ECE/MATH 520, Spring 2008

Homework Problems 2

Solutions (version: February 19, 2008, 12:9)

7.2a,b,d Let $f(x) = x^2 + 4\cos x$, $x \in \mathbb{R}$. We wish to find the minimizer x^* of f over the interval $[1, 2]$. (Calculator users: Note that in $\cos x$, x is in rad).

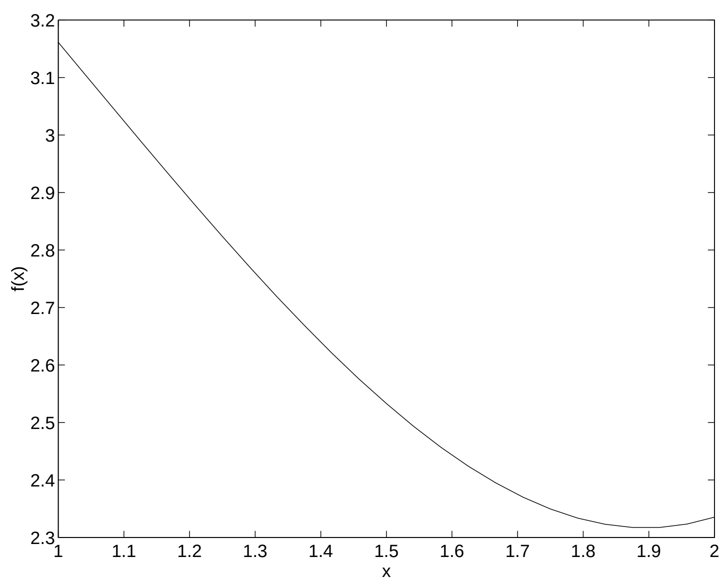
a. Plot $f(x)$ versus x over the interval $[1, 2]$.

b. Use the Golden Section method to locate x^* to within an uncertainty of 0.2. Display all intermediate steps using a table as follows:

Iteration k	a_k	b_k	$f(a_k)$	$f(b_k)$	New uncertainty interval
1	?	?	?	?	[?,?]
2	?	?	?	?	[?,?]
⋮	⋮	⋮	⋮	⋮	⋮

d. Apply Newton's method, using the same number of iterations as in part b, with $x^{(0)} = 1$.

Ans.: a. The plot of $f(x)$ versus x is as below:



b. The number of steps needed for the Golden Section method is computed from the inequality:

$$0.61803^N \leq \frac{0.2}{2-1} \qquad \Rightarrow \qquad N \geq 3.34.$$

Therefore, the fewest possible number of steps is 4. Applying 4 steps of the Golden Section method, we end up with an uncertainty interval of $[a_4, b_0] = [1.8541, 2.000]$. The table with the results of the intermediate steps is displayed below:

Iteration k	a_k	b_k	$f(a_k)$	$f(b_k)$	New uncertainty interval
1	1.3820	1.6180	2.6607	2.4292	[1.3820,2]
2	1.6180	1.7639	2.4292	2.3437	[1.6180,2]
3	1.7639	1.8541	2.3437	2.3196	[1.7639,2]
4	1.8541	1.9098	2.3196	2.3171	[1.8541,2]

d. We have $f'(x) = 2x - 4\sin x$, $f''(x) = 2 - 4\cos x$. Hence, Newton's algorithm takes the form:

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - 2\sin x^{(k)}}{1 - 2\cos x^{(k)}}.$$

Applying 4 iterations with $x^{(0)} = 1$, we get $x^{(1)} = -7.4727$, $x^{(2)} = 14.4785$, $x^{(3)} = 6.9351$, $x^{(4)} = 16.6354$. Apparently, Newton's method is not effective in this case.

8.1 Let $\{\boldsymbol{x}^{(k)}\}$ be a sequence that converges to $\boldsymbol{x}^*$. Show that if there exists $c > 0$ such that

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\| \geq c\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^p,$$

then the order of convergence (if it exists) is at most p.

Ans.: Let s be the order of convergence of $\{\boldsymbol{x}^{(k)}\}$. Suppose there exists $c > 0$ such that for all k sufficiently large,

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\| \geq c\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^p.$$

Hence, for all k sufficiently large,

$$\begin{aligned} \frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^s} &= \frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^p} \frac{1}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^{s-p}} \\ &\geq \frac{c}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^{s-p}}. \end{aligned}$$

Taking limits yields

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^s} \geq \frac{c}{\lim_{k\to\infty} \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^{s-p}}.$$

Since by definition s is the order of convergence,

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|^s} < \infty.$$

Combining the above two inequalities, we get

$$\frac{c}{\lim_{k\to\infty}\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\|^{s-p}} < \infty.$$

Therefore, since $\lim_{k\to\infty}\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\| = 0$, we conclude that $s \le p$, i.e., the order of convergence is at most p.

