MIT 6.7220/15.084 — Nonlinear Optimization Spring 2025

Problem Set 1

Due: Feb 18th 2025 11:59pm ET

Collaboration policy. We encourage working together whenever possible: in the recitations, problem sets, and general discussion of the material and assignments. Keep in mind, however, that for the problem sets the solutions you hand in should reflect your own understanding of the class material, and should be written solely by you. It is not acceptable to copy (in whole or in part) a solution that somebody else has written.

1. Normal cone for the unit ball [20pts]

(1.a) [10pts] Define the n-dimensional unit ball $\mathbb{B}_n := \{x \in \mathbb{R}^n : \|x\|_2 \le 1\}$. For any point $\overline{x} \in \mathbb{B}_n$, find the normal cone of $\mathbb{B}_n$ at $\overline{x}$ (please give a rigorous proof).

> *Solution.* If $\|\overline{x}\|_2 < 1$, then $\mathcal{N}_{\mathbb{B}_n}(\overline{x}) = \{0\}$ since $\overline{x}$ is in the interior.
>
> If $\|\overline{x}\|_2 = 1$, we prove that $\mathcal{N}_{\mathbb{B}_n}(\overline{x}) = \{c\overline{x} : c \ge 0\}$.
>
> - $\mathcal{N}_{\mathbb{B}_n}(\overline{x}) \subseteq \{c\overline{x} : c \ge 0\}$: Pick any $z \in \mathcal{N}_{\mathbb{B}_n}(\overline{x})$. If $z = 0$, the statement is clearly true. So, assume $z \ne 0$. Since $z \in \mathcal{N}_{\Omega}(\overline{x})$, it must satisfy
>
> $$\langle z, x - \overline{x}\rangle \le 0$$
>
>   for all $x \in \mathbb{B}_n$. In particular, consider the point $x := \frac{\overline{x}+\frac{z}{\|z\|_2}}{2}$, then we have $\|x\|_2 \le \frac{\|\overline{x}\|_2+\frac{\|z\|_2}{\|z\|_2}}{2} = 1$, so $x \in \mathbb{B}_n$. From $z \in \mathcal{N}_{\mathbb{B}_n}(\overline{x})$, we must have $\langle z, x - \overline{x}\rangle \le 0$. And hence,
>
> $$\begin{aligned} 0 &\ge \langle z, x - \overline{x}\rangle \\ &= \frac{1}{2}\left\langle z, \frac{z}{\|z\|_2} - \overline{x}\right\rangle \\ &= \frac{1}{2}(\|z\|_2 - \langle z, \overline{x}\rangle) \\ &= \frac{\|z\|_2}{2}\left(1 - \left\langle \frac{z}{\|z\|_2}, \overline{x}\right\rangle\right) \\ &= \frac{\|z\|_2}{4}\left(\left\|\frac{z}{\|z\|_2}\right\|_2^2 - 2\left\langle \left\|\frac{z}{\|z\|_2}\right\|_2, \overline{x}\right\rangle + \|\overline{x}\|_2^2\right) \\ &= \frac{\|z\|_2}{4}\left\|\frac{z}{\|z\|_2} - \overline{x}\right\|^2. \end{aligned}$$
>
>   However, notice that this inequality holds true only for $\frac{z}{\|z\|_2} = \overline{x}$. Hence $z = \|z\|_2\overline{x}$ and therefore, $z \in \{c\overline{x} : c \ge 0\}$. Note that 4th equality follows from $\left\|\frac{z}{\|z\|_2}\right\|_2 = \|x\|_2 = 1$
>
> - $\mathcal{N}_{\mathbb{B}_n}(\overline{x}) \supseteq \{c\overline{x} : c \ge 0\}$: For any $z = c\overline{x}$ with $c \ge 0$ and any $x \in \mathbb{B}_n$, we have
>
> $$\begin{aligned} \langle z, x - \overline{x}\rangle &= c(\langle \overline{x}, x\rangle - \|\overline{x}\|_2^2) \\ &\le \frac{c}{2}(2\langle \overline{x}, x\rangle - \|\overline{x}\|_2^2 - \|x\|_2^2) \\ &= -\frac{c}{2}\|x - \overline{x}\|_2^2 \le 0, \end{aligned}$$
>
>   where we have used $\|x\|_2 \le 1 = \|\overline{x}\|_2^2$ in the first inequality. As a result,

$z \in \mathcal{N}_{\mathbb{B}_n}(\overline{x})$.

