MATH/AMATH 516 SPRING 2005
SECOND HOMEWORK SET

Problems 1–3 Due Tuesday, April 17. and Problem 4 Due Thursday, April 19.

1. Consider minimizing the continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ on $\mathbb{R}^n$. Let $x \in \mathbb{R}^n$ and $d \in \mathbb{R}^n$ be such that $\nabla f(x)^T d < 0$. Recall that in the backtracking line–search, we are given parameters $0 < \gamma < 1$ and $0 < c_1 < 1$ and we obtain an update to x, say x_+, of the form $x_+ = x + \lambda d$ where

$$\begin{array}{rll} \lambda = & \max \gamma^k & \\ & \text{subject to} & k \in \{1, 2, \ldots\} \text{ and } f(x + \gamma^k d) - f(x) \le c_1 \gamma^k \nabla f(x)^T d \;. \end{array}$$

The key inequality

$$f(x + \lambda d) - f(x) \le c_1 \lambda \nabla f(x)^T d \tag{1}$$

is called the Armijo–Goldstein inequality. A possible shortcoming of this step length is that it is not ties to the one dimensional optimality condition $\nabla f(x + \lambda d)^T d = 0$. In this regard, we consider methods that require the step length λ to satisfy both the Armijo–Goldstein inequality and an inequality of the form

$$\nabla f(x + \lambda d)^T d \ge c_2 \nabla f(x)^T d \tag{2}$$

for a given $c_2 \in (0, 1)$. The conditions (1) and (2) taken together are called the *weak Wolfe conditions.*

(a) Show that if $0 < c_1 < c_2 < 1$ and the set $\{f(x + \lambda d) : \lambda \ge 0\}$ is bounded below, then the two conditions (1) and (2) can be satisfied simultaneously. In particular, show that the set

$$\left\{ \lambda \;\middle|\; \begin{array}{c} \lambda > 0, \nabla f(x + \lambda d)^T d \ge c_2 \nabla f(x)^T d, \text{ and} \\ f(x + \lambda d) - f(x) \le c_1 \lambda \nabla f(x)^T d \end{array} \right\}$$

has non–empty interior.

(b) Show that the following bisection method is either finitely terminating at a value for t at which the weak Wolfe conditions are satisfied, or generates an infinite sequence of trial values for t with $f(x + td) \downarrow -\infty$.

A Bisection Method for the Weak Wolfe Conditions

```
INITIALIZATION: Choose $0 < c_1 < c_2 < 1$, and set $\alpha = 0$, $t = 1$, and $\beta = +\infty$.

REPEAT
    If $f(x + td) > f(x) + c_1 t f'(x; d)$,
        set $\beta = t$ and reset $t = \frac{1}{2}(\alpha + \beta)$.
    Elseif $f'(x + td; d) < c_2 f'(x; d)$,
        set $\alpha = t$ and reset
            $t = \begin{cases} 2\alpha, & \text{if } \beta = +\infty \\ \frac{1}{2}(\alpha + \beta), & \text{otherwise.} \end{cases}$
    Else, STOP.
END REPEAT
```

2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function

$$f(x) = \frac{1}{2}x^T Q x + g^T x + \alpha \ ,$$

where we will assume that Q is symmetric and positive definite. Let x_0 and d_0 be given vectors in $\mathbb{R}^n$. Find a closed form representation for the global solution to the problem

$$\min\{\phi(\lambda) : \lambda \in \mathbb{R}\}$$

where the function $\phi : \mathbb{R} \to \mathbb{R}$ is given by $\phi(\lambda) = f(x_0 + \lambda d_0)$. A step length λ_0 chosen in this way is called the *Curry step size.*

3. Let $Q \in \mathbb{R}^{n\times n}$ be symmetric and positive definite and consider the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) := \frac{1}{2}x^T Q x - b^T x$$

where $b \in \mathbb{R}^n$.

   (a) What is the unique point $\bar{x} \in \mathbb{R}^n$ that achieves the global minimum value of f on $\mathbb{R}^n$?

   (b) A set of vectors $\{d^0, d^1, \ldots, d^k\} \subset \mathbb{R}^n \backslash \{0\}$ are said to be *Q-conjugate* if

$${d^i}^T Q d^j = 0 \quad \text{whenever } i \neq j \ .$$

   If the vectors $\{d^0, d^1, \ldots, d^k\}$ are Q-conjugate, show that they are necessarily linearly independent.

