MIT 6.7220/15.084 — Nonlinear Optimization (Spring '25) Tue, Apr 1st 2025

Midterm Analysis

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

This short note discusses several aspects of the midterm exam:

1. What each problem was designed to test;
2. The distribution of scores on each question;
3. A discussion of the failure modes that were encountered for each question.

MA.1 Problem 1: Optimality conditions

Let $\Omega \subseteq \mathbb{R}^n$ be closed and convex, and $f : \Omega \to \mathbb{R}$ convex and differentiable.

(1.a) [15pts] Let $\Pi_\Omega(y)$ denote the Euclidean projection of y onto Ω, and $\eta > 0$ be an arbitrary positive real. Prove that $x \in \Omega$ is a minimizer of f on Ω if and only if

$$x = \Pi_\Omega(x - \eta \nabla f(x)).$$

(1.b) [10pts] If f is differentiable but not convex, is the condition in Problem (1.a):

- Still necessary?
- Still sufficient?

For both questions, either produce an argument or show a counterexample.

*These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

Problem 1 was designed to test understanding of first-order optimality conditions, and especially their sufficiency in the case of convex functions.

MA.1.1 The solution to (1.a)

There are two optimization problems going on here: x being a minimizer of f, and x being a projection of $x - \eta\nabla f(x)$, *i.e.*, a minimizer of the function

$$g(y) := \frac{1}{2}\|y - x + \eta\nabla f(x)\|_2^2$$

on Ω. Both problems involve convex objective functions being minimized on convex sets. Hence, we know from Theorem L4.2 that first-order optimality conditions for these problems are both necessary and sufficient—and hence *equivalent*—to optimality. Writing optimality conditions, we find

$$x \quad \text{optimal for} \quad \min_{y\in\Omega} f(y) \qquad\Updownarrow\qquad \langle \nabla f(x), y - x\rangle \geq 0 \qquad \forall y \in \Omega.$$

$$x \quad \text{optimal for} \quad \min_{y\in\Omega} \frac{1}{2}\|y - x + \eta\nabla f(x)\|_2^2$$

$$\Updownarrow$$

$$\begin{aligned} & \langle \nabla g(x), y - x\rangle \geq 0 \quad \forall y \in \Omega \\ \iff\ & \langle x - x + \eta\nabla f(x), y - x\rangle \geq 0 \quad \forall y \in \Omega \\ \iff\ & \eta\langle \nabla f(x), y - x\rangle \geq 0 \quad \forall y \in \Omega \\ \iff\ & \langle \nabla f(x), y - x\rangle \geq 0 \quad \forall y \in \Omega. \end{aligned}$$

Since the first-order optimality conditions are the same, and they are *equivalent* to statements "x optimal for f on Ω" and "x optimal for g on Ω", the result follows.

MA.1.2 The solution to (1.b)

The convexity of $g(y)$ is independent on the convexity of f, and always holds. So, the equivalence

$$x = \Pi_\Omega(x - \eta\nabla f(x)) \qquad \iff \qquad \langle \nabla f(x), y - x\rangle \geq 0 \quad \forall y \in \Omega$$

remains true no matter the convexity of f. However, the latter condition is only necessary, and no longer sufficient in general, for x to be a minimizer of f on Ω.

It is easy to construct examples of nonconvex functions f for which a point satisfying the first-order optimality condition is not a minimizer. For example, consider

$$f(x) = -x^2 \quad \text{on} \quad \Omega = \mathbb{R}.$$

The point $x = 0$ is certainly not a minimizer of f, but it satisfies the first-order optimality condition $\nabla f(x) = 0$. And indeed we have $\Pi_\Omega(x - \eta\nabla f(x)) = x$.

MA.1.4 Analysis of performance in (1.a)

Most students recognized that writing the first-order optimality condition for the projection problem was a good starting point. From there, seeing that the condition simplifies to the first-order optimality condition for f was direct. Overall, the only consistent failure mode was due to people failing to write first-order optimality conditions.

