MIT 6.7220/15.084 — Nonlinear Optimization Mar 21st 2024

In-Class Midterm

Instructor: Prof. Gabriele Farina (✉ `gfarina@mit.edu`)

Your information

Name: ....

MIT ID: ....

Instructions

You have 80 minutes to solve the following five problems.

This is a closed-book midterm.

Keep stapled.

Do not distribute outside of class.

You can use this page as scratch paper if you want. Scratch space will not be graded.

More scratch space is available at the end of the document. Additional scratch paper is available on request.

1 Existence and uniqueness of Euclidean projections

Let $\Omega \subseteq \mathbb{R}^n$ be a nonempty set.

(a) [7pts] Give a formal argument as to why, when Ω is *closed and convex*, the Euclidean projection of any point $x \in \mathbb{R}^n$ exists and is unique. As a reminder, the Euclidean projection of x is a solution to the nonlinear optimization problem

$$\begin{aligned} \min_y \ & \|y - x\|_2 \\ \text{s.t.} \ & y \in \Omega. \end{aligned}$$

(b) [7pts] If the assumption of *closedness* is removed, is the existence of a projection still guaranteed? For those cases in which a projection exists, is such a projection still guaranteed to be unique? For each, give either a proof or a counterexample.

(c) [7pts] If the assumption of *convexity* is removed, is the existence of a projection still guaranteed? For those cases in which a projection exists, is such a projection still guaranteed to be unique? For each, give either a proof or a counterexample.

(Remember: a set is closed *if "it contains its boundary", that is, the limit of any converging sequence of points that belong to the set belongs to the set. For example, the interval* $[1, 2] := \{x \in \mathbb{R} : 1 \le x \le 2\}$ *is a closed set, while the interval* $(1, 2] := \{x \in \mathbb{R} : 1 < x \le 2\}$ *is not. You can take for granted that the union of two closed sets is closed.)*

> *Solution.*
>
> (a) **Existence**: we choose a point $y_0 \in \Omega$ and define $\Omega_0 = \Omega \cap \{y : \|x - y\|_2 \le \|x - y_0\|_2\}$. It's clear that for any point y outside Ω_0, y cannot be the minimizer of the projection problem. Therefore, we can restrict the minimization of $\|y - x\|_2$ to Ω_0, and the Weierstrass Theorem guarantees the existence of a minimizer.
> **Uniqueness**: let z be one of the minimizers of the problem. The first-order optimality condition at z reads
>
> $$\langle x - z, y - z \rangle \le 0 \qquad \forall y \in \Omega.$$
>
> For any $y \neq z$,
>
> $$\|x - y\|_2^2 = \|x - z + (z - y)\|_2^2 = \|x - z\|_2^2 + \|z - y\|_2^2 - 2\langle x - z, y - z \rangle > \|x - z\|_2^2.$$
>
> Therefore, the projection is unique.
>
> (b) **No, existence is not guaranteed anymore if the assumption of closedness is removed**. For example, consider projecting the point 0 onto the open interval $(1, 2) \subset \mathbb{R}$. The projection problem has no solution, since for any candidate solution point $y \in (1, 2)$, the point
>
> $$y' := 1 + \frac{y - 1}{2}$$
>
> belongs to $(1, 2)$ and achieves a strictly lower objective. **However, in those cases where a projection does exist, then the projection would be unique**, again by the proof of uniqueness in (a) (the uniqueness argument does not depend on closedness).
>
> (c) **Yes. The existence argument, which is based on the Weierstrass theorem, does not require convexity, only closedness**. So, the existence of a projection is guaranteed even for nonconvex feasible sets Ω. **However, uniqueness is not guaranteed**. For example, consider projecting the origin $x = (0, 0)$ onto the circle

$$\Omega := \{y \in \mathbb{R}^2 : \|y\|_2 = 1\}.$$

All points on Ω have the same (minimum) distance of 1 from the origin, and so infinite projections exist.

□

2 Lipschitz continuity of projections

Let $\Omega \subseteq \mathbb{R}^n$ be a closed convex set, and let

$$\begin{aligned} \Pi_\Omega(x) \coloneqq \arg\min_y \ & \frac{1}{2}\|y - x\|_2^2 \\ \text{s.t.} \ & y \in \Omega \end{aligned}$$

denote the projection of point $x \in \mathbb{R}^n$ onto Ω. (Recall that Euclidean projections onto a closed and convex set exist and are unique.)

