ECE/MATH 520, Spring 2008
Exam 1: Due Session 15

Solutions (version: March 25, 2008, 11:21) 75 mins.; Total 50 pts.

1. (10 pts.) Consider the problem

$$\begin{array}{ll} \text{maximize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega, \end{array}$$

where $\Omega \subset \{\boldsymbol{x} \in \mathbb{R}^2 : x_1 > 0, x_2 > 0\}$ and $f : \Omega \to \mathbb{R}$ is given by $f(\boldsymbol{x}) = \log(x_1) + \log(x_2)$ with $\boldsymbol{x} = [x_1, x_2]^\top$, where "$\log$" represents natural logarithm. Suppose that $\boldsymbol{x}^*$ is an optimal solution. Answer each of the following questions, showing complete justification.

a. Is it possible that $\boldsymbol{x}^*$ is an interior point of Ω?

b. At what point(s) (if any) is the second-order necessary condition satisfied?

Ans.: a. We have $\nabla f(\boldsymbol{x}^*) = [1/x_1^*, 1/x_2^*]^T$. If $\boldsymbol{x}^*$ were an interior point, then $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$. But this is clearly impossible. Therefore, $\boldsymbol{x}^*$ cannot possibly be an interior point.

b. We have $\boldsymbol{F}(\boldsymbol{x}) = -\,\text{diag}[1/x_1^2, 1/x_2^2]$, which is negative definite everywhere. Therefore, the second-order necessary condition is satisfied everywhere. (Note that because we have a maximization problem, *negative* definiteness is the relevant condition.)

2. (10 pts.) Derive a one-dimensional search (minimization) algorithm based on quadratic fit with only objective function values. Specifically, derive an algorithm that computes $x^{(k+1)}$ based on $x^{(k)}$, $x^{(k-1)}$, $x^{(k-2)}$, $f(x^{(k)})$, $f(x^{(k-1)})$, and $f(x^{(k-2)})$.
Hint: To simplify, use the notation $\sigma_{ij} = (x^{(k-i)})^2 - (x^{(k-j)})^2$ and $\delta_{ij} = x^{(k-i)} - x^{(k-j)}$. You might also find it useful to experiment with your algorithm by writing a MATLAB program (this is optional). Note that three points are needed to initialize the algorithm.

Ans.: The quadratic function that matches the given data $x^{(k)}$, $x^{(k-1)}$, $x^{(k-2)}$, $f(x^{(k)})$, $f(x^{(k-1)})$, and $f(x^{(k-2)})$ can be computed by solving the following three linear equations for the parameters a, b, and c:

$$a(x^{(k-i)})^2 + bx^{(k-i)} + c = f(x^{(k-i)}), \quad i = 0, 1, 2.$$

Then, the algorithm is given by $x^{(k+1)} = -b/2a$ (so, in fact, we only need to find the *ratio* of a and b). With some elementary algebra (e.g., using Cramer's rule without needing to calculate the determinant in the denominator), the algorithm can be written as:

$$x^{(k+1)} = \frac{\sigma_{12} f(x^{(k)}) + \sigma_{20} f(x^{(k-1)}) + \sigma_{01} f(x^{(k-2)})}{2(\delta_{12} f(x^{(k)}) + \delta_{20} f(x^{(k-1)}) + \delta_{01} f(x^{(k-2)}))}$$

where $\sigma_{ij} = (x^{(k-i)})^2 - (x^{(k-j)})^2$ and $\delta_{ij} = x^{(k-i)} - x^{(k-j)}$.

3. (20 pts.) Consider the two sequences $\{\boldsymbol{x}^{(k)}\}$ and $\{\boldsymbol{y}^{(k)}\}$ defined iteratively as follows:

$$\begin{aligned}\boldsymbol{x}^{(k+1)} &= a\boldsymbol{x}^{(k)}\\ \boldsymbol{y}^{(k+1)} &= (\boldsymbol{y}^{(k)})^b\end{aligned}$$

where $a \in \mathbb{R}$, $b \in \mathbb{R}$, $0 < a < 1$, $b > 1$, $\boldsymbol{x}^{(0)} \neq \mathbf{0}$, $\boldsymbol{y}^{(0)} \neq \mathbf{0}$, and $|\boldsymbol{y}^{(0)}| < 1$.

