MIT 6.7220/15.084 — Nonlinear Optimization Spring 2025

Problem Set 2

Due: Mar 11th 2025 11:59pm ET

Instructions. This problem set is slightly larger than the previous one and contains 6 problems. For this reason, it is due in two weeks (14 days). The total number of points is 105.

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. Projection onto the ℓ_1-ball [15pts]

Let

$$\Omega := \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$$

be the unit ball with respect to the ℓ_1 norm, where as a reminder

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|.$$

(1.a) [5pts] Show that Ω can be expressed as the intersection of a finite number of halfspaces (dependent on n), and give an expression for the normal cone at any point $x \in \Omega$.

> *Solution.* By the definition of $\|\cdot\|_1$, we can write
>
> $$\|x\|_1 = \max_{e \in \{-1,1\}^n} e^\top x,$$
>
> where the max occurs for $e \odot \operatorname{sgn}(x) \geq 0$ with $\odot$ being the elementwise product. Therefore, the set Ω can be written as
>
> $$\Omega = \bigcap_{e \in \{-1,1\}^n} \{x \in \mathbb{R}^n \mid e^\top x \leq 1\},$$
>
> which is the intersection of 2^n half-spaces.
>
> If $\|x\|_1 < 1$, x is an interior point and therefore
>
> $$\mathcal{N}_\Omega(x) = \{0\}$$
>
> Now assume $\|x\|_1 = 1$ and define set $I = \{i \in [n] : x_i \neq 0\}$, then let
>
> $$E_x = \{\forall e \in \{-1,1\}^n \mid \forall i \in I : e_i = \operatorname{sgn}\ (x_i)\}$$
>
> E_x is the set of all inequalities that are tight at the point x, hence the normal cone can be written as
>
> $$\begin{aligned}\mathcal{N}_\Omega(x) &= \left\{ \sum_{e \in E_X} \lambda_e \cdot e \mid \forall e : \lambda_e \geq 0 \right\} \\ &= \left\{ d \mid \exists \lambda_e \geq 0 : \forall i \in I : d_i = \operatorname{sgn}(x_i) \cdot \sum_{e \in E_X} \lambda_e, \text{and } \forall i \notin I : d_i = \sum_{e \in E_X} \lambda_e \cdot \operatorname{sgn}(e_i) \right\} \\ &= \left\{ d \mid \exists \lambda_e \geq 0 : \forall i \in I : d_i = \operatorname{sgn}(x_i) \cdot \sum_{e \in E_X} \lambda_e, \text{and } \forall i \notin I : \mid d_i \mid \leq \sum_{e \in E_X} \lambda_e \right\} \\ &= \{ d \mid \exists \lambda \geq 0 : \forall i \in I : d_i = \lambda \cdot \operatorname{sgn}(x_i), \text{and } \forall i \notin I : \mid d_i \mid \leq \lambda \}\end{aligned}$$

Where in the last equality $\lambda = \sum_{e \in E_X} \lambda_e$. ◂

(1.b) [5pts] Let $y \in \mathbb{R}^n$ be a given point. Show that the Euclidean projection x of y onto Ω must satisfy the equality

$$x_i = \operatorname{sgn}(y_i)[|y_i| - \lambda]^+ \qquad \forall i \in \{1, ..., n\}$$

for some $\lambda \in \mathbb{R}_{\geq 0}$, where $[\cdot]^+$ is a shorthand for $\max\{\cdot, 0\}$ and sgn is the sign function defined as

$$\operatorname{sgn}(z) := \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z = 0 \\ -1 & \text{if } z < 0. \end{cases}$$

Solution. Euclidean projection of the given point y is the solution to the following optimization problem

$$\min_{\|x\|_1 \leq 1} \ \tfrac{1}{2}\|x - y\|_2^2.$$

If the projection x is an interior point, then

$$y - x \in \mathcal{N}_\Omega(x) = \{0\} \Rightarrow x = y \Rightarrow \|y\|_1 \leq 1$$

Therefore, if the point is in the domain, the solution to the projection problem is the point itself. Now assume $\|y\|_1 > 1$, then the optimal point can not be in the interior of the feasible set, and the first-order optimality condition implies that

$$y - x \in \mathcal{N}_\Omega(x) = \{d \mid \exists \lambda \geq 0 : \forall i \in I : d_i = \lambda \cdot \operatorname{sgn}(x_i), \text{and } \forall \notin I : \mid d_i \mid \leq \lambda\}$$

Therefore, at the optimal point, we must have

$$\begin{cases} |y_i - x_i| \leq \lambda & \text{if } x_i = 0 \\ y_i - x_i = \lambda \cdot \operatorname{sgn}(x_i) & \text{if } x_i \neq 0 \end{cases}$$

These are equivalent to

$$\begin{cases} |y_i| \leq \lambda & \text{if } x_i = 0 \\ y_i = x_i + \lambda \cdot \operatorname{sgn}(x_i) & \text{if } x_i \neq 0 \end{cases}$$

