Optimization

Homework 2

(Due Day: 9:00 AM, Nov 5, 2008, hardcopies in the class)

1. Suppose that we wish to minimize a function $f : \mathbb{R} \to \mathbb{R}$ that has a derivative f'. A simple line search method, called *derivative descent search* (DDS), is described as follows: given that we are at a point $x^{(k)}$, we move in the direction of the negative derivative with step size α; that is, $x^{(k+1)} = x^{(k)} - \alpha f'(x^{(k)})$, where $\alpha > 0$ is a constant.

In the following parts, assume that f is quadratic: $f(x) = \frac{1}{2}ax^2 - bx + c$ (where a, b, and c are constants, and $a > 0$).

a. Write down the value of x^* (in terms of a, b, and c) that minimizes f.

b. Write down the recursive equation for the DDS algorithm explicitly for this quadratic f.

c. Assuming the DDS algorithm converges, show that it converges to the optimal value x^* (found in part a).

d. Find the order of convergence of the algorithm, assuming it does converge.

e. Find the range of values of α for which the algorithm converges (for this particular f) for all starting points $x^{(0)}$.

2. Consider the optimization problem:

$$\text{minimize } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2,$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $m \geq n$, and $\boldsymbol{b} \in \mathbb{R}^m$.

a. Show that the objective function for the above problem is a quadratic function, and write down the gradient and Hessian of this quadratic.

b. Write down the fixed step size gradient algorithm for solving the above optimization problem.

c. Suppose

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Find the largest range of values for α such that the algorithm in part b converges to the solution of the problem.

3. 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 $\boldsymbol{Q}$.

4. 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. Consider the algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \beta\alpha_k\boldsymbol{g}^{(k)},$$

where $\boldsymbol{g}^{(k)} = \boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{b}$, $\alpha_k = \boldsymbol{g}^{(k)T}\boldsymbol{g}^{(k)}/\boldsymbol{g}^{(k)T}\boldsymbol{Q}\boldsymbol{g}^{(k)}$, and $\beta \in \mathbb{R}$ is a given constant. (Note that the above reduces to the steepest descent algorithm if $\beta = 1$.)

Show that $\{\boldsymbol{x}^{(k)}\}$ converges to $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{b}$ for any initial condition $\boldsymbol{x}^{(0)}$ if and only if $0 < \beta < 2$.

5. 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. Consider a gradient algorithm

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

where $\boldsymbol{g}^{(k)} = \boldsymbol{Q}\boldsymbol{x}^{(k)} - \boldsymbol{b}$ is the gradient of f at $\boldsymbol{x}^{(k)}$, and α_k is some step size.

Show that the above algorithm has the descent property (i.e., $f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$ whenever $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$) if and only if $\gamma_k > 0$ for all k.

6. 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.

7. Consider the modified Newton's algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)},$$

where $\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} - \alpha\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)})$. Suppose that we apply the algorithm to a quadratic function $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$, where $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$. Recall that the standard Newton's method reaches the point $\boldsymbol{x}^*$ such that $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ in just one step starting from any initial point $\boldsymbol{x}^{(0)}$. Does the above modified Newton's algorithm possess the same property? Justify your answer.