[20pts] Show that Π_Ω is 1-Lipschitz continuous, that is,

$$\left\|\Pi_\Omega(x) - \Pi_\Omega(x')\right\|_2 \le \left\|x - x'\right\|_2 \qquad \forall x, x' \in \mathbb{R}^n.$$

(Hint: start by writing the first-order optimality conditions for $z \coloneqq \Pi_\Omega(x)$ and $z' \coloneqq \Pi_\Omega(x')$. In particular, two vectors must have a nonpositive inner product (obtuse angle) with $z' - z$. What happens if you sum the inequalities and use Cauchy-Schwarz?)

> *Solution.* Let $x, x' \in \mathbb{R}^n$ be arbitrary, and $z \coloneqq \Pi_\Omega(x)$ and $z' \coloneqq \Pi_\Omega(x')$ denote their projections onto Ω. If $z = z'$, that is, $\|z - z'\| = 0$, the result follows immediately from the nonnegativity of $\left\|x - x'\right\|_2$. So, we will focus on the case $\|z - z'\| > 0$.
>
> The first-order optimality conditions for z, z' are both necessary and sufficient since the Euclidean projection problem is convex. These conditions are
>
> $$\begin{aligned} \langle x - z, y - z\rangle &\le 0 \qquad \forall y \in \Omega, \quad \text{and} \\ \langle x' - z', y - z'\rangle &\le 0 \qquad \forall y \in \Omega, \end{aligned}$$
>
> respectively. Plugging $y = z'$ in the first inequality and $y = z$ in the second inequality, we find in particular
>
> $$\langle x - z, z' - z\rangle \le 0, \quad \text{and} \quad \langle z' - x', z' - z\rangle \le 0.$$
>
> Summing the inequalities yields
>
> $$\langle (z' - z) + (x - x'), z' - z\rangle \le 0,$$
>
> which implies
>
> $$\|z' - z\|_2^2 \le \langle x' - x, z' - z\rangle \le \left\|x' - x\right\|_2 \left\|z' - z\right\|_2,$$
>
> where the second inequality follows from applying the Cauchy-Schwarz inequality. Dividing by $\left\|z' - z\right\|_2$ yields the statement. □

3 An optimization problem

Consider the optimization problem

$$\begin{aligned} \min_{x,y} \quad & \tfrac{1}{2}x^2 + \tfrac{1}{2}(y+3)^2 \\ \text{s.t.} \quad & x^2 + y^2 \geq 2 \\ & x + y \geq 0 \end{aligned}$$

(a) [3pts] Is the objective function, seen as a function from $\mathbb{R}^2$ to $\mathbb{R}$, convex? Provide a proof or a counterexample.

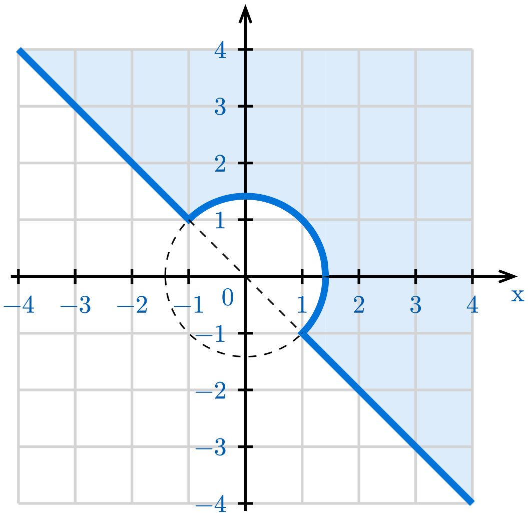
(b) [3pts] Is the feasible set a convex subset of $\mathbb{R}^2$? Provide a proof or a counterexample.

(c) [8pts] Write down the KKT conditions at a generic feasible point (x, y).

(d) [6pts] Check whether the KKT conditions hold at points $(x, y) = (1, -1)$ and $(x, y) = (-1, 1)$.

Solution. (a) The function is twice differentiable on an open set, with Hessian matrix equal to the identity, which is positive semidefinite. Hence, the objective function is convex.

(b) No, the feasible set is not convex. This becomes apparent from sketching the feasible set.



In particular, the points $(-1, 1)$ and $(1, -1)$ are both feasible. Yet, their midpoint $(0, 0)$ is not feasible.

(c) The KKT conditions at a generic feasible point (x, y) are the existence of (α, β) (the "Lagrange multipliers") such that

$$\begin{aligned} \begin{pmatrix} -x \\ -y-3 \end{pmatrix} &= \alpha \begin{pmatrix} -2x \\ -2y \end{pmatrix} + \beta \begin{pmatrix} -1 \\ -1 \end{pmatrix}, && \text{(Stationarity)}, \\ \alpha, \beta &\geq 0, && \text{(Dual feasibility)}, \\ \alpha(-2 + x^2 + y^2) &= 0, && \text{(Complementary slackness)} \\ \beta(x + y) &= 0. && \end{aligned}$$

