STAT 7934: HW 1

Due in my office on Fr Feb 6 by 5 pm

Group Project:

Consider a classification problem, where the response is binary $y \in \{0, 1\}$ and the predictor variables numerical. .Generate a design matrix X of size 25,000 by 3,000 (i.e. 25,000 observations and 3,000 variables). Generate the corresponding response y by selecting an appropriate coefficient vector β and adding a noise term.
Hint: Think what the appropriate criterion (likelihood) function is for this problem.

(a) Generate three versions of the data set, one corresponding to a well-conditioned Gram (i.e.$X'X$) matrix (i.e. its condition number close to one), one fairly well conditioned (condition number close to 30) and one corresponding to an ill-conditioned Gram matrix (i.e. condition number around 500). Explain carefully how the data were generated.

(b) Solve the problem for all instances by implementing the steepest descent algorithm for the following choices of step size: constant, diminishing step size (explain which choice you made) and Armijo'ss rule. Discuss what you observe both in terms of the choice of your step size and that of the conditioning of the Gram matrix.

(c) For the ill-conditioned version, implement the steepest descent method with the three choices of step size adding a ridge penalty on the β coefficient;

$$\min_{\beta} f(\beta) + \lambda ||\beta||_2^2,$$

for an appropriate choise of $f : \mathbb{R}^p \to \mathbb{R}$ (see part (a)).

Discuss how the choice of the tuning parameter λ impacts the results.

Individual Question:

Apply Newtons method to minimizing the function

$$f(x) = ||x||^3, \quad x \in \mathbb{R}^2$$

using a constant step size and where $|| \cdot ||$ denotes the Euclidean norm. For what range of step sizes the method converges?