A very small number of people got confused with variable names: the projection of $x - \eta \nabla f(x)$ onto Ω is the solution to the optimization problem

$$\begin{aligned} \min_y \quad & \frac{1}{2}\|y - x + \eta \nabla f(x)\|_2^2. \\ \text{s.t.} \quad & y \in \Omega \end{aligned}$$

Some people, however, reused x as the name of the optimization variable, leading to the very different optimization problem

$$\begin{aligned} \min_x \quad & \frac{1}{2}\|x - x + \eta \nabla f(x)\|_2^2 = \frac{\eta^2}{2}\|\nabla f(x)\|_2^2. \\ \text{s.t.} \quad & x \in \Omega \end{aligned}$$

MA.1.5 Analysis of performance in (1.b)

Most students correctly identified that the condition in (1.a) is still necessary, but not sufficient, for x to be a minimizer of f on Ω. This was true also of students that struggled with the first part of the question.

A somewhat large number of students failed to produce a counterexample to the sufficiency. This was penalized very lightly (*i.e.*, only one point) in cases where the solution made otherwise clear what specifically in the proof of (1.a) breaks when convexity of f is removed.

MA.2 Problem 2: Separation

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex differentiable function, and consider the set

$$\Omega := \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : t \ge f(x)\}.$$

This set is closed.

(2.a) [6pts] Is the set Ω guaranteed to be

- (2.a) Convex?
- (2.b) A cone?

For both of the questions, either produce a proof or a counterexample.

(2.b) [14pts] Given a point $(y, w) \notin \Omega$, prove that the vector

$$u := (-\nabla f(y),\ 1)$$

is always a direction separating (y, w) from Ω.

(2.c) [5pts] Construct an efficient separation oracle for the set

$$\Omega := \left\{(x, t) \in \mathbb{R}^n \times \mathbb{R} : t \ge \|x\|_2^2\right\}.$$

Problem 2 was designed to test three important concepts: the definition of convex set and cone; that convex functions are lower bounded by their linearization at any point; and separation oracles.

MA.2.1 The solution to (2.a)

The set is guaranteed to be convex. To see this, let $(x_1, t_1), (x_2, t_2) \in \Omega$, and $\lambda \in [0, 1]$ be arbitrary. Then,

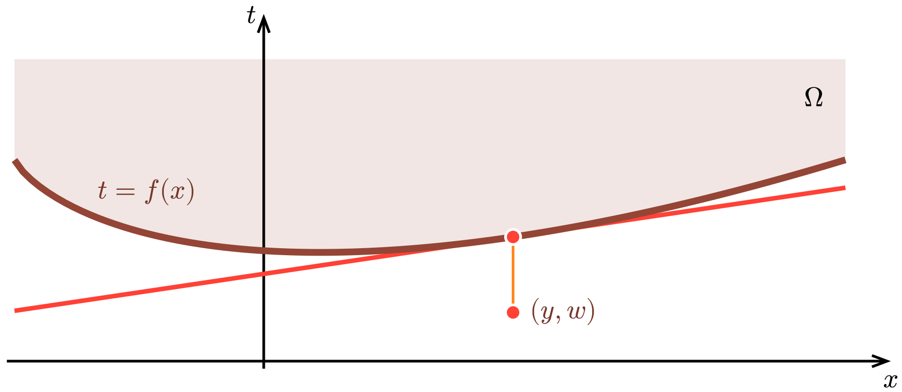
$$\begin{aligned} \lambda t_1 + (1-\lambda)t_2 &\ge \lambda f(x_1) + (1-\lambda) f(x_2) && \text{(by definition of } \Omega) \\ &\ge f(\lambda x_1 + (1-\lambda)x_2). && \text{(by convexity of } f) \end{aligned}$$

This shows that $\lambda(x_1, t_1) + (1-\lambda)(x_2, t_2) \in \Omega$, and so Ω is convex.