Taking $\operatorname{sgn}(\cdot)$ from both sides of the second case results in $\operatorname{sgn}(y_i) = \operatorname{sgn}(x_i)$ and hence the conditions can be rewritten as

$$\begin{cases} |y_i| \leq \lambda & \text{if } x_i = 0 \\ x_i = y_i - \lambda \cdot \operatorname{sgn}(y_i) = \operatorname{sgn}(y_i)[|y_i| - \lambda] & \text{if } x_i \neq 0 \end{cases}$$

Which can be written as

$$x_i = \operatorname{sgn}(y_i) \cdot [|y_i| - \lambda]^+ \qquad \forall i \in [n]$$

Note that whenever $|y_i| \leq \lambda$, this function outputs zero and would output $\operatorname{sgn}(y_i) \cdot (|y_i| - \lambda)$ otherwise, which is equivalent to what we derived from the first-order optimality condition. ◂

(1.c) [5pts] Based on the result above, give an algorithm with complexity $O(n \log n)$ for computing the projection of any given $y \in \mathbb{R}^n$ onto Ω.

Solution. If the Euclidean projection of y onto the ℓ_1-ball is an interior point, then by using the result of Problem (1.a), we must have $y = x$ and subsequently $\|y\|_1 \leq 1$, in which case the point is already in Ω and the projection is the same point.

Now we consider the case that the projection, x, is not an interior point, and hence $\|x\|_1 = 1$, Problem (1.b) states that the solution must satisfy the form

$$x_i = \operatorname{sgn}(y_i) \cdot [|y_i| - \lambda]^+ \qquad \forall i \in [n]$$

With this formulation, by having λ, the projection x can be computed in $O(n)$. We should find λ such that $\|x\|_1 = 1$

$$\begin{aligned} 1 = \|x\|_1 &= \sum_{i \in [n]} | \operatorname{sgn}(y_i) \cdot [|y_i| - \lambda]^+ \, | \\ &= \sum_{i \in [n]} [|y_i| - \lambda]^+ \\ &= f(\lambda) \end{aligned}$$

To find λ, we first sort $|y|$, which takes $O(n \log n)$ and denote the i-th largest absolute value in y is denoted by $|y|_{(i)}$. Let

$$S_j = \sum_{i=1}^{j} |y|_{(i)}$$

be the cumulative sum of the largest j elements. All S_j's can be computed by a pass over sorted $|y|$ and hence requires $O(n)$. To find λ we condition on the index k such that

$$|y|_{(k)} > \lambda \geq |y|_{(k+1)}$$

Knowing i, the equality constraint can be written as

$$1 = f(\lambda_k) = \sum_{i \in [n]} (|y_i| - \lambda_k)^+ = \sum_{i \in [k]} \left(|y|_{(i)} - \lambda_k \right)$$

Hence

$$\lambda_k = \frac{\sum_{i\in[k]} |y|_{(i)} - 1}{k} = \frac{S_k - 1}{k}$$

And we only need to check if the calculated value λ_k satisfies the corresponding inequality

$$|y|_{(k)} > \lambda_k \geq |y|_{(k+1)}$$

Note that calculating and checking λ_k's also takes $O(n)$, and therefore the total complexity is $O(n \log n)$. ◂

2. Nonexpansive maps and separation of fixed points [20pts]

A continuous function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be *nonexpansive* if it satisfies

$$\|F(x) - F(y)\|_2 \le \|x - y\|_2 \qquad \forall x, y \in \mathbb{R}^n.$$

An important example of nonexpansive functions are Euclidean projections, as shown in the next subproblem.

(2.a) [6pts] Let $\Omega \subseteq \mathbb{R}^n$ be closed and convex, and let $\Pi_\Omega(z)$ denote the Euclidean projection of $z \in \mathbb{R}^n$ onto Ω. Show that Π_Ω is nonexpansive.

- *Hint:* Take any two $x, x' \in \mathbb{R}^n$, and consider their respective projections $z := \Pi_\Omega(x)$ and $z' := \Pi_\Omega(x')$ onto Ω. Write the first-order optimality conditions at z and z'. 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?

- *Remark:* This question, with the same hint, was in the 2024 Midterm exam.

> *Solution.* Let $x, x' \in \mathbb{R}^n$ be arbitrary, and $z := \Pi_\Omega(x)$ and $z' := \Pi_\Omega(x')$ denote their projections onto Ω. If $z = z'$, that is, $\|z - z'\| = 0$, the result follows immediately from the nonnegativity of $\|x - x'\|_2$. 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
>
> $$\langle x - z, y - z\rangle \le 0 \qquad \forall y \in \Omega, \quad \text{and}$$
> $$\langle x' - z', y - z'\rangle \le 0 \qquad \forall y \in \Omega,$$
>
> 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
>
> $$\|z' - z\|_2^2 \le -\langle x' - x, z' - z\rangle \le \|x' - x\|_2 \|z' - z\|_2,$$
>
> where the second inequality follows from applying the Cauchy-Schwarz inequality. Dividing by $\|z' - z\|_2$ yields the statement. ◀

Now, consider a generic nonexpansive function $F : \mathbb{R}^n \to \mathbb{R}^n$, and assume the set of its fixed points $\Omega_\star := \{x \in \mathbb{R}^n : F(x) = x\}$ is nonempty.

