ECE 6500 — **Assignment # 1**

Issued: Jan 12, 2023 — Due: Jan 31, 2023

Reading Assignments:

- Read Boyd & Vandenberghe, Chapter 1, Sections 2.1-2.3, Sections 3.1-3.2, and 3.5.

Note: The softcopy of the book is accessible at:
https://web.stanford.edu/∼boyd/cvxbook/bv_cvxbook.pdf

(1) (Existence of X_{opt}) Is the (i) minimum and (ii) maximum for the function

$$f(\mathbf{x}) = x_1^2 + e^{x_2} + e^{-x_2} + 3x_3^4$$

exist for

(a) over the set $\mathcal{X} = \{\mathbf{x} \in \mathbf{R}^3 : x_1^2 + 2x_2^2 + 3x_3^2 \leq 1\}$?

(b) over the set $\mathbf{R}^3$?

(2) (GM for a non-convex function) Consider the implementation of the Gradient Descent Method (GM) to $f : \mathcal{R}^2 \to \mathcal{R}$ given by

$$f(\mathbf{x}) = \frac{1}{2}x_1^2 + \frac{1}{4}x_2^4 - \frac{1}{2}x_2^2$$

(a) Plot the function in a 3-D figure and identify the local minima.

(b) Write the GM iteration, and consider the trajectory of GM with a small enough (identify how small it should be for guaranteed convergence) constant stepsize for the initial state $\mathbf{x}^{(0)} = (1, 0)^T$. Does $\{\mathbf{x}^{(t)}\}_t$ converge to one of the identified minima? Does your observation conflict with the claimed results in class?

(3) (GM for a convex non-Lipschitz gradient function) Consider the operation of GM with constant step-size on the function $f : \mathcal{R}^2 \to \mathcal{R}$ given by

$$f(\mathbf{x}) = \|\mathbf{x}\|^{3/2}.$$

(a) Show that this function does not have Lipschitz ∇f for any choice of $C < \infty$.

(b) Show that for any value of the constant stepsize $\gamma > 0$, the method either converges in a *finite* number of iterations to the unique minimizing point $\mathbf{x}^* = \mathbf{0}$, or else it does not converge to $\mathbf{x}^*$.

(4) (GM for Linear Regression) Recall the Least-Square minimization problem formulated in Lecture 3 for Linear Regression with $h_\theta(\mathbf{x}) = \theta_0 + \sum_{j=1}^n \theta_j x_j$. Given a set $\{(\mathbf{x}^{(m)}, y^{(m)})\}_{m=1,\ldots,M}$ of M input-output observations, suppose you want to apply steepest descent to finding the optimal vector $\theta^\star$ that minimizes the quadratic cost. Express the update rule for the vector θ under GM.

(5) (Convex Sets) Problem 2.11 from B & V.

(6) (Convex Functions) Problem 3.16 from B & V [specify: convex, concave, both, or neither.]

(7) (Logistic Regression convexity) Recall the Logistic Regression cost function expressed in Lecture 3. In particular, for the following function:

$$f(\theta) = -y\log(h_\theta(\mathbf{x})) - (1-y)\log(1-h_\theta(\mathbf{x}), \qquad \text{where} \qquad h_\theta(\mathbf{x}) = \frac{1}{1+\exp(-\theta^T\mathbf{x})},$$

prove that this cost is a convex function of the vector $\theta \in \mathbb{R}^n$ for each $\mathbf{x} \in \mathbb{R}^n$ and $y \in [0,1]$.

(You can use the fact that sum of two convex functions is also convex without proof.)

(8) (GM Implementation-Investigation) Recall that the steepest gradient descent algorithm with a constant gradient is given as:

$$x^{k+1} = x^k - \gamma\nabla f(x^k).$$

We showed in class that if f is a function with a Lipschitz gradients with constant L, then

$$f(x^{k+1}) \leq \quad f(x^k) - \gamma(1-\frac{\gamma L}{2})\|\nabla f(x^k)\|^2 \tag{1}$$

Consider the two-dimensional quadratic function $f(x,y) = x^2 + 20y^2$. Run the "gradient-constant" algorithm provided.

(a) Determine μ and L, where μ and L are the least and the greatest eigenvalues of the Hessian $\nabla^2 f(x,y)$, respectively. (That is, $\mu\mathrm{I} \leq \nabla^2\mathrm{f(x,y)} \leq \mathrm{LI}$.) Set the variables mu and L in the MATLAB code "mainOpt.m" to these values.

(b) Recall the interval (A,B) of the step size γ for which convergence is feasible. Set $\gamma_1 = B$, the upper limit of the interval .

(c) Using the MATLAB code provided, investigate the behavior of convergence in an ϵ-neighborhood of B. Use the row vector $gammaVec$ in "mainOpt.m" to assign the competing step sizes. What do you observe? Comment on these observations.

(Note: Choose $\epsilon < 0.001$ to make your observations.)

(d) Find the step size γ_3 using the above upper bound 1 in terms of L that the achieves fastest convergence rate for the steepest descent algorithm. Compare and illustrate with a plot the convergence speed of this choice with respect to other extremes (in particular, compare to too small and barely stabilizing choices of γ)

(e) $f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|^2$, since $\nabla^2 f(z) \geq \mu\mathrm{I}$, for all z.

Use the row vector $gammaVec$ to assign competing step sizes, γ_1, $\gamma_2 = \frac{2}{L+\mu}$ and γ_3. What do you observe? Do these observations align with your findings in (b)-(d). Comment.