ISyE 6663 Optimization III

Spring 2011

Assignment 3

Issued: March 1, 2011

Due: March 10, 2011

Problem 1

Consider a constant stepsize algorithm, so that $x_{k+1} = x_k + sd_k$ for some constant stepsize $s > 0$. There are a variety of conditions under which the sequence of iterates $\{x_k\}$ of such an algorithm converges to a stationary point of f.

(1) Consider the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ given by $f(x) := \|x\|_2^{2+a}$, with $a \geq 0$. Consider the application of the steepest descent algorithm with constant stepsize to f, that is, $x_{k+1} = x_k - s\nabla f(x_k)$ for some constant stepsize $s > 0$. Determine for which values of s and x_0 the sequence of iterates $\{x_k\}$ converges to $x^* = 0$.

(2) Consider the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ given by $f(x) := \|x\|_2^{3/2}$.

(a) Show that f is not Lipschitz continuously differentiable, that is, there is no constant L such that

$$\|\nabla f(x) - \nabla f(y)\|_2 \quad \leq \quad L\|x - y\|_2$$

for all $x, y \in \mathbb{R}^n$. (In fact, f is not even locally Lipschitz continuously differentiable at the optimal solution $x^* = 0$, that is, there is no neighborhood of the optimal solution $x^* = 0$ and constant L such that the Lipschitz inequality above holds for all x, y in the neighborhood.)

(b) Consider the application of the steepest descent algorithm with constant stepsize to f, that is, $x_{k+1} = x_k - s\nabla f(x_k)$ for some constant stepsize $s > 0$. Show that, for any value of $s > 0$, the sequence of iterates $\{x_k\}$ either converges to $x^* = 0$ in a finite number of iterations (and only in a very special case), or else the iterates do not converge to x^*.

(3) Consider a quadratic function $f : \mathbb{R}^n \mapsto \mathbb{R}$ given by $f(x) := \frac{1}{2}x^T Gx + d^T x$, where $G \in \mathbb{R}^{n\times n}$ is symmetric positive definite and $d \in \mathbb{R}^n$. In the previous homework you were asked to show that f is Lipschitz continuously differentiable, that is, there is a constant L such that

$$\|\nabla f(x) - \nabla f(y)\|_2 \quad \leq \quad L\|x - y\|_2$$

for all $x, y \in \mathbb{R}^n$, with the smallest such constant L given by the largest eigenvalue of G.

(a) Consider a steepest descent algorithm with a constant stepsize s applied to f. Show that $\{x_k\}$ converges to $x^* = -G^{-1}d$ for any starting point x_0 if and only if $0 < s < 2/L$.

(b) Consider a gradient search algorithm with a constant stepsize s and constant symmetric positive definite deflection matrix B applied to f, that is, $x_{k+1} = x_k - sB\nabla f(x_k)$. Let L be the largest eigenvalue of $B^{1/2}GB^{1/2}$. Show that $\{x_k\}$ converges to $x^* = -G^{-1}d$ for any starting point x_0 if and only if $0 < s < 2/L$.

(4) Consider a quadratic function $f : \mathbb{R}^n \mapsto \mathbb{R}$ given by $f(x) := \frac{1}{2}x^T Gx$, where $G \in \mathbb{R}^{n\times n}$ is nonsingular symmetric indefinite. Consider a steepest descent algorithm with a constant stepsize s applied to f. Show that if the starting point x_0 does not belong to the subspace spanned by the eigenvectors corresponding to the nonnegative eigenvalues of G, the sequence $\{x_k\}$ diverges.

(5) Suggest some conditions that you think would be required for a constant stepsize algorithm to converge to a stationary point of f.

Problem 2

Program the steepest descent and Newton algorithms, both using the Armijo (backtracking) line search method with initial step size $\bar{\alpha} = 1$, decrease factor $\beta = 1/2$, and constant for sufficient decrease condition $c_1 = 0.1$. Use the two algorithms to minimize the Rosenbrock function

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

First try the initial point $x^0 = (1.2, 1.2)$, and then the initial point $x^0 = (-1.2, 1.0)$. Print the step length α^k used by each algorithm at each iteration k, and the distance $\|x^k - x^*\|_2$ between the iterate x^k at each iteration k and the optimal solution x^*.

Problem 3

Prove that $\|Bx\| \geq \|x\|/\|B^{-1}\|$ for any $x \in \mathbb{R}^n$, $B \in \mathbb{R}^{n\times n}$ nonsingular, and any norm $\|\cdot\|$.

Problem 4

Kantorovich inequality: Show that for any symmetric positive definite matrix $Q \in \mathbb{R}^{n\times n}$ and any $x \in \mathbb{R}^n$, it holds that

$$\frac{(x^T x)^2}{(x^T Qx)(x^T Q^{-1}x)} \quad \geq \quad \frac{4\lambda_{\min}\lambda_{\max}}{(\lambda_{\min} + \lambda_{\max})^2}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues of Q respectively.