◂

(1.b) [10pts] Find the solution of the problem

$$\begin{aligned} \min_{x_1, x_2} \quad & f(x) := \frac{1}{2}(x_1^2 - x_2^2) + 2\sqrt{3}x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \le 1 \end{aligned}$$

Solution. Let $x^* = (x_1^*, x_2^*)$ be the minimizer, then by the optimality condition,

$$-\nabla f(x^*) = \begin{pmatrix} -x_1^* - 2\sqrt{3} \\ x_2^* - 1 \end{pmatrix}$$

lies in the normal cone of $\mathbb{B}_2$. If x^* is in the interior of $\mathbb{B}_2$, then $\nabla f(x^*) = 0$, which implies $x_1^* = -2\sqrt{3}$ and $x_2^* = 1$, but this contradicts $x^* \in \mathbb{B}_2$.

If x^* is on the boundary of $\mathbb{B}_2$, then from the previous subproblem there exists some $\lambda \ge 0$ such that $-\nabla f(x^*) = \lambda x^*$, which implies

$$x_1^* = -\frac{2\sqrt{3}}{\lambda + 1}, \qquad x_2^* = -\frac{1}{\lambda - 1}.$$

Since x^* is on the boundary of $\mathbb{B}_2$, we have

$$\frac{12}{(\lambda+1)^2} + \frac{1}{(\lambda-1)^2} = 1 \tag{1}$$

Let $g(t) := \frac{12}{(t+1)^2} + \frac{1}{(t-1)^2}$. Note that for any $t \in [0, 1)$, by Cauchy-Schwarz inequality,

$$g(t) \ge 2\sqrt{\frac{12}{(t^2-1)^2}} = \frac{4\sqrt{3}}{|t^2 - 1|} > 1$$

For $t > 1$, $g(t)$ is decreasing as t increases. We can observe that $\lambda = 3$ is the only value for which (1) holds true. As a result, $x_1^* = -\frac{\sqrt{3}}{2}$ and $x_2^* = -\frac{1}{2}$. ◂

2. Existence of entropy-regularized best responses [20pts]

Let $\mathring{\Delta}^n := \{x \in \mathbb{R}^n_{>0} : x_1 + \ldots + x_n = 1\}$ be the set of probability distributions over n actions that put *strictly positive* mass on each action.

Let $v \in \mathbb{R}^n$ be a given vector with entries in some interval $[-C, C]$, and consider the optimization problem

$$\begin{aligned} \min_x \ & \textstyle\sum_{i=1}^n v_i x_i + \sum_{i=1}^n x_i \log x_i \\ \text{s.t.} \ & x \in \mathring{\Delta}^n. \end{aligned}$$

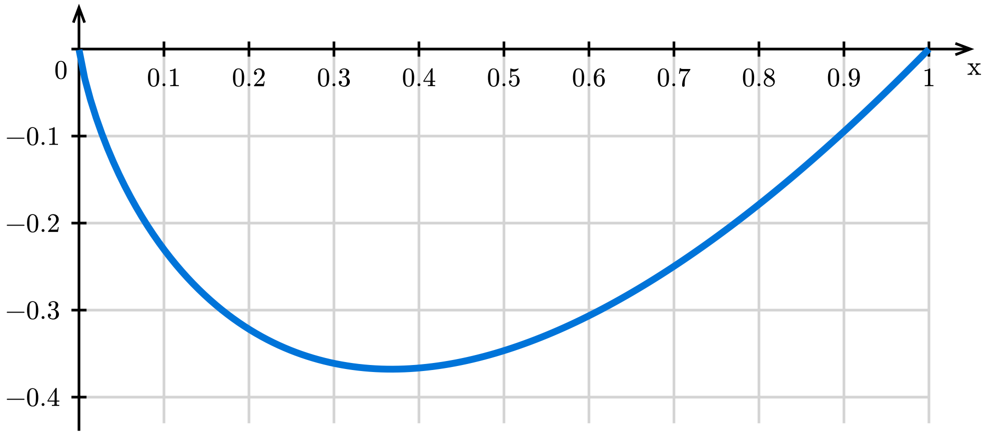
(2.a) [2pts] Is the set $\mathring{\Delta}^n$ closed? bounded? convex? (Just give a yes/no answer, no need to prove anything.)

