

Lecture 1

Introduction

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

L1.1 Mathematical optimization, and this course

One cannot overstate how pervasive and far-reaching mathematical optimization is. Optimization is everywhere. This is true in two senses:

- Nearly all human endeavors and designs are driven by an aspiration to optimize: minimize risk, maximize reward, reduce energy consumption, train a neural network to minimize model loss, *et cetera*.
- Nature’s phenomena themselves seem to be guided by optimization—the principle of least action, Hamiltonian mechanics, the Rayleigh-Ritz variational principle (from which the Schrödinger equation can be derived), *et cetera*.¹

But the benefits of an optimization mindset are also more subtle. For example, optimization in the space of semidefinite matrices has led to important breakthroughs in classical combinatorial optimization problems in graph theory, such as the MAX-CUT problem [GW95]. Optimization theory can also be used to show the *fundamental theorem of algebra*, and amusingly Güler presents that on page 34 (!) of his “Foundations of Optimization” book [Gül10], adapting some ideas by Terkelsen, F. [Ter76] and Fefferman, C. [Fef67].

The literature on optimization dates back to (at least) the ancient Greeks, with their work on the *isoperimetric problem*. The problem, which asks to construct the shape that maximizes the area given a fixed perimeter, is at the center of the legend of how Carthage—one of the most powerful empires of the classical world—was founded.² It was solved (to the standards of rigor of those times, at least) by Zenodorus.³

Fast forward to today, mathematical optimization is a vast field, subdivided into a myriad of subfields depending on what types of optimization problems are being solved; here are just a few:

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

¹See https://en.wikipedia.org/wiki/Variational_principle for a longer list.

²Dido was the legendary founder—and first queen—of Carthage, a Phoenician city-state that would then become one of the most powerful empires in the world, and will be destroyed by the Roman empire. As recounted by Blåsjö, V. [Blå05], the legend says that “Dido, daughter of the king of Tyre, fled her home after her brother had killed her husband”. She later “ended up on the north coast of Africa, where she bargained to buy as much land as she could enclose with an oxhide. So she cut the hide into thin strips, and then she faced, and presumably solved, the isoperimetric problem.”

³Approx 150BC, [https://en.wikipedia.org/wiki/Zenodorus_\(mathematician\)](https://en.wikipedia.org/wiki/Zenodorus_(mathematician))

- Linear optimization refers to problems that seek to minimize or maximize linear objectives subject to linear constraints imposed on continuous optimization variables.
- Integer linear optimization is akin to linear optimization, but variables can be further restricted to only taking integer values.
- Combinatorial optimization more generally studies optimization problems over finite sets.
- Convex optimization studies optimization of convex objectives on convex sets.
- Calculus of variations studies optimization problems in function spaces (that is, where the optimal object that is sought is a function).

L1.1.1 Focus and objectives of this course

This is a graduate-level course on *nonlinear* optimization. The main focus of this course is *continuous optimization in finite dimensional spaces*. As such, nonlinear optimization encompasses a large variety of topics and algorithms, with applications that range from semidefinite relaxations of combinatorial problems, to regression problems, to training of deep neural networks. I foresee that this course attracts students from (at least) two very different categories:

- Some of you are interested in *nonlinear optimization theory*. Typically, you're facing an optimization problem that you hope to solve in closed form, so that you can plug in or reason about the solution as part of a larger mathematical endeavor.
- Some of you are interested in *nonlinear optimization algorithms*. Typically, you are facing an optimization problem for which you simply want a numerical solution.

In this course, the TAs and I have worked hard to cater to *both* of these categories of students.

L1.1.2 Prerequisites

This class is pretty light on prerequisites. Fundamentally, the only prerequisites are Linear algebra (18.06) and Real analysis (18.100A/B/Q). Beyond some basic mathematical concepts (continuity, vector spaces, some basic point-set topology), we will try to keep this course as self-contained as possible.

L1.2 Optimization problems: Preliminary considerations

In general, an optimization problem has the form

$$\begin{array}{ll} \min_x & f(x) \\ \text{s.t.} & x \in \Omega \end{array}$$

In terms of nomenclature:

- The function we are trying to minimize (or maximize), in this case $f(x)$, is called the *objective function*.
- The entries of the vector x , are called *optimization variables*.
- The set Ω is called the *feasible set* or *constraint set*.