(d) Plugging $(x, y) = (1, -1)$ into the KKT conditions yields the system

$$\begin{pmatrix}-1\\-2\end{pmatrix} = \alpha\begin{pmatrix}-2\\2\end{pmatrix} + \beta\begin{pmatrix}-1\\-1\end{pmatrix},$$

$$\alpha, \beta \geq 0,$$

$$\alpha \cdot 0 = 0,$$

$$\beta \cdot 0 = 0.$$

The stationarity condition has the unique solution $\alpha = -\frac{1}{4}, \beta = \frac{3}{2}$, which does not satisfy dual feasibility. **Hence, the KKT conditions do NOT hold at $(1, -1)$.**

Plugging $(x, y) = (-1, 1)$ into the KKT conditions yields the system

$$\begin{pmatrix}1\\-4\end{pmatrix} = \alpha\begin{pmatrix}2\\-2\end{pmatrix} + \beta\begin{pmatrix}-1\\-1\end{pmatrix},$$

$$\alpha, \beta \geq 0,$$

$$\alpha \cdot 0 = 0,$$

$$\beta \cdot 0 = 0.$$

The stationarity condition has the unique solution $\alpha = \frac{5}{4}, \beta = \frac{3}{2}$, which satisfies all other conditions. **Hence, the KKT conditions hold at $(-1, 1)$.** $\square$

4 Optimality conditions for semidefinite programs

Consider the optimization problem

$$\begin{aligned} \min_{X} \quad & f(X) \\ \text{s.t.} \quad & X \succeq 0 \\ & X \in \mathbb{S}^n, \end{aligned}$$

where f is a convex and differentiable function and $\mathbb{S}^n$ denotes the set of symmetric $n \times n$ matrices with real coefficients.

(a) [10pts] Consider the point $X_1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

- If $\nabla f(X_1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, can you conclude that X_1 is a minimizer of f? Can you conclude that X_1 is *not* the minimizer of f?
- What about $\nabla f(X_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$?

(b) [10pts] Consider the point $X_2 := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

- If $\nabla f(X_2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, can you conclude that X_2 is a minimizer of f? Can you conclude that X_2 is *not* the minimizer of f?
- What about $\nabla f(X_2) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$?

For each answer, provide a (short) formal argument.

(Hint: consider the first-order optimality conditions for conic problems. You do not need to prove the expression for the dual/polar of the positive semidefinite cone.)

> *Solution.* Let $\mathcal{S}^n$ denote positive semidefinite cone. As proven in Lecture 7, the normal cone at any $X \succeq 0$ is given by
>
> $$\mathcal{N}_{\mathcal{S}^n}(X) = \left\{ Y : \langle X, Y \rangle = 0, Y \in (\mathcal{S}^n)^{\perp} \right\},$$
>
> where $(\mathcal{S}^n)^{\perp}$ denotes the cone polar to $\mathcal{S}^n$. Since the positive semidefinite cone is self-dual, then $(\mathcal{S}^n)^{\perp} = -\mathcal{S}^n$ and the normal cone at X can be written as
>
> $$\mathcal{N}_{\mathcal{S}^n}(X) = \{ Y : \langle X, Y \rangle = 0, Y \preceq 0 \},$$
>
> Since the problem is convex, the first-order optimality conditions in this case are both necessary and sufficient and are
>
> $$-\nabla f(X) \in \mathcal{N}_{\mathcal{S}^n}(X).$$
>
> Using the characterization of $\mathcal{N}_{\mathcal{S}^n}(X)$ given above, we therefore have that at optimality it is necessary and sufficient that
>
> $$\langle \nabla f(X), X \rangle = 0, \quad \text{and} \quad \nabla f(X) \succeq 0.$$
>
> (a) Using the above criterion, we find that in both cases we can conclude that the point X_1 is *not* the minimizer of f. In the first bullet point, the condition $\langle \nabla f(X_1), X_1 \rangle = 0$ was not satisfied. In the second bullet point, the condition $\nabla f(X_1) \succeq 0$ was not satisfied.
>
> (b) Using the above criterion, we find that in the first bullet point, X_2 is not the minimizer of f, since the condition $\langle \nabla f(X_2), X_2 \rangle = 0$ is not satisfied. In the second bullet point, $\langle \nabla f(X_2), X_2 \rangle = 0$ and $\nabla f(X_2) \succeq 0$, so X_2 is a minimizer of f. □

5 Gradient descent

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable, L-smooth function ($L > 0$) attaining a minimum at a point $x_\star \in \mathbb{R}^n$. Consider the iterates $\{x_t\}$ produced by the gradient descent algorithm run with constant stepsize $\eta = 1/L$ starting from the initial point $x_0 = 0 \in \mathbb{R}^n$.

