CSE257 Fall 2021: Assignment 1

Deadline: Oct-24 11:59pm. Submit a PDF of your answers on Gradescope (Course Code: 74J38G). Email a zip file of all your source code for the questions that involve implementation to `ucsdcsegaoclasses@gmail.com`, by the same deadline. All of the code should be in Python and you can use standard libraries such as Numpy, Scipy, Matplotlib. There is no need for deep learning libraries such as tenslorflow or Pytorch. Feel free to search online and in textbooks for help with proofs (searching for references is by itself part of the learning process), but you are not allowed to copy any code from other sources. Finish all the answers and implementation completely by yourself.

Question 1 (4 Points, implementation involved)**.** Plot some level sets of the following three functions (when we say "plot something" you need to show the plots in the PDF file):

- $f_1(x_1, x_2) = (x_1 - x_2)^2$
- $f_2(x_1, x_2) = x_1^2 - 3x_2^2$
- $f_3(x_1, x_2, x_3) = x_1^2 + 5x_2^2 + 3x_3^2$

for the range of $x_i \in [-10, 10]$ for all i. You only need to plot the level sets of the function roughly taking integer values, i.e. $f(x) = 1, -1, 0$ etc. (small numerical errors are allowed, we only care about the shapes of the curves).

Question 2 (8 Points, implementation involved)**.** Consider $f(x_1, x_2) = x_1^2 - x_1x_2 + 3x_2^2 + 5$ and let the initial point be $x = (2, 2)$. Implement the following procedures for minimizing f:

1. (2 Points) Perform gradient descent using a fixed step size $\alpha = 0.3$.
2. (3 Points) Perform gradient descent using optimal step sizes in each step.
3. (3 Points) Perform Newton descent using step size $\alpha = 1$.

For each case, plot two graphs. First, the value of f with respect to the number of iterations (i.e., let the x-axis be the number of iterations starting from 0, and y-axis shows the f value). Second, plot the sequence of points you get in each iteration and the gradient descent direction, in the x_1-x_2 plane. If there are many iteration steps you only need to show up to 20 steps in each plot.

Question 3 (3 Points)**.** Prove that if a function is convex, then any of its local minimum is also a global minimum. Hint: suppose a function has two local minima that correspond to different function values (so not both are global minima), then prove that the function can not be convex.

Question 4 (7 Points + 5 Extra Points, coding involved)**.** Consider the following regression problem. We aim to train a model $f_{A,B}(x) = Ax+B$ where $x \in \mathbb{R}^3$ and $f(x) \in \mathbb{R}^3$, where the parameter matrices A and B are of the appropriate sizes, to fit the following dataset:

$$D = \{(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\-3\\1\end{bmatrix}), (\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}1\\-3\\1\end{bmatrix}), (\begin{bmatrix}-1\\2\\-1\end{bmatrix}, \begin{bmatrix}0\\1\\2\end{bmatrix})\}.$$

Namely, these are the three data points in D (let's label them (x_i, y_i) for $i = 1, 2, 3$, respectively); for each pair in D, the first component x_i is the input vector, and the second component y_i is the label vector that we hope to predict. We aim to find the parameter matrices A and B such that the mean squared error over D is minimized. That is, we need to minimize the loss function

$$L(A, B) = \sum_{i=1}^{3} (f_{A,B}(x_i) - y_i)^2$$

where $(x_i, y_i) \in D$. Answer the following:

1. (3 Points) Prove that any local minimum of $L(A, B)$ is also a global minimum of $L(A, B)$. The proof can either involve numerical calculations or a general argument that only uses the symbolic formulas.
2. (3 Points) Perform gradient descent to find one solution A, B that minimizes $L(A, B)$ and plot the value of L over the number of iterations until convergence (you can choose any initial point and any step size you like).
3. (1 Point) Can you use Newton directions for this minimization? Why?
4. (Extra 5 Points) Perform conjugate gradient descent from the same initial point as what you chose for gradient descent. You can give it more data points for the problem to be solvable with conjugate descent methods. Plot the change of loss over number of iterations.

Question 5 (4 Points)**.** Prove that gradient descent with exact line search always takes orthogonal steps. That is, suppose p_k is the gradient descent direction in the kth iteration of gradient descent for minimizing an arbitrary continuously differentiable and lower-bounded function $f(x) : \mathbb{R}^n \to \mathbb{R}$. After we use a step size α that minimizes $f(x_k + \alpha p_k)$ along the direction of p_k (i.e. $\min_{\alpha \geq 0} f(x_k + \alpha p_k)$), the next direction p_{k+1} in gradient descent always satisfies $p_k p_{k+1} = 0$.

Question 6 (4 Points)**.** Follow the proof sketches in the slides and rigorously prove the first-order and second-order necessary conditions for local minima.