The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381V-11: Large Scale Optimization — Fall 2012

Problem Set Three

Caramanis/Sanghavi Due: Thursday, September 27, 2012.

Reading Assignments

1. Reading: Boyd & Vandenberghe: Chapters 9.1 - 9.6.

Written Problems

1. **Gradient descent with diminishing step size**

   Recall the proof, done in class, of the convergence of gradient descent with fixed step size. We showed that if a function has L-lipschitz gradients, and its minimum is finite-valued, and the step size $\eta < 2/L$, then gradient descent converges to a stationary point. (This has the advantage of not having to know L in advance)

   Prove that the same holds true for any sequence $\{\eta^{(k)}\}$ of step sizes such that $\eta^{(k)} \to 0$ and $\sum_0^\infty \eta^{(k)} = \infty$.

   You can use the following fact: with $\eta^{(k)}$ as above, if there is a sequence $a^{(k)}$ such that $\sum_{k=0}^{k=\infty} a^{(k)}\eta^{(k)} < \infty$, then it has to be that $a^{(k)} \to 0$.

2. **Gradient descent and non-convexity**

   Consider the gradient descent algorithm with fixed step size η for the function $f(x) = x'Qx$, where Q is symmetric but not positive semidefinite. (i.e., Q has some negative eigenvalues). Exactly describe the set of initial points from which gradient descent, with *any* positive step size, will diverge. What happens at the other points, if the step size is small enough ?

3. **Jacobi Method**

   Recall that coordinate descent (with exact line search) involves minimizing over one coordinate at a time, keeping the other coordinates fixed. The Jacobi method involves, in a sense, doing all minimizations simultaneously. In particular, given a point x, define the vector $\bar{x}$, in which the value at every coordinate i is determined by the corresponding *individual* coordinate descent update

   $$\bar{x}_i \;:=\; \arg\min_{\psi} f(x_1, \ldots, x_{i-1}, \psi, x_{i+1}, \ldots, x_n)$$

   Thus, potentially, every coordinate of $\bar{x}$ could be different from that of x.

   The Jacobi method is defined by the iteration

   $$x_+ \;=\; x + \alpha(\bar{x} - x)$$

   Prove that, for a convex continuously differentiable f, and a step size $\alpha = 1/n$ where n is the number of coordinates, the next iterate of the Jacobi method produces a lower function value than x, provided x does not already minimize the function.
   (Hint: express x_+ as a convex combination of n points.)

4. **Step size in Newton**

   Consider the use of Newton's method with constant step size t to minimize the function $\|x\|^3$.

   (a) For what values of t do we obtain global convergence to the minimum (i.e. $x^* = 0$) ? What happens for the other values of t ?

   (b) For the values of t for which it does converge, why is the convergence not quadratic ?

5. **Composite functions**

   Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ be a convex function, $\phi : \mathbb{R} \to \mathbb{R}$ be both convex and increasing, and define $g(x) = \phi(f(x))$. Assume that both f and g are twice differentiable everywhere. Note that this means g too is convex.

   (a) Consider an initial point $x^{(0)}$, and run two gradient descent with exact line search iterations: one on f, and the other one on g, with this same initial point. Show that the entire sequence of iterates will then be the same.

   (b) Is the same true for the Newton methods with exact line search ? Prove your answer, or provide a simple counter-example. *(Hint: use the matrix inversion lemma of Appendix C.4.3 in the text)*