Bowen Weng
Department of Computer Science
Iowa State University

HW3 — Optimization

1 Definitions

1. Are the following sets convex?
   (a) $\{x \mid x \in \mathbb{R}^2, x^T x \leq 2\}$
   (b) $\{x \mid x \in \mathbb{R}^2, x^T x \geq 2\}$
2. Are the following sets strictly convex?
   (a) $\{x \mid x \in \mathbb{R}^2, x^T x \leq 2\}$
   (b) $\{x \mid Ax \leq 0\}$ ($A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$)
3. Are the following functions convex?
   (a) $f(x) = x^2, x \in \mathbb{R}$
   (b) $f(x) = x^2, x \in [0, 1]$
   (c) $f(x) = x^T x + 4, x \in \mathbb{R}^2$
   (d) $f(x) = x^T x + 4, x \in \mathbb{R}^2, x^T x \geq 2$
4. Are the following matrices Positive Definite (PD), Positive Semi-Definite (PSD), Negative Definite (ND), or Negative Semi-Definite (NSD)? Please explain your answers.
   (a) $\begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$
   (b) $\begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$
   (c) $\begin{bmatrix} -5 & 2 & 0 \\ 2 & -3 & 1 \\ 0 & 1 & -2 \end{bmatrix}$
5. Are the following functions Lipschitz continuous? Why?
   (a) $f(x) = \sqrt{x}, x \in [0, 1]$
   (b) $f(x) = x^2, x \in \mathbb{R}$
   (c) $f(x) = x^2, x \in [0, 1]$
   (d) $f(x) = |x|, x \in \mathbb{R}$
   (e) $f(x) = \sqrt{7x^2 + 4}, x \in \mathbb{R}$
   (f) $f(x) = \sin x, x \in \mathbb{R}$
   (g) $f(x) = x^5, x \in \mathbb{R}$
6. Are the following functions Lipschitz smooth? Why?
   (a) $f(x) = \sqrt{x}, x \in [0, 1]$
   (b) $f(x) = x^2, x \in \mathbb{R}$

(c) $f(x) = x^2, x \in [0, 1]$

(d) $f(x) = |x|, x \in \mathbb{R}$

(e) $f(x) = \sqrt{7x^2 + 4}, x \in \mathbb{R}$

(f) $f(x) = \sin x, x \in \mathbb{R}$

(g) $f(x) = x^5, x \in \mathbb{R}$

7. Recall the (second) definition of the convex function given in the lecture, consider the following lemma (note $\langle a, b\rangle$ denotes $a^T b$ for some vectors a and b):

   Lemma. *If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and convex, then*

   $$\forall x, y \in \mathbb{R}^n, f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle. \tag{1-1}$$

   (a) What does the lemma mean (hint: in the perspective of the first definition of the convex function given in the lecture)?

   (b) Prove the lemma (hint: use the convexity definition and the notion of limits).

8. Recall the smooth function definition, let $f : \mathbb{R}^n \to \mathbb{R}$ be L-smooth, convex, with the minimum value at $x = x^*$, prove

   $$\|\nabla f(x)\|^2 \leq 2L(f(x) - f(x^*)) \tag{1-2}$$

9. How many feasible solutions do the following optimization problems have? Why?

   (a) $\min_x x^4, x \in \mathbb{R}$

   (b) $\min_x x^T x, x \in \{x \mid x \in \mathbb{R}^2, x^T x \geq 2\}$

   (c) $\min_x f(x), f(x) = \begin{cases} x^2, \text{if } x \geq 0 \\ 0, \text{otherwise} \end{cases}$

   (d) $\min_x x^4 - 4x^2 + 4, x \in \mathbb{R}$

   (e) $\min_x \sin(x), x \in \mathbb{R}$

2 Stochastic Gradient Descent (or Life)

1. Consider the function

   $$f(x) = \frac{x^2}{4} + 1 - \cos(2\pi x), x \in [-4.5, 4.5] \tag{1-3}$$

   (a) Draw $f(x)$ versus x. You can use a picture of hand-drawing (no need to be perfect) or with the help of the computer program.

   (b) Find an initial point x_0 where gradient descent could end up converging to a local minimum that is not the global minimum.

   (c) Find a subset of $[-4.5, 4.5]$ such that f(x) has a non-unique global minimum.

   (d) Find a subset of $[-4.5, 4.5]$ such that f(x) has a unique global minimum.

2. Consider the linear regression problem in the previous lectures. Recall the "real estate" data set (see code on Canvas and more details in the slides). Consider the data set as it stands (i.e., incorporating all six x features and the target y being the house price of unit area). Given the candidate function

   $$y = \omega^T x + b, \omega \in \mathbb{R}^6, b \in \mathbb{R}, \tag{1-4}$$

   and the cost function defined as RSS (unless noted otherwise), consider the following questions. Note the gradient formulation (if used) is expected to be derived analytically (then implemented in code), but the calculation of gradients can be done numerically in a variety of ways. As a result, your Python code is expected to use only standard Python libraries and `numpy` only.

(a) Starting from the initial point at $\omega = \begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1\end{bmatrix}^T$ and $b = 10$, execute the gradient descent algorithm for 4140 steps at a constant learning rate $\gamma = 0.001$, what is the updated solution (you are extremely likely to need some help from your computer here)?

(b) If you follow the configuration described in the previous question (i.e., same initial point, same learning rate), for a sufficiently large running steps, will Gradient Descent lead you to the same optimal solution you have derived from the regression lecture (in other words, will it converge?)? Why?

(c) Starting from the initial point at at $\omega = \begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1\end{bmatrix}^T$ and $b = 10$, execute the stochastic gradient descent algorithm for 4140 steps at a constant learning rate $\gamma = 0.001$ (let's say an invisible magical force is controlling the generation of all randomness on our planet, and your "stochastic" samples end up being the enumeration of the 414 points in the real estate data set following the exact order as it stands in the .csv file, thus to have the 4140 steps, you will have to repeat the enumeration at the same order for 10 times), what is the updated solution? Is your solution better than, the same with, or worth than the solution from 2(a)? Why?

(d) Let's consider a cost that is not RSS, but takes the following form:

$$J(X, y; \omega, b) = \sum_{i=1}^{n} |y_i - \omega^T x_i - b|. \tag{1-5}$$

i. Is the optimal solution unique? If the optimal solution is not unique, why? If it is unique, is it still the same with the original one derived from the RSS cost?

ii. For sufficiently large running steps with learning rate $\gamma = 0.001$, can you prove the convergence of Gradient Descent using the updated cost function (1-5)? Why?

3. A manufacturing company produces a product using different materials. The goal is to minimize the production cost while meeting certain quality constraints. The company uses three materials, A, B, and C, in varying quantities to produce one unit of the product. The cost per unit of each material is as follows:

- Material A: \$4 per unit
- Material B: \$7 per unit
- Material C: \$3 per unit

Additionally, the company must meet the following quality constraints for each unit produced:

- The product must contain at least 20 units of Material A.
- The product must contain at least 15 units of Material B.
- The product must contain at least 10 units of Material C.

The company wants to determine the optimal quantities of each material to use in production to minimize the total cost while satisfying the quality constraints.

(a) Formulate the above problem as an optimization problem.

(b) Solve it with Gradient Descent.

3 A "Bonus" Question, This is the Last Time (1 pt)

From a scale of 1 to 5, how difficult is HW3? 1 is "I can do it in my sleep". 5 is "Bowen is ridiculous". 0 is "I refuse to answer this question". This is for my own reference to improve the quality of future assignments. Thanks!

Submitted by Bowen Weng on .