The set is *not* guaranteed to be a cone. For example, take the function $f(x) = \|x\|_2^2$. For any $x \neq 0$, the point $(x, \|x\|_2^2)$ belongs to Ω. Yet, $(2x, 2\|x\|_2^2)$ does *not* belong, since $2\|x\|_2^2 \not\ge \|2x\|_2^2$.

MA.2.2 The solution to (2.b)

This is just a direct application of the linear lower bound for convex functions (see Lecture 4). It becomes even more evident if we try to draw the set Ω.



Concretely, consider the linearization of f around y. By convexity,

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \qquad \forall x \in \mathbb{R}^n.$$

Since $t \geq f(x)$ for all $(x, t) \in \Omega$, and $w < f(y)$ by the hypothesis that $(y, w) \notin \Omega$, the previous inequality then implies

$$t > w + \langle \nabla f(y), x - y \rangle \qquad \forall (x, t) \in \Omega,$$

which is exactly the statement unpon rearranging.

MA.2.3 The solution to (2.c)

This is a straightforward application of (2.b). Given a point (y, w), we can first check whether $w \geq \|y\|_2^2$. If yes, we return "$(y, w) \in \Omega$". Else, we return "$(y, w) \notin \Omega$", together with the separating direction $(-2y, 1)$.

MA.2.5 Analysis of performance in (2.a)

Most students did well in this question.

A small number of students claimed that the set Ω is a sublevel set of f. This is incorrect. A sublevel set has the form $\{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ for some $\alpha \in \mathbb{R}$. Instead, the set Ω is what is called the *epigraph* of the function f. Nonetheless, the epigraph of f can be seen as the sublevel set of the convex function $f(x) - t$ for the value $\alpha = 0$.

A small number of people wrongly claimed Ω is always guaranteed to be a cone. Many examples could have immediately been produced to show that this is not the case, including consider a *constant* function such as $f(x) = 1$.

MA.2.6 Analysis of performance in (2.b)

This question resulted harder than expected, despite the proof being a couple of lines. Some people were on the right track, and correctly suspected that the lower bound property of convex functions would play a key role. Others even mentioned the geometric intuition but could not formalize it. Some people who claimed that Ω is a cone in Problem (2.a) tried to use specific properties of conic separation ($v = 0$).

MA.2.7 Analysis of performance in (2.c)

This question was a trivial application of the general result in (2.b), and most students got it right.

Surprisingly, around 15% of students did not calculate the gradient of $\|x\|_2^2$, or made a mistake in calculating it.

Somewhat unexpectedly, a small group of students claimed wrongly that the set Ω is the Lorentz cone. This is incorrect, as the Lorentz cone is defined as the set of points (x, t) such that $t \geq \|x\|_2$, rather than $t \geq \|x\|_2^2$ (note the square). In fact, the set in (2.c) is *not* a cone.

MA.3 Problem 3: An optimization problem

Consider the optimization problem

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

(3.a) [5pts] Is the objective function, seen as a function from $\mathbb{R}^2$ to $\mathbb{R}$, convex? Provide a proof or a counterexample.

(3.b) [7pts] Sketch the feasible set. Graphically, what is the solution of the problem?

(3.c) [5pts] Is the feasible set a convex subset of $\mathbb{R}^2$? Provide a proof or a counterexample.

(3.d) [8pts] Write the KKT conditions at the optimal point. Do the KKT conditions hold?

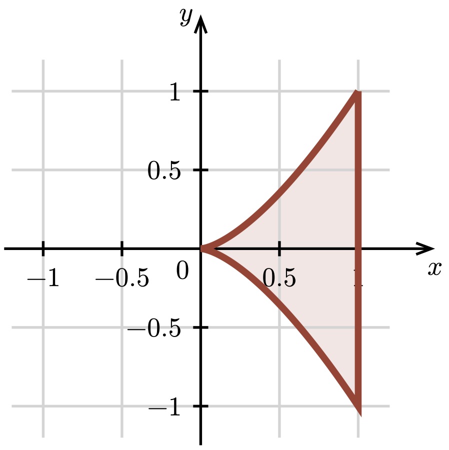
Problem 3 was designed to test the understanding of KKT conditions. It was not designed to test whether students are able to sketch simple functions, although a failure mode we observed was that some students had trouble sketching the function $f(x) = x^{1.5}$.

