ISyE 6663 Nonlinear Optimization

Spring 2011

Assignment 2

Issued: February 1, 2011

Due: February 10, 2011

Problem 1
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 $G \in \mathbb{R}^{n\times n}$ and $d \in \mathbb{R}^n$.

(1) Show that we can assume, without loss of generality, that G is symmetric.

(2) Verify that $\nabla f(x) = Gx + d$.

(3) 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 $x, y \in \mathbb{R}^n$.

(4) What is the smallest constant L such that the Lipschitz inequality above holds?

Problem 2
Consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ defined by $f(x) = \|x\|_2^2$. Consider the sequence $\{x_k\}$ given by

$$x_k \quad = \quad \left(1 + \frac{1}{2^k}\right) \begin{bmatrix} \cos(k) \\ \sin(k) \end{bmatrix}$$

for $k = 0, 1, 2, \ldots$.

1. Show that the sequence $\{x_k\}$ satisfies $f(x_{k+1}) < f(x_k)$ for all k.
2. Show that every point on the unit circle $\{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$ is a limit point of the sequence $\{x_k\}$. You may use the following result: The sequence $\{\xi_k\}$ defined by $\xi_k = k(\mod 2\pi)$ is dense in $[0, 2\pi]$. The point of the exercise is to show an example of a sequence $\{x_k\}$ of iterates such that the objective value converges (maybe to the minimum objective value), but the sequence of iterates does not converge.

Problem 3
Prove that all isolated local minimizers are strict. That is, for any isolated local minimizer x^*, show that there is a neighborhood $\mathcal{N}$ of x^* such that for all $x \in \mathcal{N} \setminus \{x^*\}$, it holds that $f(x) > f(x^*)$.

Problem 4
Consider a convex function f. Show that $\arg\min_{x \in \mathrm{dom}(f)} f(x)$ is a convex set.

Problem 5
Suppose that $f : \mathbb{R}^m \mapsto \mathbb{R}$ is continuously differentiable. Consider the function $\tilde{f} : \mathbb{R}^n \mapsto \mathbb{R}$ given by $\tilde{f}(x) = f(Sx + s)$ for some $S \in \mathbb{R}^{m \times n}$ and $s \in \mathbb{R}^m$. Show that

$$\nabla \tilde{f}(x) \quad = \quad S^T \nabla f(Sx + s) \qquad \text{and} \qquad \nabla^2 \tilde{f}(x) \quad = \quad S^T \nabla^2 f(Sx + s) S$$

Hint: First use the chain rule to find an expression for each partial derivative $\partial \tilde{f}(x)/\partial x_i$ in terms of the partial derivatives of f, and then verify that these expressions are consistent with the matrix expressions above.

Problem 6
Consider the symmetric-rank-one (SR1) formula for updating a Hessian approximation:

$$B_{k+1} \quad = \quad B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

Consider the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula for updating a Hessian approximation:

$$B_{k+1} \quad = \quad B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$

Show that both the SR1 update and the BFGS update are scale-invariant. That is, suppose that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is continuously differentiable, and, for each update method, consider the following two sequences $\{x_k\}$ and $\{\tilde{x}_k\}$: Choose initial point $x_0 \in \mathbb{R}^n$ and initial Hessian approximation $B_0 \in \mathbb{R}^{n \times n}$ (typically symmetric positive definite). Then the sequence $\{x_k\}$ is generated as follows:

$$x_{k+1} \quad = \quad x_k - B_k^{-1} \nabla f(x_k)$$

(For now, assume that all B_k are nonsingular.) Next, consider an affine rescaling of the decision variables x given by $\tilde{x} = S^{-1}(x - s)$, where $S \in \mathbb{R}^{n \times n}$ is nonsingular, and $s \in \mathbb{R}^n$. That is, the new objective function is given by $\tilde{f}(\tilde{x}) = f(S\tilde{x} + s)$. Choose scaled initial point $\tilde{x}_0 = S^{-1}(x_0 - s)$, and corresponding initial Hessian approximation $\tilde{B}_0 = S^T B_0 S$ ($\tilde{B}_0$ is symmetric positive definite iff B_0 is symmetric positive definite). Then the sequence $\{\tilde{x}_k\}$ is generated as follows:

$$\tilde{x}_{k+1} \quad = \quad \tilde{x}_k - \tilde{B}_k^{-1} \nabla \tilde{f}(\tilde{x}_k)$$

Show that the two sequences maintain the initial scaling relationship, that is, $\tilde{x}_k = S^{-1}(x_k - s)$ for all k.

Problem 7
Consider a quadratic function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ given by $f(x, y) := ax^2 + by^2 + cxy + dx + ey$.

(1) Suppose that $a = b = 1$. Identify and classify the stationary points of f (it may depend on the other parameters c, d, e).

(2) Identify and classify the stationary points of f as a function of all the parameters a, b, c, d, e.

Problem 8

A local minimizer in every direction is not necessarily a local minimum:
Consider a differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$. Consider a point x^* that is a local minimizer of f along every line through x^*; that is, the function $g(\alpha) := f(x^* + \alpha d)$ is minimized at $\alpha = 0$ for all $d \in \mathbb{R}^n$.

(1) Show that $\nabla f(x^*) = 0$.

(2) Consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ given by $f(x, y) := (x - ay^2)(x - by^2)$ for some $a, b > 0$, $a \neq b$. Show that $x^* = 0$ is a local minimizer of f along every line through x^*, but that x^* is not a local minimum of f.

Problem 9

Show that the sequence $\{x_k\}$ given by $x_k = 1/k$ is not Q-linearly convergent, though it does converge to 0. (It converges sublinearly to 0.)