ORIE 6365 — Continuous Optimization: Algorithms and Complexity, Spring 2026

Homework 1

Released: Feb 3, 2026 *Due: Feb 21, 2026*

Problem 1. Let $\|\cdot\|$ be an arbitrary fixed norm on $\mathbb{R}^n$. Consider the function, $\rho : \mathbb{R}^n \to \mathbb{R}$, defined by

$$\rho(s) \quad := \quad \max_{x \in \mathbb{R}^n \,:\, \|x\| \leq 1} \langle s, x \rangle.$$

- Show that $\rho(s) = \max_{x \in \mathbb{R}^n \,:\, \|x\|=1} \langle s, x \rangle$, i.e. the maximum of the linear form $\langle s, \cdot \rangle$ over the unit ball is always achieved on its boundary (the unit sphere).

- Show that $\rho(s)$ satisfies the standard axioms of a norm:

  1. **Positive definiteness:** $\rho(s) \geq 0$ for all $s \in \mathbb{R}^n$, and $\rho(s) = 0 \;\Leftrightarrow\; s = 0$.
  2. **Homogeneity:** $\rho(ts) = |t|\rho(s)$ for any $s \in \mathbb{R}^n$ and $t \in \mathbb{R}$.
  3. **Triangle inequality:** $\rho(s + y) \leq \rho(s) + \rho(y)$ for all $s, y \in \mathbb{R}^n$.

  Therefore, the *dual norm* $\|s\|_* := \rho(s)$ is a well-defined norm on $\mathbb{R}^n$.

Problem 2. Consider the logistic loss function, $\ell(t) = \ln(1 + e^t)$ for $t \in \mathbb{R}$. Show that $\ell(\cdot)$ is convex and that its derivative $\ell'(\cdot)$ is Lipschitz continuous on $\mathbb{R}$. Compute the smallest possible Lipschitz constant $L > 0$ (w.r.t. the standard Euclidean norm, i.e., the absolute value $|\cdot|$ on $\mathbb{R}$).

Problem 3. Consider the following function, often called *soft-max* or *log-sum-exp*:

$$f(x) \quad = \quad \ln\left(\sum_{i=1}^{n} \exp(x^{(i)}) \right), \qquad x \in \mathbb{R}^n.$$

Show that $0 \preceq \nabla^2 f(x) \preceq I$, for any $x \in \mathbb{R}^n$, where $I \in \mathbb{R}^{n \times n}$ is the identity matrix; i.e., all eigenvalues of $\nabla^2 f(x)$ lie in the interval $[0, 1]$. Therefore, f is convex and has a Lipschitz continuous gradient with constant $L = 1$ (w.r.t. the Euclidean norm).

Problem 4. Let $B = B^\top \succ 0$ be a symmetric positive definite matrix, $B \in \mathbb{R}^{n \times n}$. Consider the following generalized Euclidean norm: $\|x\| := \langle Bx, x \rangle^{1/2}$ for $x \in \mathbb{R}^n$.

- Show that the dual norm is given by $\|s\|_* = \langle s, B^{-1}s \rangle^{1/2}$ for $s \in \mathbb{R}^n$.

- Provide an explicit formula for the step x^+ of the gradient method with respect to this norm, for $M > 0$:

$$x^+ \quad := \quad \arg\min_{y \in \mathbb{R}^n} \Big\{ f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{M}{2}\|y - x\|^2 \Big\}.$$

Problem 5. Let $f : \mathbb{R}^n \to \mathbb{R}$ have a Lipschitz continuous gradient with constant $L_f > 0$ with respect to the standard Euclidean norm in $\mathbb{R}^n$. Let $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$ be fixed parameters, and define the function (an affine change of variable):

$$F(y) \quad = \quad f(Ay + b), \qquad y \in \mathbb{R}^m.$$

- Consider the standard Euclidean norm in $\mathbb{R}^m$, express the Lipschitz constant L_F of the gradient ∇F in terms of L_f and given parameters.
- Fix $B = A^\top A$ and assume $B \succ 0$. Consider the generalized Euclidean norm in $\mathbb{R}^m$ defined by $\|y\| := \langle By, y\rangle^{1/2}$. Show that the Lipschitz constant of ∇F with respect to this norm is exactly L_f.

Problem 6. Consider $f : \mathbb{S}^n \to \mathbb{R}$ where $\mathbb{S}^n = \{X \in \mathbb{R}^{n\times n} : X = X^\top\}$ is the space of symmetric $n \times n$ matrices. We equip $\mathbb{S}^n$ with the spectral norm:

$$\|X\| \quad := \quad \max_{u\in\mathbb{R}^n : \|u\|_2 = 1} |\langle Xu, u\rangle| \quad = \quad \max_{1\le i\le n} |\lambda^{(i)}(X)| \quad =: \quad \|\lambda(X)\|_\infty, \qquad X \in \mathbb{S}^n,$$

where $\lambda(X) = (\lambda^{(1)}(X), \ldots, \lambda^{(n)}(X)) \in \mathbb{R}^n$ is the vector of eigenvalues of X.

