Convex Programming
Problem Set 5 – CS 6515/4540 (Fall 2025)

This problem set is due on **Tuesday November 4th**. Submission is via Gradescope. Your solution must be a typed pdf (e.g. via LaTeX) – no handwritten solutions.

18 Convex Functions

1. Given two convex functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, is the sum $f + g$ also convex? Either prove it or give a counterexample.
2. Given two convex functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, is the product fg also convex? Either prove it or give a counterexample.
3. What about the convexity of the function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(x) = \max(f(x), g(x))$? Prove it or give a counterexample.
4. Show an α-strongly convex function (defined in Pb 20) $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(x) = \Omega(x^2)$

19 Gradient Descent Failure

Suppose we run an unconstrained Gradient Descent on $f(x) = \frac{1}{2}x^2$ with some arbitrary step size. Give (and justify) an example consisting of

1. a step size $\eta > 0$
2. an initial point $x_0 \in \mathbb{R}$

such that t-th **average** $\overline{x}_t = \frac{1}{t}\sum_{i \le t} x_i$ of an unconstrained Gradient Descent with the above parameters does not converge (e.g it diverges) to the optimum as $t \to \infty$.

20 Gradient Descent for Strongly-Convex Functions

A differentiable function f is α-strongly convex for $\alpha > 0$ if for all $x, y \in \mathbb{R}^n$ we have

$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|_2^2.$$

Consider an α-strongly convex differentiable function f with the 2-norm of its gradient always bounded by G. The goal is to minimize f and let x^* denote its minimum.

Show that the gradient descent algorithm with step size $\frac{1}{\alpha(t+1)}$ satisfies

$$f\Big(\frac{\sum_t x_t}{T}\Big) - f(x^*) \le \frac{G^2(1 + \log T)}{2\alpha T}.$$

Thus, strong-convexity allows us to get $1/T$ dependency in regret instead of $1/\sqrt{T}$ dependency for general convex functions.

(Hint: Change the potential function in the analysis from class to $\Phi(t) = \frac{t\alpha}{2}\|x_t - x^*\|^2$. Also, use that $\sum_{t \in \{1,\dots,T\}} \frac{1}{t} \le 1 + \log T$.)

21 Non-Convex Function

In class we assumed function f is convex. We now want to consider the non-convex case. We want to show that for L-smooth f, after t iterations with step size $\eta \leq 1/L$ we can find a point x' with

$$\|\nabla f(x')\| \leq \sqrt{\frac{2}{\eta \cdot t}(f(x^0) - f(x^*))}.$$

(Note that for a local optimum we have $\nabla f(x) = 0$, so a small norm $\|\nabla f(x')\|$ indicates that we are close to a local optimum or saddle point.)

Proving this from scratch is a bit tricky, so we provide the following subproblems to guide you to a proof. Each subproblem can be solved in a few lines of calculation/algebra.

Problem:

1. Show $f(x^{t+1}) \leq f(x^t) - \frac{\eta}{2}\|\nabla f(x^t)\|^2$ (Hint: Check the proof from class for convex functions. Does it work for non-convex functions?)
2. Show $\sum_{k=0}^{t} \|\nabla f(x^k)\|^2 \leq \frac{2}{\eta}(f(x^0) - f(x^*))$ (Hint: 1. implies $\frac{\eta}{2}\|\nabla f(x^t)\|^2 \leq ...$)
3. Show $\min_{k=0...t} \|\nabla f(x^k)\| \leq \sqrt{\frac{2}{\eta \cdot t}(f(x^0) - f(x^*))}$.

where x^* is the global optimum $f(x^*) = \min_x f(x)$.

You are allowed to use subproblems to solve later subproblems (e.g., use 1+2 to solve 3), even if you did not prove them.