> *Solution.* The set is not closed. It is bounded. It is convex. ◂

(2.b) [18pts] Prove that the optimization problem above has a solution.

▸ *Hint:* Consider the extension of the objective to the closed set $\Delta^n : \{x \in \mathbb{R}^n_{\geq 0} : x_1 + \ldots + x_n = 1\}$. Is the objective continuous? Can you use the Weierstrass theorem here? Then, argue that the minimizer cannot be on the boundary.

For reference, this is what the function $x \log x$ looks like on [0,1]:



In particular, the function is arbitrarily steep around 0. Because of this, if the minimizer had some $x_i = 0$, what would happen if you added a tiny bit of mass ϵ to x_i and removed it from some action with a lot of mass?

> *Solution.* Denote by $f(x) = \sum_{i=1}^n v_i x_i + \sum_{i=1}^n x_i \log x_i$ the objective function. We have that f is continuous over $\mathring{\Delta}^n$. Since $\lim_{x \to 0} x \log x = 0$, with the convention $0 \log 0 := 0$, the domain of definition of f can be extended to Δ^n such that f remains continuous over Δ^n.
>
> According to the Weierstrass theorem, since Δ^n is nonempty, closed, and bounded, a minimizer of f exists within Δ^n.

Thus, it suffices to show that no minimizer exists on the boundary $\partial\Delta^n$:

$$\partial\Delta^n = \{x \in \Delta^n : \exists i, x_i = 0\}.$$

Take $x \in \partial\Delta^n$ with $x_i = 0$. There always exists j with $x_j > 0$. Consider $x' \in \Delta^n$ defined as $x'_i = \epsilon$, $x'_j = x_j - \epsilon$ for some constant $0 < \epsilon < x_j$, and $x'_k = x_k$ for any other k.

We will show $f(x) > f(x')$ when ϵ is sufficiently small. By calculation,

$$\begin{aligned}
&f(x) - f(x') \\
&= v_i(x_i - x'_i) + v_j(x_j - x'_j) + (x_i \log x_i - x'_i \log x'_i) + (x_j \log x_j - x'_j \log x'_j) \\
&= (v_i - v_j)\epsilon - \epsilon \log \epsilon + x_j \log x_j - (x_j - \epsilon)\log(x_j - \epsilon) \\
&= (v_i - v_j + \log(x_j) + 1)\epsilon - \epsilon \log \epsilon + o(\epsilon),
\end{aligned}$$

where the last equality follows from Taylor expansion.

By the definition of $o(\cdot)$, there exists some constant $c > 0$ such that

$$(v_i - v_j + \log(x_j) + 1)\epsilon + o(\epsilon) > -c\epsilon.$$

For $0 < \epsilon < \min\{x_j, e^{-c}\}$, we obtain

$$f(x) - f(x') > -c\epsilon - \epsilon \log \epsilon > 0,$$

indicating x cannot be the minimizer in Δ^n. In general, we conclude that $\mathring{\Delta}^n$ contains a minimizer. ◀

3. The Rayleigh quotient [25pts]

(3.a) [1pts] Let $\Omega := \mathbb{R}^n \setminus \{0\}$, that is, the entire space without the origin. Is this set open? closed? convex? bounded? (Just give a yes/no answer, no need to prove anything.)

> *Solution.* The set is open, not closed, not convex, and not bounded. ◂

(3.b) [5pts] Let the function $f : \Omega \to \mathbb{R}$ be continuous, satisfying $f(\lambda x) = f(x)$ for all $\lambda > 0$ in $\mathbb{R}$ and nonzero $x \in \mathbb{R}^n$. Prove that f has a minimizer.

> *Solution.* Consider the restriction of the function f on the *surface* of the unit ball $C := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. Since C is nonempty, closed, and bounded, the function f has a minimizer $x^* \in C$. Since for all $x \in C$, $f(x) \geq f(x^*)$ and that for all $y \in \mathbb{R}^n \setminus \{0\}$, there exists an $x \in C$ such that $f(y) = f(x)$, we can conclude that $f(y) = f(x) \geq f(x^*)$, implying that x^* is a global minimizer of f. ◂

(3.c) [5pts] Denote with $\mathbb{S}^n$ the set of symmetric $n \times n$ matrices with real entries. Given a matrix $A \in \mathbb{S}^n$, define

$$g : \Omega \to \mathbb{R}, \qquad g(x) := \frac{x^\top A x}{\|x\|_2^2}.$$

Prove that g has a minimizer.