a. Derive a formula for $\boldsymbol{x}^{(k)}$ in terms of $\boldsymbol{x}^{(0)}$ and a. Use this to deduce that $\boldsymbol{x}^{(k)} \to 0$.

b. Derive a formula for $\boldsymbol{y}^{(k)}$ in terms of $\boldsymbol{y}^{(0)}$ and b. Use this to deduce that $\boldsymbol{y}^{(k)} \to 0$.

c. Find the order of convergence of $\{\boldsymbol{x}^{(k)}\}$ and the order of convergence of $\{\boldsymbol{y}^{(k)}\}$.

d. Calculate the smallest number of iterations k such that $|\boldsymbol{x}^{(k)}| \leq c|\boldsymbol{x}^{(0)}|$, where $0 < c < 1$. *Hint:* The answer is in terms of a and c. You may use the notation $\lceil z \rceil$ to represent the smallest integer not smaller than z.

e. Calculate the smallest number of iterations k such that $|\boldsymbol{y}^{(k)}| \leq c|\boldsymbol{y}^{(0)}|$, where $0 < c < 1$.

f. Compare the answer of part e with that of part d, focusing on the case where c is very small.

Ans.: a. We have

$$\begin{aligned}\boldsymbol{x}^{(k)} &= a\boldsymbol{x}^{(k-1)}\\ &= a \cdot a\boldsymbol{x}^{(k-2)}\\ &= a^2\boldsymbol{x}^{(k-2)}\\ &\vdots\\ &= a^k\boldsymbol{x}^{(0)}.\end{aligned}$$

Because $0 < a < 1$, we have $a^k \to 0$, and hence $\boldsymbol{x}^{(k)} \to 0$.

b. Similarly, we have

$$\begin{aligned}\boldsymbol{y}^{(k)} &= (\boldsymbol{y}^{(k-1)})^b\\ &= ((\boldsymbol{y}^{(k-2)})^b)^b\\ &= (\boldsymbol{y}^{(k-2)})^{b^2}\\ &\vdots\\ &= (\boldsymbol{y}^{(0)})^{b^k}.\end{aligned}$$

Because $|\boldsymbol{y}^{(0)}| < 1$ and $b > 1$, we have $b^k \to \infty$ and hence $\boldsymbol{y}^{(k)} \to 0$.

c. The order of convergence of $\{\boldsymbol{x}^{(k)}\}$ is 1 because

$$\lim_{k\to\infty} \frac{|\boldsymbol{x}^{(k+1)}|}{|\boldsymbol{x}^{(k)}|} = \lim_{k\to\infty} a = a,$$

and $0 < a < \infty$.

The order of convergence of $\{\boldsymbol{y}^{(k)}\}$ is b because

$$\lim_{k\to\infty} \frac{|\boldsymbol{y}^{(k+1)}|}{|\boldsymbol{y}^{(k)}|^b} = \lim_{k\to\infty} 1 = 1,$$

and $0 < 1 < \infty$.

d. Suppose $|\boldsymbol{x}^{(k)}| \le c|\boldsymbol{x}^{(0)}|$. Using part a, we have $a^k \le c$, which implies that $k \ge \log(1/c - 1/a)$. So the smallest number of iterations k such that $|\boldsymbol{x}^{(k)}| \le c|\boldsymbol{x}^{(0)}|$ is $\lceil \log(1/c - 1/a) \rceil$ (the smallest integer not smaller than $\log(1/c - 1/a)$).

e. Suppose $|\boldsymbol{y}^{(k)}| \le c|\boldsymbol{y}^{(0)}|$. Using part b, we have $|\boldsymbol{y}^{(0)}|^{b^k} \le c|\boldsymbol{y}^{(0)}|$. Taking logs (twice) and rearranging, we have

$$k \ge \frac{1}{\log(b)} \log\left(1 + \log\left[\frac{1}{c} - \frac{1}{|\boldsymbol{y}^{(0)}|}\right]\right).$$

Denote the right-hand side by z. So the smallest number of iterations k such that $|\boldsymbol{y}^{(k)}| \le c|\boldsymbol{y}^{(0)}|$ is $\lceil z \rceil$.

f. Comparing the answer in part e with that of part d, we can see that as $c \to 0$, the answer in part d is $\Omega(\log(1/c))$, whereas the answer in part e is $O(\log\log(1/c))$. Hence, in the regime where c is very small, the number of iterations in part d (linear convergence) is (at least) exponentially larger than that in part e (superlinear convergence).