(2.b) [7pts] Show that $\Omega_\star$ is a convex set.

▸ *Hint:* Let $x, x' \in \Omega_\star$, and x_t be some convex combination of the two. Then, F will try to weakly "pull" x_t towards x and at the same time towards x' (why?). What is the only possibility?

> *Solution.* Consider two points $x, x' \in \Omega_*$, and any $t \in [0,1]$. By definition of Ω_*,
>
> $$F(x) = x, \qquad F(x') = x'.$$
>
> Let $x_t := (1-t)x + tx'$. We want to show that x_t is also a fixed point of F.
>
> The key observation is that
>
> $$\begin{aligned} \|F(x_t) - x\|_2 &= \|F(x_t) - F(x)\|_2 \le \|x_t - x\|_2 \\ \|F(x_t) - x'\|_2 &= \|F(x_t) - F(x')\|_2 \le \|x_t - x'\|_2 \end{aligned}$$
>
> and therefore
>
> $$\|F(x_t) - x\|_2 + \|F(x_t) - x'\|_2 \le \|x - x'\|_2.$$
>
> On the other hand, the triangle inequality implies the reverse inequality, and therefore all inequalities must be equalities:
>
> $$\|F(x_t) - x\| = \|x_t - x\|_2, \qquad \text{and} \qquad \|F(x_t) - x'\| = \|x_t - x'\|_2.$$
>
> In other words, $F(x_t)$ must be in a ball of radius $\|x_t - x\|_2$ from x, and in a ball of radius $\|x_t - x'\|_2$ from x'. The sum of the radii is exactly equal to the distance between x and x', and therefore the only intersection between the balls is x_t.
>
> \- -
>
> In the last part of the argument we used this intuitive fact: the intersection between two balls with centers $a, b \in \mathbb{R}^n$, $a \neq b$, and radii $r_1, r_2 > 0$ with $r_1 + r_2 = \|a - b\|_2$, is unique. For completeness, we include here a proof. The proof was not required for the homework, but it is a good exercise in translating geometric intuition into a more formal algebraic proof.
>
> Suppose that two distinct intersections, say s and s', exist in the case $r_1 + r_2 = \|a - b\|_2$. By the strict convexity of the function $\|x - a\|_2^2$, it would follow that the midpoint $\frac{s+s'}{2}$ is at distance $< r_1$ from a. Similarly, the midpoint $\frac{s+s'}{2}$ is at distance $< r_2$ from b. Hence, $q := \frac{s+s'}{2}$ is such that $\|q - a\|_2 + \|q - b\|_2 < r_1 + r_2 = \|a - b\|_2$. This violates the triangle inequality, absurd. ◂

(2.c) [7pts] Construct a separation oracle for $\Omega_\star$. Your proposed algorithm should return the correct answer in time polynomial in the dimension n, ignoring the cost of evaluating F at any point in $\mathbb{R}^n$.

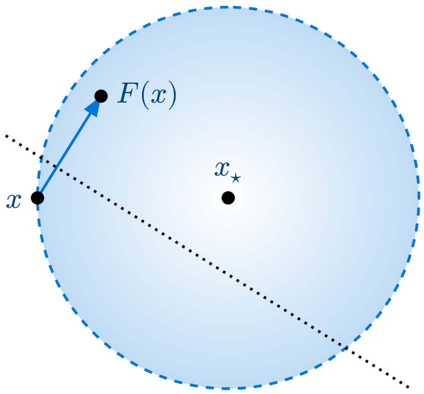
▸ *Hint:* Consider any $x_\star \in \Omega_\star$. F will keep x within a ball of radius $\|x_\star - x\|_2$ from $x_\star$. Draw a figure, and convince yourself that the direction $F(x) - x$ might be interesting...

▸ *Remark:* This implies that, provided a few minor details are in place, the ellipsoid method can efficiently compute fixed points of nonexpansive maps.

▸ *Remark:* The same question, but relative to nonexpansive maps with respect to the ℓ_∞ norm instead of the ℓ_2 norm is an important open question in theoretical computer science, related to the `SSG` complexity class.

Solution. Given a point $x \in \mathbb{R}^n$, we can check in linear time in the dimension n if $F(x) = x$. If yes, $x \in \Omega_\star$ and we are done. Otherwise, we need to find a separating hyperplane.

As hinted in the statement, the main insight is that the direction $F(x) - x$ is always separating.



