfy second-order *sufficient* conditions? Why or why not?

Solution: We cannot argue in part (b) that the limit of a sequence of positive definite matrice is positive definite — only positive semidefinite, as some of the eigenvalues may be approaching zero.