MA.3.1 The solution to (3.a)

The Hessian matrix of the function is equal to the identity, which is positive semidefinite. Hence, the objective function is convex.

MA.3.2 The solution to (3.b)

The condition $y^2 \le x^3$ is equivalent to $x \ge 0 \wedge -x^{1.5} \le y \le x^{1.5}$. The graph of the function $x^{1.5}$ passes through the points $(0,0)$ and $(1,1)$, and falls in between the graphs for the functions x and x^2.



From the plot it is apparent that the optimal solution is the point $(0, 0)$.

MA.3.3 The solution to (3.c)

The feasible set is not convex (this is visible in the above plot too). In particular, the points $(0, 0)$ and $(1, 1)$ are both feasible. Yet, their midpoint $(0.5, 0.5)$ is not feasible, as can be checked directly from the fact that $-(1/2)^3 + (1/2)^2 = 1/8 > 0$.

MA.3.4 The solution to (3.d)

The KKT conditions at the optimal point $(0, 0)$ are the existence of the Lagrange multipliers $\alpha, \beta \geq 0$ such that

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

This system has no solution, so the KKT conditions do *not* hold at the optimal point.

MA.3.6 Analysis of performance in (3.a)

Most students had no problem with this question.

MA.3.7 Analysis of performance in (3.b)

A non-negligible fraction of the students struggled to sketch the graph of the function $f(x) = x^{1.5}$. This was somewhat unexpected. Instead of drawing the graph to be somewhat intermediate between that of x and x^2, some people draw it to resemble more closely the graph of $x^{0.5}$. This led some people to wrongly claim that the set Ω is convex.

The confusion regarding whether the function $x^{1.5}$ passes below or above the line $y = x$ could have been resolved easily. Beyond the fact that $x^{1.5}$ is intermediate between x^2 and x, one approach could have been to evaluate the function at $x = 1/2$. Another approach could have been to evaluate the derivative at 0 and 1, and see that the function starts flat at $x = 0$ and has a slope of 1.5 at $x = 1$.

MA.3.8 Analysis of performance in (3.c)

Overall, most people had no problem with the question, and expectedly pointed out how the point $\left(\frac{1}{2}, \frac{1}{2}\right)$ is not contained in Ω despite $(0, 0)$ and $(1, 1)$ both being in Ω.

For those that struggled with the question, there was correlation with failing to sketch the function $f(x) = x^{1.5}$.

MA.3.9 Analysis of performance in (3.d)

This question also did not pose any particular difficulty to most students. The major source of point deductions was due to failing to write dual feasibility and complementary slackness.

MA.4 Problem 4: Convex functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and continuous, but not necessarily differentiable.

(4.a) [6pts] Is it guaranteed that f attains a minimum on $\mathbb{R}^n$? When it does, is it guaranteed that the minimizer is unique? For both questions, give either a proof or a counterexample that works for all dimensions n.

(4.b) [6pts] Assume further that f is **strictly convex**. Is it guaranteed that f attains a minimum on $\mathbb{R}^n$? When it does, is it guaranteed that the minimizer is unique? For both questions, give either a proof or a counterexample that works for all dimensions n.

▸ *Hint:* It is fine to refer to, without proof, results shown in class. You can also take for granted the fact that $\nabla^2 f(x) \succ 0$ at all $x \in \mathbb{R}^n$ is a *sufficient* condition for strict convexity.