(a) [10pts] Is it guaranteed that the function value is nonincreasing, that is,

$$f(x_{t+1}) \le f(x_t) \text{ at all times } t?$$

What if f is further assumed to be convex? For each question, either give a proof or a counterexample.

(b) [10pts] Is it guaranteed that the distance to $x_\star$ is nonincreasing, that is,

$$\left\|x_{t+1} - x_\star\right\|_2 \le \|x_t - x_\star\|_2 \text{ at all times } t?$$

What if f is further assumed to be convex? For each question, either give a proof or a counterexample.

(Remember: a differentiable function is L-smooth *when its gradient is* L*-Lipschitz continuous, that is,* $\|\nabla f(x) - \nabla f(y)\|_2 \le L \cdot \|x - y\|_2$ *for all* $x, y \in \mathbb{R}^n$*.)*

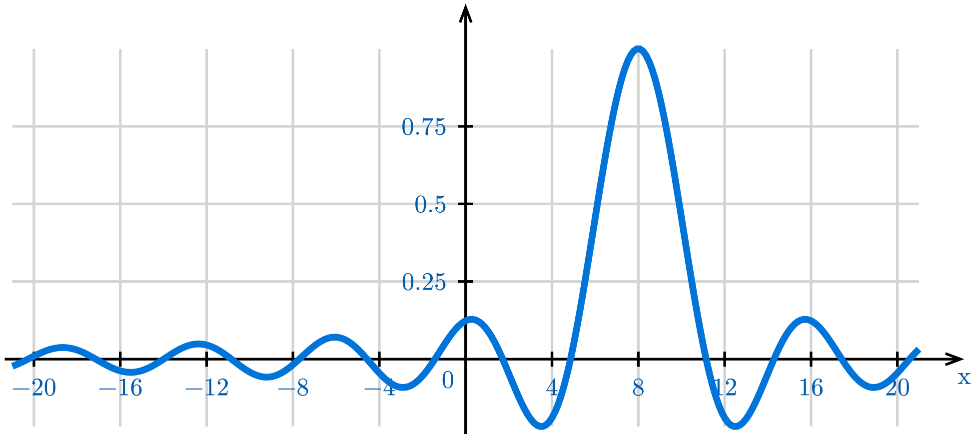
Solution. (a) **Yes**. The gradient descent lemma states that for any L-smooth function, when $\eta = \frac{1}{L}$ one has

$$f(x_{t+1}) \le f(x_t) - \frac{1}{2L}\|\nabla f(x_t)\|_2^2.$$

Hence, $f(x_{t+1}) \le f(x_t)$ at all steps, whether or not f is convex.

(b) **With only L-smoothness, the guarantee does not hold**. Many examples are possible, but just pick one, consider the function

$$f(x) = \frac{\sin(x-8)}{x-8}.$$



This function is L-smooth, but it is clear that already at time 0, x_1 will move further away from any minimizer of the function, by "rolling down" to the left of the origin.

When the function is convex, however, the guarantee holds. This is a consequence of the Euclidean mirror descent lemma, which states that

$$f(x_t) \leq f(x_\star) + \frac{L}{2}\left(\|x_\star - x_t\|_2^2 - \left\|x_\star - x_{t+1}\right\|_2^2 + \left\|x_{t+1} - x_t\right\|_2^2\right).$$

Using the fact that $x_{t+1} - x_t = \frac{1}{L}\nabla f(x_t)$, together with the gradient descent lemma, we have that

$$\begin{aligned} f(x_t) &\leq f(x_\star) + \frac{L}{2}\left(\|x_\star - x_t\|_2^2 - \left\|x_\star - x_{t+1}\right\|_2^2 + \frac{1}{L^2}\|\nabla f(x_t)\|_2^2\right) \\ &\leq f(x_\star) + \frac{L}{2}\left(\|x_\star - x_t\|_2^2 - \left\|x_\star - x_{t+1}\right\|_2^2\right) + \frac{1}{2L}\|\nabla f(x_t)\|_2^2 \\ &\leq f(x_\star) + \frac{L}{2}\left(\|x_\star - x_t\|_2^2 - \left\|x_\star - x_{t+1}\right\|_2^2\right) + f(x_t) - f(x_{t+1}). \end{aligned}$$

Rearranging the terms yields

$$\left\|x_\star - x_{t+1}\right\|_2^2 \leq \|x_\star - x_t\|_2^2 - \frac{2}{L}(f(x_{t+1}) - f(x_\star)).$$

Since by definition $x_\star$ minimizes f, then the last term is nonpositive and the guarantee holds. ☐

Scratch page (not graded).

Scratch page (not graded).

Scratch page (not graded).

Scratch page (not graded).

6.7220 Nonlinear Optimization
Spring 2025