Let $x_\star \in \Omega_\star$ be arbitrary, and let's verify that indeed $F(x) - x$ separates x from $\{x_\star\}$. Since $x_\star \in \Omega_\star$ was arbitrary, this will imply that $F(x) - x$ separates x from $\Omega_\star$. We have:

$$\begin{aligned}\langle F(x) - x, x - x_* \rangle &= \langle F(x) - x_\star, x - x_\star \rangle - \langle x - x_\star, x - x_\star \rangle \\ &= \langle F(x) - x_\star, x - x_\star \rangle - \|x - x_\star\|_2^2 \\ &\le \|F(x) - x_\star\|_2 \cdot \|x - x_\star\|_2 - \|x - x_\star\|_2^2.\end{aligned}$$

By the nonexpansiveness hypothesis and the definition of Ω_*, we have that $\|F(x) - x_\star\|_2 \le \|x - x_\star\|_2$. If $\|F(x) - x_\star\|_2 < \|x - x_\star\|_2$, we are done. Otherwise, consider the case of $\|F(x) - x_\star\| = \|x - x_\star\|_2$. Since $x \neq F(x)$ by hypothesis, then it is impossible that $F(x) - x_\star = x - x_\star$, and the Cauchy-Schwarz inequality must be strict. Hence, in either case

$$\langle F(x) - x, x - x_\star \rangle < 0,$$

> meaning that $F(x) - x$ is a separating direction. ◂

3. Convex functions [15pts]

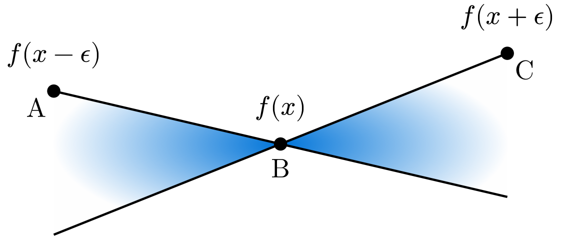
(3.a) [10pts] The other day, one of you asked if convexity in general implies continuity. As shown in class, the answer in general is no. As a reminder, the example that was drawn on the board at the end of class was the discontinuous convex function

$$f : [0,1] \to \mathbb{R}, \qquad f(x) := \begin{cases} 0 \text{ if } x \in [0,1) \\ 1 \text{ if } x = 1. \end{cases}$$

However, as it turns out, no discontinuity can happen in the *interior* of the domain.

More specifically, consider a convex function $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}$ is convex, and any x in the interior of Ω (that is, there exists $\epsilon > 0$ such that the ball of radius ϵ centered in x is fully contained in Ω). Prove that f must be continuous at x.

▸ *Hint:* Let the graphical intuition in the figure down here guide you: wherever the graph of the function f lies, it needs to be in the shaded cone below due to convexity. Indeed, convexity directly implies that the graph of f should be *below* A-B and B-C. But it also implies that it should be *above* the continuation of B-C and A-B, or otherwise the function value at B would be too high and violate the definition of convexity. The cone produces a sandwiching that prevents the value at x from jumping discontinuously.



▸ *Remark:* The result can be extended to arbitrary dimensions $n \in \mathbb{N}$, but doing so is not directly immediate since continuity on every line does not imply continuity in general. While this is *not* part of the homework, if you are curious, think about how you would extend the result you just showed to arbitrary finite dimensions.

> *Solution.* Let $t \in [0,1]$ and consider the points $x \pm t\epsilon$. We can express these points as convex combinations between x and $x \pm \epsilon$ according to
>
> $$x \pm t\epsilon = t \cdot (x \pm \epsilon) + (1-t) \cdot x.$$
>
> Therefore, we can use the convexity of f to conclude that $f(x \pm t\epsilon)$ must lie below the segments B-C and A-B in the figure:
>
> $$f(x + t\epsilon) \le t f(x + \epsilon) + (1-t) f(x)$$
> $$f(x - t\epsilon) \le t f(x - \epsilon) + (1-t) f(x).$$
>
> These inequalities imply

$$\begin{aligned}f(x+t\epsilon)-f(x)&\le t(f(x+\epsilon)-f(x))\\ f(x-t\epsilon)-f(x)&\le t(f(x-\epsilon)-f(x)).\end{aligned}$$

On the other hand, we can express the point x as convex combinations

$$x=\frac{1}{t+1}(x\pm t\epsilon)+\frac{t}{t+1}(x\mp\epsilon),$$

and using convexity of f again we have

$$\begin{aligned}f(x)&\le\frac{1}{t+1}f(x+t\epsilon)+\frac{t}{t+1}f(x-\epsilon)\\ f(x)&\le\frac{1}{t+1}f(x-t\epsilon)+\frac{t}{t+1}f(x+\epsilon).\end{aligned}$$

Multiplying by $t+1$ and rearranging, we find that $f(x\pm t\epsilon)$ must lie above the continuations of the segments A-B and B-C:

$$\begin{aligned}f(x+t\epsilon)-f(x)&\ge t(f(x)-f(x-\epsilon))\\ f(x-t\epsilon)-f(x)&\ge t(f(x)-f(x+\epsilon)).\end{aligned}$$

This shows that around x, the function f is sandwiched between linear functions that pass through x. Hence, f must be continuous at x. ◂

(3.b) [5pts] Prove that a function $f:\Omega\to\mathbb{R}$ on a convex domain Ω is convex if and only if, for all $k\in\mathbb{N}$, $x_1,\dots,x_k\in\Omega$, and $\lambda_1,\dots,\lambda_k\in\mathbb{R}_{\ge 0}$ with $\lambda_1+\cdots+\lambda_k=1$,

$$f(\lambda_1x_1+\cdots+\lambda_kx_k)\le\lambda_1f(x_1)+\cdots+\lambda_kf(x_k).$$

▸ *Hint:* The case $k=1$ is trivial. The case $k=2$ is the definition of convexity. For the general case, use induction.