(4.c) [13pts] Assume further that f is μ-**strongly convex** for some $\mu > 0$, that is,

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

Is it guaranteed that f attains a minimum on $\mathbb{R}^n$? When it does, is it guaranteed that the minimizer is unique? For both questions, give either a proof or a counterexample that works for all dimensions n.

▸ *Hint:* The problem is *not* assuming that f is differentiable. However, starting from considering the differentiable case might give useful insights.

This question was meant to probe the understanding of the existence and uniqueness of minima for convex functions. Question (4.a) was supposed to be extremely easy: under no additional assumptions, convexity does not buy anything in terms of attainment and uniqueness of minima, as linear functions already show. Question (4.b) required just a bit more thought, although the heavy lifting was a theorem we saw in class, and the hint was suggesting the answer for the first part of the question. Question (4.c) was left as a slightly more challenging question, whose solution involved ideas that were individually discussed in class or in the homework, but never required in combination.

MA.4.1 The solution to (4.a)

It is not guaranteed that convex functions attain minima on $\mathbb{R}^n$. For example, the linear (and hence convex) function $f(x) = 1^\top x$ can be made arbitrarily small. Alternatively, the function $f(x) = \sum_{i=1}^n e^{-x_i}$ can be made arbitrarily close to its infimum 0, without ever attaining the value.

Uniqueness is also not guaranteed. For example, the trivially convex function $f(x) = 0$ has infinitely many minimizers.

MA.4.2 The solution to (4.b)

Again, it is not guaranteed, even for strictly convex functions, that minima are attained. For example, the function $f(x) = \sum_{i=1}^{n} e^{-x_i}$ is strictly convex, since

$$\nabla^2 f(x) = \begin{pmatrix} e^{-x_1} & & \\ & \ddots & \\ & & e^{-x_n} \end{pmatrix} \succ 0 \qquad \forall x \in \mathbb{R}^n.$$

As noted above, the function can be made arbitrarily close to its infimum 0 without ever attaining the value.

On the other hand, uniqueness *is* guaranteed for strictly convex functions, as proved in class in Lecture L4.5.

MA.4.3 The solution to (4.c)

Since strongly convex functions are strictly convex, the uniqueness of the minimizer is guaranteed by (4.b). The non-obvious question is the attainment of the minimum. As it turns out, strongly convex functions always attain a minimum, and the proof is very similar to that of the existence of projections onto closed sets, using the Weierstrass theorem after appropriately restricting the focus on a bounded subset of the domain where the minimizer must exist (see Lecture 1).

▸ **The general idea.** By definition, a strongly convex function can be written in the form

$$f(x) = (\text{convex function}) + \frac{\mu}{2}\|x\|_2^2.$$

Since convex functions are lower bounded by their linearization at any point (see Lecture 4), for large $\|x\|_2$ the squared norm dominates the right-hand side. This immediately implies that all sublevel sets of f must be bounded, and so the Weierstrass theorem guarantees the existence of a minimizer.

▸ **Formalizing the argument when f is differentiable.** Let's formalize the above idea by following the hint and considering the case of a function f differentiable at 0. In this case, we know that the linearization around 0 gives a global linear lower bound

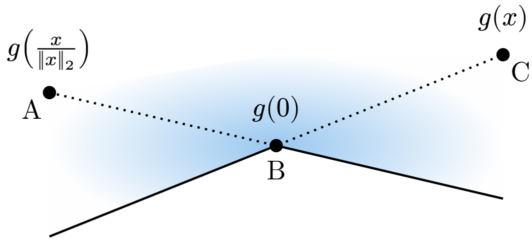
$$f(x) \geq f(0) + \langle \nabla f(0), x\rangle + \frac{\mu}{2}\|x\|_2^2 \qquad \forall x \in \mathbb{R}^n.$$

The right-hand side is dominated by the quadratic term for large $\|x\|_2$, and so we can always find a value $\alpha \geq 0$ such that $\|x\|_2 \geq \alpha$ guarantees $f(x) \geq f(0)$. If a minimizer exists, it must then belong to the closed ball $\mathbb{B}_\alpha(0)$ of radius α centered in 0, and we can apply the Weierstrass theorem.

