Homework 2

October 4, 2016

* For each question, give your answer. Justify your answer using a detailed explanation.

30% (Convergence under strong convexity)

Assume in our linear regression problem, $J(\theta)$ is strongly convex, i.e., $X^T X$ is positive definite, so we have

$$J(\xi) \geq J(\theta) + \nabla J(\theta)^T(\xi - \theta) + \frac{\delta}{2}\|\xi - \theta\|^2$$

holds true. Let θ^* be the optimal solution (minimizer). Show that if we have

$$\|\nabla J(\theta)\| \leq \sqrt{\frac{\delta\epsilon}{2}}$$

then we have

$$J(\theta) - J(\theta^*) \leq \epsilon$$

(Hint: using convexity, you can build up your first inequality $J(\theta^*) \geq J(\theta) + \ldots$; using strongly convexity and that gradient is 0 at optimal solution, you can derive your second inequality. The result can be proved by combining the two inequalities.)

70% (Gradient descent)

Program the gradient descent algorithm in Matlab with fixed stepsize α to solve least square problems $\min_\theta \ f(\theta) := \frac{1}{2}\|\mathrm{X}\theta - \mathrm{y}\|_2^2$. Your code should have format

function [theta, f, g] = GDLS(X,y,theta0,alpha,tol) where

- X can be any $m \times n$ matrix
- y can be any m dimensional vector
- theta0 is the initial point, alpha is the stepsize
- tol is the tolerance for $\nabla f(x)$
- Output theta is the final solution you find, f and g are vectors that store the value of $f(x)$ and $\|\nabla f(x)\|_2$ over the iterates $x_1, x_2, \ldots$, respectively.

In the following experiment, use tol $= 10^{-6}$.

(i) Write down the pseudo-code for your algorithm.

(ii) For

$$\mathrm{X} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \mathrm{y} = \begin{pmatrix} 10 \\ 20 \end{pmatrix}, \quad \theta^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

give the stepsize that you can use in your code, explain why you choose this value. What is the solution you find?

(iii) Plot the curve of $\log \mathrm{f}$ and $\log \mathrm{g}$, what convergence rate do you observe?

(iv) Load the data X, y and theta0 from “hw2.mat”, run your code. What is the stepsize you use?

(v) Plot $\log \mathrm{g}$, what convergence rate do you observe?.

• Notice: when I said tol is the the tolerance for $\nabla f(x)$, I didn't mean the termination criterion is

$$\|\nabla f(x^k)\| \leq \text{tol}.$$

Instead, it should be the one I mentioned in the class:

$$\frac{\|\nabla f(x^k)\|}{\max\{1, \|\nabla f(x^0)\|\}} \leq \text{tol}.$$