Solution. We prove the result by induction. The cases up to $k=2$ are trivial, so we focus on the inductive step.

Consider any convex combination $\lambda_1x_1+\cdots+\lambda_{k+1}x_{k+1}$. If $\lambda_1=1$, the statement is trivial since then $\lambda_2=\cdots=\lambda_{k+1}=0$ and the statement reduces to the case $k=1$. Hence, we focus on the case $\lambda_1\in[0,1)$. Then, we can write

$$\lambda_1x_1+\cdots+\lambda_{k+1}x_{k+1}=\lambda_1x_1+(1-\lambda_1)\left(\frac{\lambda_2}{1-\lambda_1}x_1+\cdots+\frac{\lambda_{k+1}}{1-\lambda_1}x_{k+1}\right).$$

Importantly, since $\lambda_2,\dots,\lambda_{k+1}\ge 0$ and $\lambda_2+\ldots+\lambda_{k+1}=1-\lambda_1$, the large parenthesis on the right-hand side is a convex combination of k points and belongs to Ω. Hence, using first the convexity of f and then the inductive hypothesis, we can write

$$
\begin{aligned}
&f(\lambda_1 x_1 + \cdots + \lambda_{k+1} x_{k+1}) \\
&\qquad \le \lambda_1 f(x_1) + (1-\lambda_1) f\left(\frac{\lambda_2}{1-\lambda_1} x_1 + \cdots + \frac{\lambda_{k+1}}{1-\lambda_1} x_{k+1}\right) \\
&\qquad \le \lambda_1 f(x_1) + (1-\lambda_1) \left(\frac{\lambda_2}{1-\lambda_1} f(x_1) + \cdots + \frac{\lambda_{k+1}}{1-\lambda_1} f(x_{k+1})\right) \\
&\qquad = \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_{k+1} f(x_{k+1}).
\end{aligned}
$$

This concludes the proof of the inductive step. ◂

4. Separation oracles [15pts]

(4.a) [10pts] Construct a separation oracle for the ellipsoid

$$\Omega := \{x \in \mathbb{R}^n : a_1 x_1^2 + \dots + a_n x_n^2 \le 1\},$$

where $a_1, \dots, a_n > 0$ are given positive coefficients. Your proposed algorithm should return a correct answer in time polynomial in n.

Solution. This problem has multiple solutions, including the rather tedious approach of trying to compute an explicit projection onto Ω.

One of the most direct ways of constructing a separation oracle in this case is to realize that an ellipsoid is a linear transformation of a unit ball. In particular, consider the matrix $A = \mathrm{Diag}(\sqrt{a_1}, \sqrt{a_2}, \dots, \sqrt{a_n})$. The set $A\Omega = \{Ax : x \in \Omega\}$ is the unit ball. Let y be the input of the separation oracle, and define $z := Ay$. If z is in the unit ball, we return that $y \in \Omega$. Otherwise, we know that

$$u = -z$$

is a separating direction for the unit ball, meaning that

$$u^\top (z - Ax) < 0 \qquad \forall x \in \Omega.$$

Expanding the definition of z and u, we therefore have

$$0 > u^\top (z - Ax) = (-Ay)^\top (Ay - Ax) = (-A^\top Ay)^\top (y - x)$$

for all $x \in \Omega$. This shows that the direction $-A^\top Ay = -(a_1 y_1, \dots, -a_n y_n)$ is a separating direction.

Though not required, we can also check directly that the separating direction above works. Since $\sum_{i=1}^n a_i y_i^2 > 1$,

$$u^\top y = -\sum_{i=1}^n a_i y_i^2 < -\frac{1 + \sum_{i=1}^n a_i y_i^2}{2} =: v.$$

On the other hand, it follows from Cauchy–Schwarz inequality that for any $x \in \Omega$,

$$\begin{aligned}
u^\top x &= -\sum_{i=1}^{n} a_i y_i x_i = -\sum_{i=1}^{n} \sqrt{a_i} y_i \sqrt{a_i} x_i \\
&\geq -\left(\sum_{i=1}^{n} a_i y_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} a_i x_i^2\right)^{\frac{1}{2}} \\
&\geq -\left(\sum_{i=1}^{n} a_i y_i^2\right)^{\frac{1}{2}} \\
&> -\frac{1 + \sum_{i=1}^{n} a_i y_i^2}{2}.
\end{aligned}$$

◀

(4.b) [5pts] Suppose that two sets $\Omega_1, \Omega_2 \subseteq \mathbb{R}^n$ each admits a separation oracle. Show that the separation oracles can be efficiently combined in a black-box fashion to obtain a separation oracle for the set $\Omega_1 \cap \Omega_2$.