Specifically, it is easy to see that $\alpha = \frac{2}{\mu}\|\nabla f(0)\|_2$ suffices, as any x with $\|x\|_2 \geq \alpha$ satisfies

$$\begin{aligned} \langle \nabla f(0), x\rangle + \frac{\mu}{2}\|x\|_2^2 &\geq -\|\nabla f(0)\|_2\|x\|_2 + \frac{\mu}{2}\|x\|_2^2 && \text{(Cauchy-Schwarz)} \\ &\geq -\|\nabla f(0)\|_2\|x\|_2 + \frac{\mu}{2}\alpha\|x\|_2 = 0. && \text{(since } \|x\|_2 \geq \alpha) \end{aligned}$$

▸ **The general case.** When f is not assumed differentiable, we might not be able to write the linearization of a convex function using the gradient at our favorite point (in the proof above, 0). Several technical ways to sidestep the issue are possible and quite straightforward. For those with more background in real analysis, the use of Rademacher's theorem ensures that f is differentiable almost everywhere (in the technical sense that the points of non-differentiability have measure 0), and so the proof could continue as in the previous point upon picking one of the points at which the function is differentiable. For a proof from first principles, we can use the same idea as Problem 3 in PSet 2, where we argued that the graph of the function must be lower bounded by a cone



Concretely, any convex function g must satisfy, by definition of convexity,

$$g\left(\frac{x}{\|x\|_2}\right) \le \left(1 - \frac{1}{\|x\|_2}\right) g(0) + \frac{1}{\|x\|_2} g(x) \qquad \forall x \in \mathbb{R}^n : \|x\|_2 \ge 1,$$

which implies

$$g(x) \ge g(0) + \|x\|_2 \left(g\left(\frac{x}{\|x\|_2}\right) - g(0) \right).$$

Letting $M := \min_{\|x\|_2 \le 1} \{g(x) - g(0)\}$ (note that this value exists thanks to the Weierstrass theorem, and furthermore $M \le 0$), we can then write

$$g(x) \ge g(0) + M\|x\|_2 \qquad \forall x \in \mathbb{R}^n : \|x\|_2 \ge 1.$$

Hence, we obtain a "conic" first-order lower bound for g that does not rely on the gradient of g. We can therefore continue the proof as in the case of differentiability. In particular, applying the result we just established to the convex function $f(x) - \frac{\mu}{2}\|x\|_2^2$, we find that

$$f(x) \ge f(0) + M\|x\|_2 + \frac{\mu}{2}\|x\|_2^2 \qquad \forall x \in \mathbb{R}^n : \|x\|_2 \ge 1.$$

By restricting to the ball of radius $\alpha := \max\{1, -2M/\mu\}$ centered at 0, we can then apply the Weierstrass theorem to conclude that f attains a minimum.

MA.4.5 Analysis of performance in (4.a)

Most people had not problems with this question. The only real failure mode that was observed was that some students claimed that convex functions always attain minima. This is not true, and it is clearly already false in the special case of *linear* functions. This was more of a conceptual question than a question on specific knowledge.

Out of those that failed the question, a large fraction invoked the Weierstrass theorem, without realizing that the theorem only guarantees the existence of minima on *compact* (*i.e.*, closed and *bounded*) sets. Obviously, $\mathbb{R}^n$ is not bounded, and so the Weierstrass theorem is not applicable.

The functions used in this subproblem were very complicated for no apparent reason.

MA.4.6 Analysis of performance in (4.b)

Most people had no problem with this question.

Some students claimed wrongly that the function $\exp\left(\sum_i x_i\right)$ is strictly convex.

MA.4.7 Analysis of performance in (4.c)

Most people got the uniqueness part correct by realizing that strong convexity implies strict convexity. Most people did not answer the existence part.