In particular, in this class we are interested in optimization problems that satisfy certain assumptions:

- First of all, we assume that Ω is a subset of some ambient *finite* Euclidean space E . Without loss of generality, $E = \mathbb{R}^n$.
- The objective function $f(x)$ is a *continuous* real-valued function. Very often, we also assume it is differentiable.
- A category of problems we are paying attention to are those with *functional constraints*, that is those in which the feasible set is defined as

$$\Omega := \left\{ x \in E \mid \begin{array}{ll} g_i(x) \leq 0, & i = 1, \dots, r \\ h_j(x) = 0, & j = 1, \dots, m \end{array} \right\}. \quad (1)$$

When faced with an optimization problem, two questions are immediate:

1. Does a solution *exist*?
2. If yes, can we *compute* it?

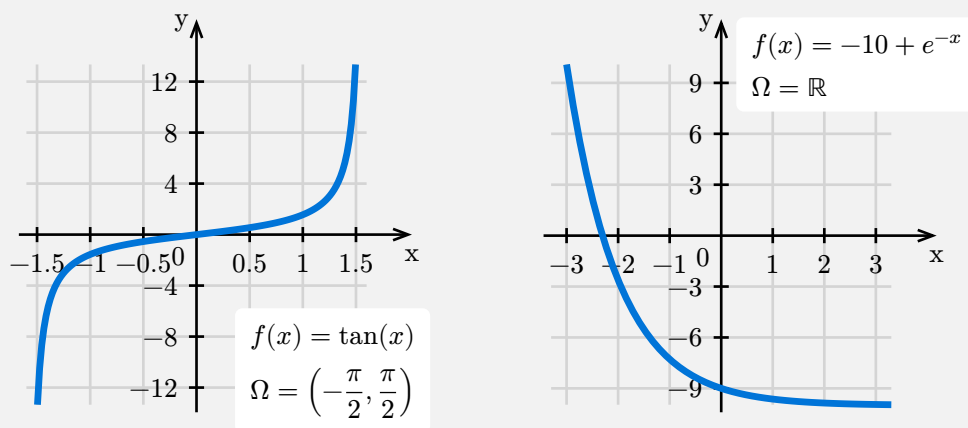
We will look into some preliminary observations regarding these two questions in Sections L1.2.1 and L1.2.2, respectively.

L1.2.1 Existence of solutions: Weierstrass theorem

There are two reasons why an optimization problem might not attain a minimum:

- Ω is empty, that is, no feasible points exist at all; or
- for any candidate minimizer, there always exists another point that achieves a lower value, as shown in the following two examples.

Example L1.1. Consider minimization of the following two functions, both over their domain:



In the plot on the left, the objective function is *unbounded*: values can be made arbitrarily small. No minimizer exists.

In the plot on the right, the objective has a lower bound, but for any value of x , all the values $x' > x$ achieve a *strictly smaller value* than $f(x)$. Again, no minimizer exists.

So, assuming that the objective f is continuous, when can we be certain that an optimizer exists? If we inspect Example L1.1, we notice that both functions are continuous. In the first example, the domain is open (the extrema $\pm\frac{\pi}{2}$ are not in the domain) and bounded. In the second example, the domain is unbounded and closed. This shows that, if there is any hope of establishing a general result on the existence of minimizers, such a result must require *both closedness and boundedness*. We are in luck:

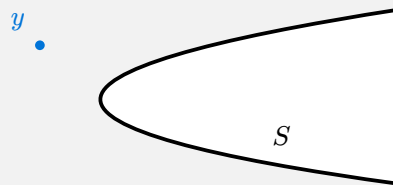
Theorem L1.1 (Weierstrass theorem). Let $f : \Omega \rightarrow \mathbb{R}$ be a *continuous* function defined on a nonempty and *compact* (i.e., closed and bounded) set Ω . Then, there exists a minimizer $x^* \in \Omega$ of f on Ω , that is,

$$f(x^*) \leq f(x) \quad \text{for all } x \in \Omega.$$

A note on boundedness. Weierstrass theorem is a pretty universal tool that you can keep in your toolbox for when you need to argue that an optimization problem has a solution. However, many optimization problems do not start off with a bounded feasible set Ω . Consider the following example:

Example L1.2. Let $y \in \mathbb{R}^n$ be a point, and $S \subseteq \mathbb{R}^n$ be a *closed* but not necessarily bounded set. Prove that there always exists a projection point

$$\begin{aligned} \min_x \quad & \|x - y\| \\ \text{s.t.} \quad & x \in S. \end{aligned}$$



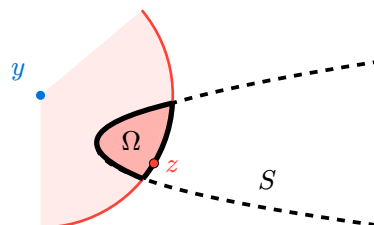
Solution. The idea is to show that we can safely restrict our attention to a compact subset of the feasible set of the problem—in this case, S . By “safely”, we mean that the minimizer does not change before and after making the modification.