> *Solution.* We have $g(\lambda x) = \frac{(\lambda x)^\top A(\lambda x)}{(\lambda x)^\top(\lambda x)} = \lambda^2 \frac{x^\top A x}{\lambda^2 x^\top x} = \frac{x^\top A x}{x^\top x} = g(x)$ so it is covered by the Problem (3.b) as the domain of g is the non zero reals. ◂

(3.d) [5pts] Calculate $\nabla g(x)$ for nonzero x.

> *Solution.* First, we note that $\nabla(x^\top A x) = (A + A^\top)x = 2Ax$, last step is due to the symmetry of A.
>
> $$\begin{aligned} \nabla g(x) &= \frac{1}{\|x\|_2^4}(\|x\|_2^2 \nabla(x^\top A x) - (x^\top A x)\nabla(\|x\|_2^2)) \\ &= \frac{2}{\|x\|_2^4}(\|x\|_2^2 A x) - (x^\top A x)x)) \\ &= \frac{1}{\|x\|_2^2}(2Ax - 2g(x)x). \end{aligned}$$
>
> ◂

(3.e) [5pts] Deduce that the minimizers of g must be eigenvectors, and calculate the minimum value.

> *Solution.* Since $\|x\|_2 \neq 0$ then first order optimality conditions in the open domain imply $\nabla g(x) = 0$ which implies that $Ax = g(x)x$ by Problem (3.d). So, x is an eigenvector of A corresponding to the eigenvalue $g(x)$. In other words, the first-order optimality conditions are saying that only eigenvectors can possibly be minimizers. On the other hand, if x is an eigenvector of A, say for some eigenvalue λ, then
>
> $$g(x) = \frac{x^\top Ax}{\|x\|_2^2} = \frac{\lambda x^\top x}{\|x\|_2^2} = \lambda.$$
>
> So, only the eigenvectors for the minimum eigenvalue $\lambda_{\min}$ of A can be minimizers, and they all would guarantee the same value to g, equal to $\lambda_{\min}$. Since g attains a minimum, it follows that the minimum of g is equal to $\lambda_{\min}$, and that the minimizers of g are exactly the eigenvectors for $\lambda_{\min}$. ◂

(3.f) [2pts] Recall that a real symmetric matrix A is said to be *positive semidefinite* when all of its eigenvalues are nonnegative. Conclude from Problem (3.e) that a matrix is positive semidefinite if and only if $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^n$.

> *Solution.* If $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^n$ then surely $g(x) \geq 0$ for all $x \in \Omega$. Since we established in Problem (3.e) that $\min_{x\in\Omega} g(x)$ is the smallest eigenvalue of A, then all of the eigenvalues of A must be nonnegative.
>
> The converse is also direct. Indeed, if all eigenvalues of A are nonnegative, then $\min_{x\in\Omega} g(x) \geq 0$ by Problem (3.e). This implies that $g(x) \geq 0$ for all $x \in \mathbb{R}^n$, which in turn implies that $x^\top Ax \geq 0$ for all $x \in \Omega$. The only case left to check is that $x^\top Ax$ for $x = 0$, which is trivial. ◂

(3.g) [2pts] Find an alternate proof of Problem (3.e) using a spectral decomposition of the symmetric matrix A.

▸ *Hint:* In a precise sense, the spectral decomposition allows you to pretend, without loss of generality, that A is a *diagonal* matrix (*i.e.*, that A has zeros everywhere outside of the diagonal). Why is that?

> *Solution.* Since A is symmetric, the real spectral theorem guarantees that it can be written in the form
>
> $$A = Q\Lambda Q^\top,$$
>
> where Q is an *orthogonal* matrix of eigenvectors (disposed on the columns), and Λ is a diagonal matrix of eigenvalues. Crucially, the matrix Q is orthog-

onal, meaning that Q and $Q^\top$ preserve norms and are each the inverse of the other. Hence,

$$\frac{x^\top A x}{\|x\|_2^2} = \frac{x^\top Q \Lambda Q^\top x}{\|x\|_2^2} = \frac{(Q^\top x)^\top \Lambda (Q^\top x)}{\|x\|_2^2} = \frac{(Q^\top x)^\top \Lambda (Q^\top x)}{\|Q^\top x\|_2^2}.$$

