Problem Set 8

This problem set is due **at 10:00pm** on **Wednesday, April 29, 2020**.

Please make note of the following instructions:

- This assignment, like later assignments, consists of *exercises* and *problems*. **Hand in solutions to the problems only.** However, we strongly advise that you work out the exercises for yourself, since they will help you learn the course material. You are responsible for the material they cover.
- Remember that the problem set must be submitted on Gradescope. If you haven't done so already, please signup for 6.046 Spring 2020 on Gradescope, with the entry code MNEBKP, to submit this assignment.
- We require that the solution to the problems is submitted as a PDF file, **typeset on LaTeX**, using the template available in the course materials. Each submitted solution should start with your name, the course number, the problem number, your recitation section, the date, and the names of any students with whom you collaborated.
- You will often be called upon to "give an algorithm" to solve a certain problem. Your write-up should take the form of a short essay. A topic paragraph should summarize the problem you are solving and what your results are. The body of your essay should provide the following:
  1. A description of the algorithm in English and, if helpful, pseudocode.
  2. A proof (or indication) of the correctness of the algorithm.
  3. An analysis of the asymptotic running time behavior of the algorithm.
  4. Optionally, you may find it useful to include a worked example or diagram to show more precisely how your algorithm works.

EXERCISES (NOT TO BE TURNED IN)

Continuous Optimization

- Show that for a β-smooth convex function, to maximize the guarantee of how much you move each iteration, $f(x^{t-1}) - f(x^t)$, the optimal step size setting is $O(\frac{1}{\beta})$.
- Show that always for a β-smooth α-strongly convex function, $\beta \geq \alpha$.

Problem 8-1. The Gradient Descent Party [40 points] Bob is holding a party with his favorite computer science researchers! This one is themed around solving systems of linear equations. Unfortunately, all of the partygoers have been conducting deep learning research, and everyone forgot all of their algorithms except gradient descent. Knowing gradient descent is truly the mother of all algorithms, Bob is unafraid.

To get the party started, Bob would like to solve the system $Ax = b$, where

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 2 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Bob comes up with the objective function

$$f_{A,b}(x) = \frac{1}{2} x^T (A^T A) x - b^T A x.$$

Now, all Bob needs to do is run gradient descent.

(a) [5 points] Compute

$$\frac{\partial f_{A,b}}{\partial x_1} \text{ and } \frac{\partial f_{A,b}}{\partial x_2}.$$

(b) [5 points] Show that for any critical point $\hat{x}$ of $f_{A,b}$, we have $x_1 = -1, x_2 = 1$ (which are the solutions to the system $Ax = b$).

(c) [20 points] Argue that $f_{A,b}$ is α strongly convex, and find the largest such α. Also show that the function is β smooth, and find the smallest such β.

Recall that a function f with Hessian H is α strongly convex if $x^T H x \geq \alpha$ for all unit norm x, and that f is β smooth if $x^T H x \leq \beta$, for all unit norm x.

Unit norm here means $(\sum_{i=1}^{d} x_i^2)^{1/2} = 1$.

(d) [10 points] How many iterations will it take for $f_{A,b}$ to converge ϵ close to the optimal solution x^*? You can express you answer in big O notation in terms of x_0 and ϵ, where x_0 is your initialization for gradient descent.

Problem 8-2. Not (Strongly) Convex Enough [50 points]

Greta Entz wants to use gradient descent to minimize β-smooth convex function $f(x)$. However, as the function f is *not* α-strongly convex for any $\alpha > 0$, she is having trouble applying the convergence bound she learned in 6.046, and this makes her deeply unhappy.

To remedy this problem, Greta has come up with the following approach. Instead of minimizing the function f directly, she will use the gradient descent method to minimize a function $g_t(x)$ instead, where

$$g_t(x) = f(x) + t \cdot x^2$$

for some parameter t.

Assume that the minimizer x^* of the function f is guaranteed to be near the origin, with $|x^*| < B$. What is the (asymptotically) best iteration bound you can establish for the above scheme to be guaranteed to converge to an ϵ-approximate minimizer of f? What is the value of t that corresponds to that scheme?
Note: The value of t can depend on ϵ.

Problem 8-3. Feedback Form [10 points] Please fill out a feedback form about this problem set at

`https://forms.gle/tdzQHbeoq2JkNNy38`.

It should not take more than a few minutes and will greatly help us improve teaching and material for future semesters!