   (c) Given $x, d \in \mathbb{R}^n$ compute the unique solution to the one dimensional optimization problem

$$\min_{t\in\mathbb{R}} f(x + td) \ .$$

   (Hint: Review Part (a) of problem 2.)

   (d) Consider the following algorithm:

```
Initialization: $x_0 \in \mathbb{R}^n$, $\{d^0, d^1, \ldots, d^k\} \in \mathbb{R}^n$.
Having $x^i$ obtain $x^{i+1}$ as follows:
    $x^{i+1} := x^i + t_i d^i$
where $t_i$ solves $\min_{t\in\mathbb{R}} f(x^i + td^i)$.
```

   If the vectors $\{d^0, d^1, \ldots, d^k\}$ are Q-conjugate show that x^{k+1} solves the problem

$$\begin{array}{ll} \min f(x) & \\ \text{subject to} & x \in x^0 + \text{Span}\{d^0, d^1, \ldots, d^k\} \ . \end{array}$$

   (e) If $x^0 \in \mathbb{R}^n$ and $\{d^0, d^1, \ldots, d^{n-1}\}$ are Q-conjugate, show that the n^{th} iterate of the algorithm described above solves the problem $\min_{x\in\mathbb{R}^n} f(x)$.

4. Given a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, a point $x \in \mathbb{R}^n$, a direction $d \in \mathbb{R}^n$ such that $f'(x;d) < 0$, and scalars $\gamma \in (0,1)$ and $c_1 \in (0,1)$, the backtracking line-search requires one to solve the problem

$$\max\{\gamma^s \,:\, s = 0, 1, \ldots,\ f(x + \gamma^s d) \le f(x) + c_1 \gamma^s f'(x;d)\}\ .$$

A sample m-file for solving this problem is provided on the course webpage under *Matlab Sample Code*. The sample code is written for Matlab 5. You may choose to upgrade the code to Matlab 6 (or later) if you wish.

In this problem your task is to develop a comparable subroutine (complete with comments and error checking) that implements a line search based on the bisection method for the weak Wolfe conditions described in problem 1. Your m-file should conform to the following guidelines:

(i) The m–file should take as input the values of the variables $x,\ d,\ 0 < c_1 < c_2 < 1$ and the value of the value $f'(x;d)$. In addition, the m-file should also take as input a character string naming the function f and its gradient ∇f. The function and gradient can then be evaluated using the character strings as pointers for f and ∇f. In Matlab 5 this is done using the *feval* command.

(ii) The output should be the vector $x + td$ and the values $f(x + td)$ and $\nabla f(x + td)$.

(iii) In order to evaluate the function within the m–file use pointers.

(iv) Limit the outer loop to 100 trials.

Compare the performance of the backtracking routine with your bisection method for the weak Wolfe conditions as follows.

(a) Apply the subroutines to minimize the test function

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2\ .$$

This function is called *Rosenbrock's function.* It has a global minimum value of $f(x) = 0$ uniquely occuring at the point $(x_1, x_2) = (1, 1)$. Use the steepest descent direction as the search direction $d = -\nabla f(x)$ (or $d = -\nabla f(x)/\|\nabla f(x)\|$) and take $x^0 = (-2, 2)$. Compare the performance for

$$(\gamma, c_1, c_2) = (.8, .01, .5), (.8, .2, .5), (.5, .01, .5), (.5, .2, .5).$$

Of course, γ is only to be used for the backtracking line search while c_2 is only to be used for the weak Wolf line search. But it would be interesting to know how often the backtracking line search fails the weak Wolfe conditions.

(b) Use the following stopping criteria:

$$\begin{aligned} \|\nabla f(x)\| &\le 10^{-6}\ , \\ \text{number of iterations} &\le 100 \quad \text{or,} \\ \|x^{n+1} - x^n\| &\le 10^{-14}\ . \end{aligned}$$

Describe the performance of your subroutine using the *Programming Guidelines* from the course webpage. Please keep in mind that I do not wish to see page after page of numerical output. Graphs of this output are all that is needed and is often much more insightful.