Solution. Given a point, y, and a set Ω, a separation oracle outputs either $y \in \Omega$ or a vector u such that $\langle u, y\rangle < \langle u, x\rangle \quad \forall x \in \Omega$. Let O_1 and O_2 be the separation oracles for the sets Ω_1 and Ω_2. Notice that if $\langle u_1, y\rangle < \langle u_1, x\rangle \quad \forall x \in \Omega_1$ is satisfied, then so is $\langle u_1, y\rangle < \langle u_1, x\rangle \quad \forall x \in \Omega_1 \cap \Omega_2$ since $\Omega_1 \cap \Omega_2 \subset \Omega_1$. Hence u_1 separates y from $\Omega_1 \cap \Omega_2$ if it separates y from Ω_1. The Ω_2 counterpart of this statement also holds. So, a separation oracle can efficiently be built for the set $\Omega_1 \cap \Omega_2$ as follows:

$$\begin{cases}
y \in \Omega_1 \cap \Omega_2 & \text{if } O_1 = y \in \Omega_1 \wedge O_2 = y \in \Omega_2 \\
(y \notin \Omega_1 \cap \Omega_2, u_1) & \text{if } O_1 = (y \notin \Omega_1, u_1) \\
(y \notin \Omega_1 \cap \Omega_2, u_2) & \text{otherwise.}
\end{cases}$$

◀

5. Equivalent conditions to strict and strong convexity [25pts]

Recall that a function $f : \Omega \to \mathbb{R}$, with $\Omega \subseteq \mathbb{R}^n$ convex, is called *strictly* convex if

$$f((1-t)x + ty) < (1-t)f(x) + tf(y) \qquad \forall x, y \in \Omega, t \in (0,1), x \neq y.$$

The function f is called *strongly convex with modulus* $\mu > 0$, or μ-*strongly convex* for short, if

$$f(x) - \frac{\mu}{2}\|x\|_2^2 \quad \text{is convex.}$$

(5.a) [2pts] Prove that if f is twice differentiable and Ω is open, μ-strong convexity is equivalent to the condition

$$\nabla^2 f(x) \succeq \mu I \qquad \forall x \in \Omega.$$

▸ *Hint:* Use Condition (4) in Theorem L4.3.

> *Solution.* We know from Theorem L4.3 that convexity for twice differentiable functions defined on open domains is equivalent to a positive semidefinite Hessian matrix at all points in the domain. Hence,
>
> $$\begin{aligned} f(x) - \frac{\mu}{2}\|x\|_2^2 \text{ convex} \qquad &\iff \qquad \nabla^2 f(x) - \mu I \succeq 0 \\ &\iff \qquad \nabla^2 f(x) \succeq \mu I, \end{aligned}$$
>
> where we used the fact that the Hessian matrix of $\|x\|_2^2$ is the identity. ◂

(5.b) [5pts] Is the condition $\nabla^2 f(x) \succ 0$ at all $x \in \Omega$ sufficient for strict convexity, for a twice differentiable function f, independently of whether Ω is open or not?

> *Solution.* By hypothesis, for any $x, y \in \Omega, x \neq y$, and $\tau \in [0,1]$,
>
> $$0 < \left\langle y - x, \nabla^2 f(x + \tau \cdot (y - x)) \cdot (y - x) \right\rangle.$$
>
> Hence, taking the integral,
>
> $$\begin{aligned} 0 &< \int_0^1 \left\langle y - x, \nabla^2 f(x + t \cdot (y - x)) \cdot (y - x) \right\rangle \mathrm{d}t \\ &= \left\langle y - x, \int_0^1 \underbrace{\nabla^2 f(x + t \cdot (y - x)) \cdot (y - x)}_{= \frac{\mathrm{d}}{\mathrm{d}t} \nabla f(x + t \cdot (y - x))} \mathrm{d}t \right\rangle = \left\langle y - x, \nabla f(y) - \nabla f(x) \right\rangle. \end{aligned}$$
>
> This establishes *strict monotonicity*. In particular, defining the point $x_t := x + t \cdot (y - x)$ for $t \in (0,1)$ (so that $x_t \neq x$ since $y \neq x$), we have
>
> $$0 < \langle \nabla f(x_t) - \nabla f(x), x_t - x \rangle = t \cdot \langle \nabla f(x_t) - \nabla f(x), y - x \rangle,$$

which implies that $\langle \nabla f(x_t) - \nabla f(x), y - x\rangle > 0$ for all $t \in (0,1)$.

