Continuous Optimization **Spring 2025**

Homework 1: Due 24/03/2025 before 23.55

Lecturer: Aurelien Lucchi

Problem 1 (Convex Sets, 10 points):

a. Consider the set (3 Pts)

$$S = \Big\{(x, y) \in \mathbb{R}^2 \;\Big|\; x \geq 0,\ x^2 + y^2 \leq 4x\Big\}.$$

Show whether S is convex or not. (Hint: notice that $x^2 + y^2 \leq 4x$ can be rewritten as a circle shifted from the origin.)

b. Define (3 Pts)

$$T = \{\,(x_1, x_2, x_3) \in \mathbb{R}^3 \;\mid\; (x_1, x_2) \in \operatorname{conv}\{(0,0),(2,0),(0,2)\},\ x_3 \geq x_1 + x_2\}.$$

Prove that T is convex.

c. Let (4 Pts)

$$U := \{\,(x, y) \in \mathbb{R}^2 : \min\{x^2 + y^2,\, (x-1)^2 + y^2\} \leq 1\}.$$

Show that U is the union of two disks of radius 1, centered at $(0, 0)$ and $(1, 0)$. Check whether this union is convex or not.

Problem 2 (Function Regularity and the PL Inequality, 10 points):

a. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by (4 Pts)

$$f(x) \;=\; \tfrac{x^4}{4} \;-\; x^2.$$

i) *(Smoothness)* Is f L-smooth on all of $\mathbb{R}$? If yes, determine such an L. If no, explain which calculations imply that no global L can exist.

ii) *(PL Inequality)* The Polyak-Łojasiewicz (PL) condition says there is a $\mu > 0$ such that

$$\tfrac{1}{2}\left|f'(x)\right|^2 \;\geq\; \mu\big(f(x) - f^*\big) \quad \text{for all } x \in \mathbb{R},$$

where $f^* = \min_{x\in\mathbb{R}} f(x)$. Find f^* by direct calculation (use $f'(x) = 0$). Then, check whether you can identify a positive μ that satisfies the PL condition for every $x \in \mathbb{R}$. Make sure you include explicit computations of both sides to conclude whether such a μ exists or not.

b. *(Example of PL without Strong Convexity)* (3 Pts)
Define $g : \mathbb{R}^n \to \mathbb{R}$ by

$$g(\mathbf{x}) \;=\; \sum_{i=1}^{n} \max\{\,0,\, x_i\}^2.$$

(i) Show that g is *not* strongly convex: for example, any vector $\mathbf{x}$ with all $x_i \leq 0$ has $g(\mathbf{x}) = 0$, so there is a whole "flat" (plateau) region where g is constant. This violates the strict curvature needed for strong convexity.

(ii) Verify that g *does* satisfy the PL inequality

$$\tfrac{1}{2}\,\|\nabla g(\mathbf{x})\|^2 \;\geq\; \mu\left(g(\mathbf{x}) \;-\; g^*\right),$$

for some $\mu > 0$, where $g^* = \min_{\mathbf{x}} g(\mathbf{x}) = 0$.

c. *(Linear Convergence via PL)* (3 Pts)
Suppose that $g(\mathbf{x})$ in part (b) is also L-smooth (or pick another suitable g that is both PL and L-smooth). With the standard gradient descent update

$$\mathbf{x}^{k+1} \;=\; \mathbf{x}^k \;-\; \alpha\,\nabla g(\mathbf{x}^k),$$

demonstrate, by explicit calculation, that for a proper step size α there is a linear rate of convergence:

$$g(\mathbf{x}^k) - g^* \;\;\leq\;\; \rho^{\,k}\left(g(\mathbf{x}^0) - g^*\right), \quad \text{for some } \rho \in (0, 1).$$

Make sure to use both L-smoothness and the PL condition in your derivation.

Problem 3 (Convergence of Gradient Descent, 10 points):

In this problem, we assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is differentiable, L-smooth and μ-strongly convex. Let us define the Bregman divergence associated with the function f as follows

$$D_f(\mathbf{x}, \mathbf{y}) \coloneqq f(\mathbf{x}) - f(\mathbf{y}) - \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle . \tag{1}$$

Bregman divergence is useful in deriving convergence guarantees of gradient descent with tighter dependencies on the problem-specific constants (e.g., L or μ). In particular, following (6) from Lecture 2: Gradient Descent the smoothness, the L-smoothness and μ-strong convexity imply

$$\frac{\mu}{2}\|\mathbf{x} - \mathbf{y}\|^2 \leq D_f(\mathbf{x}, \mathbf{y}) \leq \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2. \tag{2}$$

Moreover, a convex and L-smooth function satisfies

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \leq 2L D_f(\mathbf{x}, \mathbf{y}). \tag{3}$$

We will prove the following convergence rate for the gradient descent which is equivalent to Theorem 1 from Lecture 2: Gradient Descent using Bregman divergence

$$\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq (1 - \eta\mu)^k \|\mathbf{x}_0 - \mathbf{x}^*\|^2, \tag{4}$$

where the stepsize $\eta \leq \frac{1}{L}$.

a. Using (2) and definition of Bregman divergence (1) demonstrate that the following inequality holds (2 Pts)

$$\frac{\mu}{2}\|\mathbf{x} - \mathbf{y}\|^2 + D_f(\mathbf{x}, \mathbf{y}) \leq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle .$$

b. Demonstrate that the following equality holds (1 Pt)

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 = \|\mathbf{x}_k - \mathbf{x}^*\|^2 - 2\eta \langle \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}^*), \mathbf{x}_k - \mathbf{x}^* \rangle + \eta^2 \|\nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}^*)\|^2.$$

c. Using the result of 1. demonstrate that (2 Pts)

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 \leq (1 - \eta\mu)\|\mathbf{x}_k - \mathbf{x}^*\|^2 - 2\eta D_f(\mathbf{x}_k, \mathbf{x}^*) + \eta^2 \|\nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}^*)\|^2.$$

d. Using (3) demonstrate that (2 Pts)

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 \leq (1 - \eta\mu)\|\mathbf{x}_k - \mathbf{x}^*\|^2 - 2\eta(1 - \eta L) D_f(\mathbf{x}_k, \mathbf{x}^*).$$

e. Using the stepsize restriction derive the recursion (2 Pts)

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 \leq (1 - \eta\mu)\|\mathbf{x}_k - \mathbf{x}^*\|^2.$$

f. Unrolling the recursion obtain the final convergence result (1 Pt)

$$\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq (1 - \eta\mu)^k \|\mathbf{x}_0 - \mathbf{x}^*\|^2.$$