- Compute the dual norm $\|Y\|_* := \max\limits_{X\in\mathbb{S}^n : \|X\|\le 1} \langle Y, X\rangle$, where $\langle Y, X\rangle = \mathrm{tr}(YX)$ is the standard inner product on $\mathbb{S}^n$.
- Provide an explicit formula for a step X^+ of the gradient method with respect to this norm, for $M > 0$:

$$X^+ \quad := \quad \arg\min_{Y\in\mathbb{S}^n} \Big\{ f(X) + \langle \nabla f(X), Y - X\rangle + \tfrac{M}{2}\|Y - X\|^2 \Big\}.$$

Problem 7. Fix the standard Euclidean norm $\|\cdot\|_2$ on $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function and denote by $L > 0$ the Lipschitz constant of $\nabla f(\cdot)$. Assume f is bounded from below: $f^\star := \inf\limits_{x\in\mathbb{R}^n} f(x) > -\infty$. Let $e_1, \ldots, e_n \in \mathbb{R}^n$ be the standard basis. Consider the following randomized algorithm for finding a stationary point of f:

Algorithm: *Coordinate Descent.*

> **Initialization:** $x_0 \in \mathbb{R}^n$, the Lipschitz constant $L > 0$, number of iterations $K \ge 1$.
> **For** $k = 0 \ldots K-1$ **iterate:**
> 1. Sample coordinate $i_k \in \{1, \ldots, n\}$ uniformly at random.
> 2. Compute the i_k-th partial derivative and form the vector $g_k := \frac{\partial f}{\partial x^{(i_k)}}(x_k) \cdot e_{i_k} \in \mathbb{R}^n$.
> 3. Update $x_{k+1} := x_k - \frac{1}{L} g_k$.
>
> Sample $j \in \{0, \ldots, K-1\}$ uniformly at random and **return** $\bar{x}_K := x_j$.

We denote by $\bar{x}_K$ the result of the algorithm after running for $K \ge 1$ iterations. Note that $\bar{x}_K$ is a random vector. We let $\mathbb{E}\big[\cdot\big]$ denote the expectation with respect to all randomness in the method.

- Show the following progress of each step, $0 \leq k \leq K-1$:

$$\mathbb{E}\Big[f(x_k) - f(x_{k+1}) \,\big|\, x_k \Big] \quad \geq \quad \tfrac{1}{2nL} \|\nabla f(x_k)\|_2^2,$$

  where $\mathbb{E}\big[\cdot \,|\, x_k \big]$ is the conditional expectation, when point x_k is fixed.

- Show that

$$\mathbb{E}\Big[\|\nabla f(\bar{x}_K)\|_2^2 \Big] \quad \leq \quad \tfrac{2nL(f(x_0) - f^\star)}{K}.$$

- Show that for a given $\varepsilon > 0$, it is enough to set $K := \big\lfloor \frac{2nL(f(x_0)-f^\star)}{\varepsilon^2} \big\rfloor + 1$ in order to obtain $\mathbb{E}\big[\|\nabla f(\bar{x}_K)\|_2\big] \leq \varepsilon$. Compare this complexity with that one for the standard gradient descent.

Problem 8. Assume that the gradient $\nabla f(\cdot)$ of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is Hölder continuous of degree $0 \leq \nu \leq 1$ with constant $H_\nu > 0$, with respect to the standard Euclidean norm:

$$\|\nabla f(y) - \nabla f(x)\|_2 \quad \leq \quad H_\nu \|y - x\|_2^\nu, \qquad x, y \in \mathbb{R}^n.$$

- Show that, for any $x, y \in \mathbb{R}^n$:

$$f(y) \quad \leq \quad f(x) + \langle \nabla f(x), y - x \rangle + \tfrac{H_\nu}{1+\nu} \|y - x\|_2^{1+\nu}. \tag{1}$$

- Consider the method based on minimizing the global upper bound in (1), starting from some initialization $x_0 \in \mathbb{R}^n$, for $k \geq 0$:

$$x_{k+1} \quad = \quad \arg\min_{y \in \mathbb{R}^n} \Big[f(x_k) + \langle \nabla f(x_k), y - x_k \rangle + \tfrac{H_\nu}{1+\nu} \|y - x_k\|_2^{1+\nu} \Big]. \tag{2}$$

  Show that each step can be written in the form $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$ and provide an explicit formula for the step-size $\alpha_k > 0$.

- What is the first-order oracle complexity of this method to find a stationary point $\bar{x}$ such that $\|\nabla f(\bar{x})\|_2 \leq \varepsilon$, for $0 < \nu < 1$?