Since Q is invertible, the transformation $x \mapsto Q^\top x =: y$ is a bijection of Ω (indeed, note that $Q^\top x = 0 \iff x = 0$ since $Q^\top$ is invertible), and we can perform a change of variables:

$$\min_{x \in \Omega} \frac{x^\top A x}{\|x\|_2^2} = \min_{y \in \Omega} \frac{y^\top \Lambda y}{\|y\|_2^2}.$$

Letting $\lambda_1, ..., \lambda_n \in \mathbb{R}$ be the eigenvalues of A, that is, the diagonal entries of Λ, we can then rewrite the objective as

$$\frac{y^\top \Lambda y}{\|y\|_2^2} = \sum_{i=1}^{n} \left(\frac{y_i^2}{\sum_j y_j^2} \right) \lambda_i.$$

The coefficients $\frac{y_i^2}{\sum_j y_j^2}$ are nonnegative and sum to 1, and therefore it is immediate that

$$\frac{y^\top \Lambda y}{\|y\|_2^2} \geq \min_{i=1}^{n} \{\lambda_i\},$$

holding with equality if and only if y is supported only on whatever indices are associated with the minimum eigenvalue. Inverting the change of variables, we obtain that the optimal x is a linear combination of eigenvectors for the minimum eigenvalue, meaning that any optimal x is itself an eigenvector for the minimum eigenvalue of A. ◀

4. Convex sets [15pts]

(4.a) [5pts] Show that the intersection of two convex sets is convex. Use this fact to conclude that any nonempty set of the form

$$\Omega := \left\{x \in \mathbb{R}^d : \langle a_i, x\rangle \leq b_i \quad \forall i = 1, ..., k\right\}$$

is convex, no matter the choice of $k \in \mathbb{N}$, real vectors $a_i \in \mathbb{R}^d$, and scalars $b_i \in \mathbb{R}$.

> *Solution.* Let $S = S_1 \cap S_2$ be the intersection of two convex sets. If $x, y \in S$ then by definition $x, y \in S_i$ for $i \in \{1, 2\}$. By convexity of S_1, S_2, this implies that for all $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda)y \in S_i$ for $i = 1, 2$. Hence, $\lambda x + (1 - \lambda)y \in S_1 \cap S_2 = S$.
>
> By induction, this shows that the intersection of any *finite* number of convex sets is itself convex.
>
> Given that the halfspace $\Omega_i = \left\{x \in \mathbb{R}^d : \langle a_i, x\rangle \leq b_i\right\}$ is convex no matter the choice of vector a_i and scalar b_i, any finite intersection of halfspaces $\Omega = \bigcap_{i\in[n]} \Omega_i$ is convex. ◂

(4.b) [5pts] Show that the set of all positive semidefinite matrices is convex.

▸ *Hint:* Problem (3.f) might come in handy.

> *Solution.* This follows directly from the definition using Problem (3.f). In particular, let A, B be positive semidefinite. Pick any $\lambda \in [0, 1]$ and $x \in \mathbb{R}^n$; then,
>
> $$\begin{aligned} x^\top(\lambda A + (1 - \lambda)B)x &= \lambda(x^\top Ax) + (1 - \lambda)(x^\top Bx) \\ &\geq \lambda \cdot 0 + (1 - \lambda) \cdot 0 \\ &= 0, \end{aligned}$$
>
> where we used the fact that $x^\top Ax, x^\top Bx \geq 0$ since A and B are positive semidefinite (Problem (3.f)). Therefore, $x^\top(\lambda A + (1 - \lambda)B)x \geq 0$ for all x, which is equivalent to the statement that the convex combination of A and B is itself semidefinite. ◂

(4.c) [5pts] Show that the set $\left\{(x, y) \in \mathbb{R}^2_{\geq 0} : xy \geq 1\right\}$ is convex.

▸ *Hint:* At some point, the AM-GM inequality $a + b \geq 2\sqrt{ab}$, valid for all $a, b \geq 0$, might come in handy. (If you don't know why the AM-GM inequality holds, just observe that it is nothing but a restatement of $(\sqrt{a} - \sqrt{b})^2 \geq 0$...)