Letting t range from 0 to 1 and integrating,

$$\begin{aligned}0 &< \int_0^1 \langle y - x, \nabla f(x_t) - \nabla f(x)\rangle \,\mathrm{d}t \\ &= -\langle y - x, \nabla f(x)\rangle + \int_0^1 \langle y - x, \nabla f(x + t \cdot (y - x))\rangle \,\mathrm{d}t \\ &= -\langle y - x, \nabla f(x)\rangle + f(y) - f(x).\end{aligned}$$

Rearranging yields $f(y) > f(x) + \langle \nabla f(x), y - x\rangle$. Finally, pick any $x, y \in \Omega$ and $t \in (0,1)$, and consider the point

$$\Omega \ni z \coloneqq t \cdot x + (1-t) \cdot y.$$

Using the previously obtained expression we can write that

$$\begin{aligned}f(x) &> f(z) + \langle \nabla f(z), x - z\rangle, \\ f(y) &> f(z) + \langle \nabla f(z), y - z\rangle.\end{aligned}$$

Multiplying the first inequality by t and the second by $1 - t$, and summing, we obtain

$$t \cdot f(x) + (1-t) \cdot f(y) > f(z) + \langle \nabla f(z), t \cdot a + (1-t) \cdot y - z\rangle = f(z),$$

where the equality follows since by definition $z = t \cdot x + (1-t) \cdot y$ Hence, we obtained the definition of strict convexity. ◀

(5.c) [3pts] Prove that any strongly convex function (for any $\mu > 0$) is strictly convex.

Solution. By definition, f is μ-strongly convex if $f(x) - \frac{\mu}{2}\|x\|_2^2$ is convex. Pick any $x, y \in \Omega$, $x \neq y$, and $t \in (0,1)$. By convexity of $f(x) - \frac{\mu}{2}\|x\|_2^2$, we must have

$$\begin{aligned}&f((1-t)x + ty) - \frac{\mu}{2}\|(1-t)x + ty\|_2^2 \\ &\qquad\qquad \le (1-t)\Big(f(x) - \frac{\mu}{2}\|x\|_2^2\Big) + t\Big(f(y) - \frac{\mu}{2}\|y\|_2^2\Big).\end{aligned}$$

Rearranging,

$$\begin{aligned}&f((1-t)x + ty) \le (1-t)f(x) + tf(y) \\ &\qquad\qquad -\frac{\mu}{2}\Big((1-t)\|x\|_2^2 + t\|y\|_2^2 - \|(1-t)x + ty\|_2^2\Big).\end{aligned}$$

The proof is then complete if we can show that

$$(1-t)\|x\|_2^2 + t\|y\|_2^2 - \|(1-t)x + ty\|_2^2 > 0$$

for all $x \neq y, t \in (0,1)$. Expanding the last squared norm, we have

$$\begin{aligned} &(1-t)\|x\|_2^2 + t\|y\|_2^2 - \|(1-t)x + ty\|_2^2 \\ &= (1-t)\|x\|_2^2 + t\|y\|_2^2 - (1-t)^2\|x\|_2^2 - t^2\|y\|_2^2 - 2t(1-t)\langle x, y\rangle \\ &= t(1-t)\|x\|_2^2 + t(1-t)\|y\|_2^2 - 2t(1-t)\langle x, y\rangle \\ &= t(1-t)\|x-y\|_2^2. \end{aligned}$$

The last quantity is indeed strictly positive, as $t \in (0,1)$ and $x \neq y$. ◂

(5.d) [5pts] Prove that the function x^4 on $\mathbb{R}$ is strictly convex but not strongly convex. From the strict convexity of x^4, conclude that the condition $\nabla^2 f(x) \succ 0$ at all $x \in \Omega$ is *not necessary* for strict convexity.

▸ *Hint:* Problem (5.c) implies that the function x^2 is strictly convex.

Solution. Let $x \neq y \in \mathbb{R}$, and $t \in (0,1)$. Then,

$$\begin{aligned} ((1-t)x + ty)^4 &= \left(((1-t)x + ty)^2\right)^2 \\ &< \left((1-t)x^2 + ty^2\right)^2 && \text{strict conv. of inner square} \\ &< (1-t)x^4 + ty^4 && \text{strict conv. of outer square} \end{aligned}$$

which is the definition of strict convexity. We remark that the first inequality used the fact that the argument of the outer square is nonnegative (otherwise, in general, it is not true that $a < b \implies a^2 < b^2$).

The function's second derivative is $12x^2$ which is zero at $x = 0$ so the function is not strongly convex by 5.a. ◂

(5.e) [5pts] Prove that if f is differentiable, μ-strong convexity is equivalent to the global linear lower bound

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|_2^2 \qquad \forall x, y \in \Omega.$$

Solution. Let $g(x) \coloneqq f(x) - \frac{\mu}{2}\|x\|_2^2$. By definition, $g(x)$ is convex. Since f is differentiable, then so if g, and we can use Theorem L4.3 to establish the equivalent characterization, valid for all $x, y \in \Omega$,

$$\begin{aligned}
& g(y) \geq g(x) + \langle \nabla g(x), y - x \rangle \\
\iff \quad & f(y) - \frac{\mu}{2}\|y\|_2^2 \geq f(x) - \frac{\mu}{2}\|x\|_2^2 + \langle \nabla f(x) - \mu x, y - x \rangle \\
\iff \quad & f(y) \geq f(x) - \langle \nabla f(x), y - x \rangle + \mu\left(\frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|x\|_2^2 - \langle x, y - x \rangle\right) \\
\iff \quad & f(y) \geq f(x) - \langle \nabla f(x), y - x \rangle + \mu\left(\frac{1}{2}\|y\|_2^2 + \frac{1}{2}\|x\|_2^2 - \langle x, y \rangle\right) \\
\iff \quad & f(y) \geq f(x) - \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|_2^2,
\end{aligned}$$

as we wanted to show. ◂

(5.f) [5pts] Show that if Ω is not open, the condition $\nabla^2 f(x) \succeq 0$ not necessary for convexity. Similarly, the condition $\nabla^2 f(x) \succeq \mu I$ is not necessary for μ-strong convexity.

