Nonlinear Optimization I, Fall 2021
Midterm
Due before lecture on 10/19
Please only in file formats .pdf, .zip, standard image formats (.jpg, .png, etc.).

Your submitted solutions to the midterm should be entirely your own work. Do not consult with anyone (other than the professor) or any online source about exam questions. Feel free to cite any result presented in class or in homeworks without proof. You may access any material posted on the course blackboard.

You can write solutions by hand or type them up (the LaTeX code for this pdf is on blackboard).

Q1. (True/False and Why, 25pts) For each of the following claims, state whether it is true or false (2pts) provide a short proof or counterexample justifying your claim (3pts). An insightful justification for a wrong true/false answer can still get you the latter 3pts.

(a) Every global minimizer $x^* \in \mathbb{R}^d$ of a function $f\colon \mathbb{R}^d \to \mathbb{R}$ has $\nabla f(x^*) = 0$.

(b) For any function $f\colon \mathbb{R}^d \to \mathbb{R}$ and point $\bar{x} \in \mathbb{R}^d$, $\mathrm{prox}_f(\bar{x})$ is a singleton.

(c) For any twice continuously differentiable function $f\colon \mathbb{R}^d \to \mathbb{R}$, if $x^* \in \mathbb{R}^d$ has $\nabla^2 f(x^*) \succeq 0$ and $\nabla f(x^*) = 0$, then x^* is a local minimizer of f.

(d) Every convex function $f\colon \mathbb{R}^d \to \mathbb{R}$ that is bounded below has a global minimizer.

(e) For any convex $f\colon \mathbb{R}^d \to \mathbb{R}$ and point $\bar{x} \in \mathbb{R}^d$, if some point $\bar{x}^* \in \mathbb{R}^d$ has $\rho(\bar{x} - \bar{x}^*) \in \partial f(\bar{x}^*)$ then $\bar{x}^*$ globally minimizes $f(x) + \frac{\rho}{2}\|x - \bar{x}\|_2^2$.

Q2. (Smoothness and Nonconvex Proximal Operators, 20pts) Consider a continuously differentiable function $f \colon \mathbb{R}^d \to \mathbb{R}$ with L-Lipschitz gradient (note f may be nonconvex) and $\alpha > 0$.

(a) (5pts) Prove that $f(x) + \frac{1}{2\alpha}\|x\|_2^2$ has $(1/\alpha + L)$-Lipschitz continuous gradient.

(b) (5pts) Prove that $f(x) + \frac{1}{2\alpha}\|x\|_2^2$ is $(1/\alpha - L)$-strongly convex if $\alpha < 1/L$.

(c) (5pts) Show that $\mathrm{prox}_{\alpha f}(\bar{x})$ is a singleton for every $\alpha < 1/L$ and $\bar{x} \in \mathbb{R}^d$.

(d) (5pts) Propose an algorithm for approximately computing $\mathrm{prox}_{\alpha f}(\bar{x})$ (given $\alpha < 1/L$, an initial point $\bar{x} \in \mathbb{R}^d$, and an oracle that computes gradients $\nabla f(x)$.)
Claim (without proof) a convergence rate for your algorithm (the faster the convergence rate you can guarantee, the more points you will get).

Q3. (LASSO Optimality Conditions, 15pts) For any matrix $A \in \mathbb{R}^{m \times n}$, vector $b \in \mathbb{R}^m$, and scalar $\gamma > 0$, consider the LASSO optimization problem defined as

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} \|Ax - b\|_2^2 + \gamma \|x\|_1.$$

Prove that a point x^* is a global minimizer of f if and only if every $i = 1 \dots n$ has

$$0 \in A_i^T(Ax^* - b) + \begin{cases} \{\gamma\} & \text{if } x_i^* > 0 \\ [-\gamma, \gamma] & \text{if } x_i^* = 0 \\ \{-\gamma\} & \text{if } x_i^* < 0 \end{cases}$$

where A_i is the ith column of A.

Q4. (Improved Nonconvex Convergence Guarantees, 20pts) For a continuously differentiable function $f\colon \mathbb{R}^d \to \mathbb{R}$, consider the following "error bound" condition for some $\mu > 0$

$$\frac{1}{2}\|\nabla f(x)\|_2^2 \geq \mu(f(x) - \min_{x'} f(x')) \tag{1}$$

for any point $x \in \mathbb{R}^d$.

(a) (5pts) Show that (1) holds for every μ-strongly convex f.

(b) (5pts) Give an example of a nonconvex function $f\colon \mathbb{R} \to \mathbb{R}$ satisfying (1).

(c) (5pts) Improve our nonconvex gradient descent convergence rate from class to the following: For any f with L-Lipschitz gradient and satisfying (1), the iterates $x_{k+1} = x_k - \nabla f(x_k)/L$ have

$$f(x_k) - \min f(x) \leq \left(1 - \frac{\mu}{L}\right)^k (f(x_0) - \min_{x'} f(x')).$$

(d) (5pts) Suppose instead of (1), we had $\|\nabla f(x)\|_2 \geq \mu(f(x) - \min_{x'} f(x'))$. Is it still true that $\lim_{k\to\infty} f(x_k) = \min f(x)$? If so, derive a convergence rate bounding $f(x_k) - \min f(x)$.

Q5. (Accelerated Gradient Norm Convergence Rates, 20pts) Recall in lecture, we showed after $k > 0$ steps, the accelerated gradient method for minimizing a convex function $f\colon \mathbb{R}^d \to \mathbb{R}$ with L-Lipschitz gradient has

$$f(y_k) - f(x^*) \leq \frac{2L\|x_0 - x^*\|^2}{k^2}$$

where x^* is any minimizer of f.

(a) (10pts) Prove that this y_k has gradient norm bounded by

$$\|\nabla f(y_k)\|_2 \leq C/k$$

for some constant C depending on L and $\|x_0 - x^*\|_2$.

(b) (10pts) Consider the following combination of the accelerated method and gradient descent for minimizing a function f with L-Lipschitz gradient: For a fixed number $k > 0$, first run k steps of the accelerated method from a given x_0. Then using the final iterate of the accelerated method y_k as an initial point $\bar{x}_0 \leftarrow y_k$, run k steps of gradient descent $\bar{x}_{i+1} = \bar{x}_i - \nabla f(\bar{x}_i)/L$.

Show the improved guarantee that some $\bar{x}_i$ with $0 \leq i < k$ of this gradient descent run has

$$\|\nabla f(\bar{x}_i)\|_2 \leq C/k^{3/2}.$$