8.3 Suppose that we use the Golden Section algorithm to find the minimizer of a function. Let u_k be the uncertainty range at the kth iteration. Find the order of convergence of $\{u_k\}$.

Ans.: We have $u_{k+1} = (1-\rho)u_k$, and $u_k \to 0$. Therefore,

$$\lim_{k\to\infty}\frac{|u_{k+1}|}{|u_k|} = 1-\rho > 0$$

and thus the order of convergence is 1.

8.13 Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$, where $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{Q}$ is a real symmetric positive definite $n \times n$ matrix. Suppose that we apply the steepest descent method to this function, with $\boldsymbol{x}^{(0)} \neq \boldsymbol{Q}^{-1}\boldsymbol{b}$. Show that the method converges in one step, that is, $\boldsymbol{x}^{(1)} = \boldsymbol{Q}^{-1}\boldsymbol{b}$, if and only if $\boldsymbol{x}^{(0)}$ is chosen such that $\boldsymbol{g}^{(0)} = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}$ is an eigenvector of $\boldsymbol{Q}$,

Ans.: The steepest descent algorithm applied to the quadratic function f has the form

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k\boldsymbol{g}^{(k)} = \boldsymbol{x}^{(k)} - \frac{\boldsymbol{g}^{(k)T}\boldsymbol{g}^{(k)}}{\boldsymbol{g}^{(k)T}\boldsymbol{Q}\boldsymbol{g}^{(k)}}\boldsymbol{g}^{(k)}.$$

$\Rightarrow$: If $\boldsymbol{x}^{(1)} = \boldsymbol{Q}^{-1}\boldsymbol{b}$, then

$$\boldsymbol{Q}^{-1}\boldsymbol{b} = \boldsymbol{x}^{(0)} - \alpha_0\boldsymbol{g}^{(0)}.$$

Rearranging the above yields

$$\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b} = \alpha_0\boldsymbol{Q}\boldsymbol{g}^{(0)}.$$

Since $\boldsymbol{g}^{(0)} = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b} \neq \boldsymbol{0}$, we have

$$\boldsymbol{Q}\boldsymbol{g}^{(0)} = \frac{1}{\alpha_0}\boldsymbol{g}^{(0)}$$

which means that $\boldsymbol{g}^{(0)}$ is an eigenvector of $\boldsymbol{Q}$ (with corresponding eigenvalue $1/\alpha_0$).

$\Leftarrow$: By assumption, $\boldsymbol{Q}\boldsymbol{g}^{(0)} = \lambda\boldsymbol{g}^{(0)}$, where $\lambda \in \mathbb{R}$. We want to show that $\boldsymbol{Q}\boldsymbol{x}^{(1)} = \boldsymbol{b}$. We have

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{x}^{(1)} &= \boldsymbol{Q}\left(\boldsymbol{x}^{(0)} - \frac{\boldsymbol{g}^{(0)T}\boldsymbol{g}^{(0)}}{\boldsymbol{g}^{(0)T}\boldsymbol{Q}\boldsymbol{g}^{(0)}}\boldsymbol{g}^{(0)}\right) \\
&= \boldsymbol{Q}\boldsymbol{x}^{(0)} - \frac{1}{\lambda}\frac{\boldsymbol{g}^{(0)T}\boldsymbol{g}^{(0)}}{\boldsymbol{g}^{(0)T}\boldsymbol{g}^{(0)}}\boldsymbol{Q}\boldsymbol{g}^{(0)} \\
&= \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{g}^{(0)} \\
&= \boldsymbol{b}.
\end{aligned}$$

8.17 Given $f : \mathbb{R}^n \to \mathbb{R}$, consider the general iterative algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$$

where $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \ldots$ are given vectors in $\mathbb{R}^n$, and α_k is chosen to minimize $f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$; that is,

$$\alpha_k = \arg\min f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}).$$

Show that for each k, the vector $\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$ is orthogonal to $\nabla f(\boldsymbol{x}^{(k+1)})$.