▸ *Hint:* Consider the function $f(x, y) := \frac{1}{2}(x^2 - y^2)$ on $\Omega := \{(x, 0) : x \in \mathbb{R}\}$. What is the Hessian matrix in this case? Is the function convex on Ω? Strongly convex?

Solution. The Hessian matrix of $f(x, y) := \frac{1}{2}(x^2 - y^2)$ is

$$\nabla^2 f(x, y) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which is *not* positive semidefinite. Nonetheless, the function f behaves exactly like the function x^2 on $\Omega = \left\{(x, 0) : x \in \mathbb{R}^2\right\}$, and is therefore strongly convex. So, a positive semidefinite Hessian is not necessary for convexity when Ω is not open. Similarly, the condition $\nabla^2 f(x) \succeq \mu I$ is not necessary for μ-strong convexity when Ω is not open. ◂

6. KKT Conditions [15pts]

Consider the problem

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

(6.a) [5pts] Write down the KKT conditions at a generic feasible point (x, y).

> *Solution.* The KKT conditions at the point (x, y) are:
>
> $$\begin{aligned} -(x+1) + 2\lambda_1 x = 0 \\ -(y+1) + 2\lambda_1 y + \lambda_2 = 0 \\ \lambda_1 \geq 0,\ x^2 + y^2 \leq 2,\ \lambda_1(x^2 + y^2 - 2) = 0 \\ \lambda_2 \geq 0,\ y \leq 1,\ \lambda_2(y-1) = 0, \end{aligned}$$
>
> where the feasibility of the point (x, y) was added explicitly to the set of conditions above. ◂

(6.b) [10pts] Find all points at which the KKT conditions hold by examining all the possible choices of active constraints. Which KKT point is the minimizer of the problem?

> *Solution.* The first two equations of the KKT conditions simplify to
>
> $$x = \frac{1}{2\lambda_1 - 1}, \qquad \text{and} \qquad y = \frac{1 - \lambda_2}{2\lambda_1 - 1} \tag{1}$$
>
> Based on whether two constraints are active, we consider the following cases:
>
> - $\lambda_1 > 0$ and $\lambda_2 > 0$: In this case we obtain $x^2 + y^2 = 2$ and $y = 1$, which give two points $(x, y) = (1, 1)$ and $(x, y) = (-1, 1)$. In the first case, we have $x = y$, and (1) implies $\lambda_2 = 0$. This contradicts our assumption that $\lambda_2 > 0$. In the second case, $(x, y) = (-1, 1)$, which implies $-1 = 2\lambda_1 - 1$ or $\lambda_1 = 0$, which is again impossible.
> - $\lambda_1 > 0$ and $\lambda_2 = 0$: We obtain from (1) that $x^2 + y^2 = 2$ and $x = y = \frac{1}{2\lambda_1 - 1}$. Thus, the points $(x, y) = (1, 1)$ and $(x, y) = (-1, -1)$ are the possible KKT points. At the point $(x, y) = (1, 1)$, we have $1 = \frac{1}{2\lambda_1 - 1}$, or $2\lambda_1 = 2$. Therefore, the point $(x, y) = (1, 1)$ is a KKT point with the corresponding multipliers $(\lambda_1, \lambda_2) = (1, 0)$. At the point $(x, y) = (-1, -1)$, we have $-1 = \frac{1}{2\lambda_1 - 1}$ or $\lambda_1 = 0$, which is impossible.

- $\lambda_1 = 0$ and $\lambda_2 > 0$: We get $y = 1$, $x = -1$ and $\lambda_2 = 2$. Thus, the point $(-1, 1)$ is a KKT point with the corresponding multipliers $(\lambda_1, \lambda_2) = (0, 2)$.
- $\lambda_1 = 0$ and $\lambda_2 = 0$: Equation (1) gives the point $(x, y) = (1, 1)$, which is a KKT point.

Therefore, the three KKT points and their corresponding multipliers are

$$\begin{pmatrix}1\\1\end{pmatrix}, \lambda = \begin{pmatrix}1\\0\end{pmatrix}; \qquad \begin{pmatrix}-1\\1\end{pmatrix}, \lambda = \begin{pmatrix}0\\2\end{pmatrix}; \qquad \begin{pmatrix}-1\\-1\end{pmatrix}, \lambda = \begin{pmatrix}0\\0\end{pmatrix}.$$

Clearly, $(x, y) = (1, 1)$ is the global minimizer. ◂

6.7220 Nonlinear Optimization
Spring 2025