> *Solution.* Pick any two points $(x_1, y_1), (x_2, y_2) \in \Omega$, and $\lambda \in [0, 1]$. Denote

$$\begin{aligned}\overline{x} &\coloneqq \lambda x_1 + (1-\lambda)x_2,\\ \overline{y} &\coloneqq \lambda y_1 + (1-\lambda)y_2.\end{aligned}$$

We have

$$\overline{xy} = \lambda^2 x_1 y_1 + (1-\lambda)^2 y_1 y_2 + \lambda(1-\lambda)(x_1 y_2 + x_2 y_1).$$

Using the AM-GM inequality, we can write

$$x_1 y_2 + x_2 y_1 \geq 2\sqrt{x_1 y_1 x_2 y_2} \geq 2,$$

where the last inequality follows from the fact that $x_1 y_1, x_2 y_2 \geq 1$ since $(x_1, y_1), (x_2, y_2) \in \Omega$ by hypothesis. Plugging into the expression for $\overline{xy}$ and using the bounds $x_1 y_1, x_2 y_2 \geq 1$, we obtain

$$\overline{xy} \geq \lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda) \geq (\lambda + (1-\lambda))^2 = 1.$$

This concludes the proof. ◂

5. Existence of points with small gradients for lower-bounded functions [20pts]

It is well understood that whenever a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ has a (local) minimum x^*, then necessarily $\nabla f(x^*) = 0$. However, we know that not all functions, not even those that are lower bounded, have a minimum (consider for example the function e^{-x}).

However, we can show that in the case of lower-bounded differentiable functions, even if the gradient might never get to 0, it must get arbitrarily small in norm.

(5.a) [20pts] Let the function $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and bounded below. Prove that for any $\epsilon > 0$, there exists some point $x_\epsilon \in \mathbb{R}^n$ such that $\|\nabla f(x_\epsilon)\|_2 \le \epsilon$.

▸ *Hint:* Consider the function $f(x) + \epsilon\|x\|_2$, and argue that it must have a global minimizer on $\mathbb{R}^n$...

Solution. For any $\epsilon > 0$, the function $g(x) := f(x) + \epsilon\|\cdot\|_2$ has bounded level sets as for any α, and $M = \inf_x f(x)$, the level set $\{x : f(x) + \epsilon\|x\|_2 \le \alpha\}$ is a subset of the (bounded) ball $\left\{x : \|x\| \le \frac{\alpha - M}{\epsilon}\right\}$. Furthermore, g is the sum of continuous functions so it must be continuous. So by the Weierstrass theorem, it has a global minimizer x_ϵ.

The function $g(x)$ is *not* differentiable in $x = 0$, so we need to be careful about handling that case specially.

Let's start with the easy case: $x_\epsilon \ne 0$. Since the domain of g is $\mathbb{R}^n$, which is open, the first-order necessary optimality conditions at x_ϵ require that

$$0 = \nabla g(x_\epsilon) \qquad \Longrightarrow \qquad \nabla f(x_\epsilon) + \epsilon \frac{x_\epsilon}{\|x_\epsilon\|_2} = 0.$$

Rearranging and taking norms, we have $\|\nabla f(x_\epsilon)\|_2 = \epsilon$.

The only case left to check is $x_\epsilon = 0$. Here, the function g is not differentiable. To circumvent the issue, we can use directional derivatives. Indeed, it is necessary for the point x_ϵ to be optimal, that

$$0 \le g'(x_\epsilon; d) := \lim_{t \downarrow 0} \frac{g(x_\epsilon + td) - g(x_\epsilon)}{t}$$

for all directions d. Expanding the right-hand side using the fact that $x_\epsilon = 0$, we require

$$\begin{aligned} 0 &\le \lim_{t \downarrow 0} \frac{f(x_\epsilon + td) - f(x_\epsilon)}{t} + \epsilon \cdot \frac{\|0 + td\|_2 - \|0\|_2}{t} \\ &= \langle \nabla f(x_\epsilon), d\rangle + \epsilon \cdot \|d\|_2. \end{aligned}$$

> In particular, setting $d = -\nabla f(x_\epsilon)$, we obtain
>
> $$0 \le -\|\nabla f(x_\epsilon)\|_2^2 + \epsilon \cdot \|\nabla f(x_\epsilon)\|_2,$$
>
> which implies $\|\nabla f(x_\epsilon)\|_2 \le \epsilon$.
>
> So, in both cases, we conclude that $\|\nabla f(x_\epsilon)\|_2 \le \epsilon$, as we wanted to show. ◂

6.7220 Nonlinear Optimization
Spring 2025