Ans.: We have

$$\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} = \alpha_k \boldsymbol{d}^{(k)}$$

and hence

$$\left\langle \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}, \nabla f(\boldsymbol{x}^{(k+1)}) \right\rangle = \alpha_k \left\langle \boldsymbol{d}^{(k)}, \nabla f(\boldsymbol{x}^{(k+1)}) \right\rangle.$$

Now, let $\phi_k(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$. Since α_k minimizes ϕ_k, then by the FONC, $\phi_k'(\alpha_k) = 0$. By the chain rule, $\phi_k'(\alpha) = \boldsymbol{d}^{(k)T} \nabla f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$. Hence,

$$0 = \phi_k'(\alpha_k) = \boldsymbol{d}^{(k)T} \nabla f(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}) = \left\langle \boldsymbol{d}^{(k)}, \nabla f(\boldsymbol{x}^{(k+1)}) \right\rangle,$$

and so

$$\left\langle \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}, \nabla f(\boldsymbol{x}^{(k+1)}) \right\rangle = 0.$$

9.1 Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = (x - x_0)^4$, where $x_0 \in \mathbb{R}$ is a constant. Suppose that we apply Newton's method to the problem of minimizing f.

a. Write down the update equation for Newton's method applied to the problem.

b. Let $y^{(k)} = |x^{(k)} - x_0|$. Show that the sequence $\{y^{(k)}\}$ satisfies $y^{(k+1)} = \frac{2}{3} y^{(k)}$.

c. Show that $x^{(k)} \to x_0$ for any initial guess $x^{(0)}$.

d. Show that the order of convergence of the sequence $\{x^{(k)}\}$ in part b is 1.

e. Theorem 9.1 states that under certain conditions, the order of convergence of Newton's method is at least 2. Why does that theorem not hold in this particular problem?

Ans.: a. We have $f'(x) = 4(x - x_0)^3$ and $f''(x) = 12(x - x_0)^2$. Hence, Newton's method is represented as

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x_0}{3},$$

which upon rewriting becomes

$$x^{(k+1)} - x_0 = \frac{2}{3} \left(x^{(k)} - x_0 \right)$$

b. From part a, $y^{(k)} = |x^{(k)} - x_0| = (2/3)|x^{(k-1)} - x_0| = (2/3)y^{(k-1)}$.

c. From part b, we see that $y^{(k)} = (2/3)^k y^{(0)}$ and therefore $y^{(k)} \to 0$. Hence $x^{(k)} \to x_0$ for any $x^{(0)}$.

d. From part b, we have

$$\lim_{k\to\infty} \frac{|x^{(k+1)} - x_0|}{|x^{(k)} - x_0|} = \lim_{k\to\infty} \frac{2}{3} = \frac{2}{3} > 0$$

and hence the order of convergence is 1.

e. The theorem assumes that $f''(x^*) \neq 0$. However, in this problem, $x^* = x_0$, and $f''(x^*) = 0$.

9.3 Consider "Rosenbrock's Function": $f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$, where $\boldsymbol{x} = [x_1, x_2]^T$ (known to be a "nasty" function—often used as a benchmark for testing algorithms). This function is also known as the banana function because of the shape of its level sets.

a. Prove that $[1, 1]^T$ is the unique global minimizer of f over $\mathbb{R}^2$.

b. With a starting point of $[0, 0]^T$, apply two iterations of Newton's method. *Hint:* $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

c. Repeat part b using a gradient algorithm with a fixed step size of $\alpha_k = 0.05$ at each iteration.

Ans.: a. Clearly $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$. We have

$$\begin{aligned} f(\boldsymbol{x}) = 0 \quad &\Leftrightarrow \quad x_2 - x_1^2 = 0 \quad \text{and} \quad 1 - x_1 = 0 \\ &\Leftrightarrow \quad \boldsymbol{x} = [1, 1]^T. \end{aligned}$$

Hence, $f(\boldsymbol{x}) > f([1, 1]^T)$ for all $\boldsymbol{x} \neq [1, 1]^T$, and therefore $[1, 1]^T$ is the unique global minimizer.

b. We compute

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= \begin{bmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200(x_2 - x_1^2) \end{bmatrix} \\ \boldsymbol{F}(\boldsymbol{x}) &= \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}. \end{aligned}$$

To apply Newton's method we use the inverse of the Hessian, which is

$$\boldsymbol{F}(\boldsymbol{x})^{-1} = \frac{1}{80000(x_1^2 - x_2) + 400} \begin{bmatrix} 200 & 400x_1 \\ 400x_1 & 1200x_1^2 - 400x_2 + 2 \end{bmatrix}.$$

Applying two iterations of Newton's method, we have $\boldsymbol{x}^{(1)} = [1, 0]^T$, $\boldsymbol{x}^{(2)} = [1, 1]^T$. Therefore, in this particular case, the method converges in two steps! We emphasize, however, that this fortuitous situation is by no means typical, and is highly dependent on the initial condition.

c. Applying the gradient algorithm $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)})$ with a fixed step size of $\alpha_k = 0.05$, we obtain $\boldsymbol{x}^{(1)} = [0.1, 0]^T$, $\boldsymbol{x}^{(2)} = [0.17, 0.1]^T$.