4. (10 pts.) Consider the following simple modification of the quasi-Newton family of algorithms. In the quadratic case, instead of the usual quasi-Newton condition $\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \Delta\boldsymbol{x}^{(i)}$, $0 \le i \le k$, $k \le n-1$, suppose there are scalars $\rho_1, \rho_2, \ldots$ such that $\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \rho_i\Delta\boldsymbol{x}^{(i)}$, $0 \le i \le k$, $k \le n-1$.

Is it true that an algorithm satisfying the modified condition above is a conjugate direction algorithm? Justify your answer fully.
Hint: Formulate a precise claim that you can prove.

Ans.: The answer is yes. To show this, we will prove the following precise statement: In the quadratic case (with Hessian $\boldsymbol{Q}$), suppose that $\boldsymbol{H}_{k+1}\Delta\boldsymbol{g}^{(i)} = \rho_i\Delta\boldsymbol{x}^{(i)}$, $0 \le i \le k$, $k \le n-1$. If $\alpha_i \ne 0$, $0 \le i \le k$, then $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k+1)}$ are $\boldsymbol{Q}$-conjugate.

We proceed by induction. We begin with the $k = 0$ case: that $\boldsymbol{d}^{(0)}$ and $\boldsymbol{d}^{(1)}$ are $\boldsymbol{Q}$-conjugate. Because $\alpha_0 \ne 0$, we can write $\boldsymbol{d}^{(0)} = \Delta\boldsymbol{x}^{(0)}/\alpha_0$. Hence,

$$\begin{aligned}
\boldsymbol{d}^{(1)\top}\boldsymbol{Q}\boldsymbol{d}^{(0)} &= -\boldsymbol{g}^{(1)\top}\boldsymbol{H}_1\boldsymbol{Q}\boldsymbol{d}^{(0)} \\
&= -\boldsymbol{g}^{(1)\top}\boldsymbol{H}_1\frac{\boldsymbol{Q}\Delta\boldsymbol{x}^{(0)}}{\alpha_0} \\
&= -\boldsymbol{g}^{(1)\top}\frac{\boldsymbol{H}_1\Delta\boldsymbol{g}^{(0)}}{\alpha_0} \\
&= -\boldsymbol{g}^{(1)\top}\frac{\rho_0\Delta\boldsymbol{x}^{(0)}}{\alpha_0} \\
&= -\rho_0\boldsymbol{g}^{(1)\top}\boldsymbol{d}^{(0)}.
\end{aligned}$$

But $\boldsymbol{g}^{(1)\top}\boldsymbol{d}^{(0)} = 0$ as a consequence of $\alpha_0 > 0$ being the minimizer of $\phi(\alpha) = f(\boldsymbol{x}^{(0)} + \alpha\boldsymbol{d}^{(0)})$. Hence, $\boldsymbol{d}^{(1)\top}\boldsymbol{Q}\boldsymbol{d}^{(0)} = 0$.

Assume that the result is true for $k-1$ (where $k < n-1$). We now prove the result for k, that is, that $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k+1)}$ are $\boldsymbol{Q}$-conjugate. It suffices to show that $\boldsymbol{d}^{(k+1)\top}\boldsymbol{Q}\boldsymbol{d}^{(i)} = 0$, $0 \leq i \leq k$. Given i, $0 \leq i \leq k$, using the same algebraic steps as in the $k = 0$ case, and using the assumption that $\alpha_i \neq 0$, we obtain

$$\begin{aligned}
\boldsymbol{d}^{(k+1)\top}\boldsymbol{Q}\boldsymbol{d}^{(i)} &= -\boldsymbol{g}^{(k+1)\top}\boldsymbol{H}_{k+1}\boldsymbol{Q}\boldsymbol{d}^{(i)} \\
&\;\;\vdots \\
&= -\rho_i\boldsymbol{g}^{(k+1)\top}\boldsymbol{d}^{(i)}.
\end{aligned}$$

Because $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(k)}$ are $\boldsymbol{Q}$-conjugate by assumption, we conclude from the expanding subspace lemma (Lemma 10.2) that $\boldsymbol{g}^{(k+1)\top}\boldsymbol{d}^{(i)} = 0$. Hence, $\boldsymbol{d}^{(k+1)\top}\boldsymbol{Q}\boldsymbol{d}^{(i)} = 0$, which completes the proof.