In this case, this is simple: pick an arbitrary point $z \in S$. Then, we know that for sure, if a minimizer exists, it must live in the intersection

$$\Omega := S \cap \overline{\mathbb{B}}_{\|z-y\|}(y)$$

between S and the closed ball centered in y of radius $\|y - z\|$ (see the figure on the side).

Now, Ω is nonempty (it contains z), closed (since both S and the ball are closed and intersection preserves closedness) and bounded (since the ball has finite radius). So, since the objective function is continuous, we can invoke Weierstrass theorem and conclude the proof. □



More generally, we have the following:

Theorem L1.2 (Weierstrass theorem for compact sublevel sets). Let f be a continuous function defined on a set S . If f has a nonempty and compact *sublevel set*, that is, there exists $\alpha \in \mathbb{R}$ such that

$$\{x \in S : f(x) \leq \alpha\}$$

is nonempty, bounded, and closed, then f has a minimizing point in S .

A note on closedness. When the feasible set is bounded but not closed, things are slightly more tricky. However, in many cases you will still be able to use Weierstrass theorem. As a rule of thumb, there are two ways to deal with this:

- Show that you can safely restrict your attention to a compact subset of the feasible set, as in Example L1.2.
- Start by considering the closure of the feasible set, and then show that the objective function is continuous on the closure. After applying Weierstrass theorem, you can then argue that the minimizer is not one of the boundary points you added when taking the closure. This approach could for example be used in the following application, which comes up naturally in online learning and reinforcement learning: entropy-regularized best responses.

Example L1.3. Let Δ^n denote the set of probability distributions over n actions $\{1, \dots, n\}$ that have *full* support, that is, the *open* set

$$\Delta^n := \{(x_1, \dots, x_n) \in \mathbb{R}_{>0}^n : x_1 + \dots + x_n = 1\}.$$

Given an assignment of *values* v_i for each action $i = 1, \dots, n$, the *entropy-regularized best response* given the values is the distribution that solves the following problem:

$$\begin{aligned} \min_x \quad & g(x) := - \sum_{i=1}^n v_i x_i + \sum_{i=1}^n x_i \log x_i \\ \text{s.t.} \quad & x \in \Delta^n, \end{aligned}$$

Show that a minimizer exists.

(You should try to solve this!)

L1.2.2 Nonlinear optimization is computationally hard

We now turn our attention to the second question: *what is reasonable to expect in terms of computational efficiency in finding the solution?*

Let me start with the good news: there are entire subclasses of nonlinear optimization problems that are computationally tractable (most notably, *convex* optimization problems, albeit with some exceptions). As we will see, even beyond these classes, usually it's possible to at least guarantee some form of convergence to some weaker, "local", notion of optimality.

However, in general, *finding minimizers of nonconvex problems is computationally intractable*. This means that we believe that there does not exist any algorithm that is able to solve *all* nonconvex optimization problems. This is because we believe that *linear integer programming* is already intractable, and nonconvex optimization captures known intractable problems such as MAX-CUT,⁴ already under benign assumptions such as quadratic functional constraints and objective.

Example L1.4 (MAX-CUT as a nonlinear optimization problem). MAX-CUT is one of the most fundamental problems in theoretical computer science and discrete optimization. The definition is very simple:

- Let $G = (V, E)$ be a graph, where $V := \{1, 2, \dots, n\}$ is the set of vertices.
- We want to partition V into two sets S and $V \setminus S$, such that *the number of edges going between S and $V \setminus S$ is maximized*.

Can you see how this optimization problem might be formulated as a nonlinear optimization problem with integer variables? Can you remove the constraint that the variables are integer, and convert it into a continuous problem?

Solution. As a general rule of thumb, the first we want to do when we go from a problem into its mathematical programming formulation, is to *pick variables*. In this case, it is intuitive that we need variables that encode whether each node belongs to S or $V \setminus S$. While several choices make perfect sense, let's go with the following. Let's assign to each node $v \in V$ a variable $x_v \in \{-1, +1\}$, with the meaning that:

- if $x_v = +1$, then $v \in S$;
- if $x_v = -1$, then $v \in V \setminus S$.

It is immediate to see that an edge $(u, v) \in E$ is a cut edge (that is, connecting one node in S to one node in $V \setminus S$) if and only if $x_u \cdot x_v = -1$. Else, $x_u \cdot x_v = +1$. Hence, we can count exactly all edges in the cut by considering the objective function

$$h(x) = \sum_{(u,v) \in E} \frac{1 - x_u \cdot x_v}{2}.$$

This leads to the *discrete* optimization problem

$$\begin{aligned} \max_x \quad & h(x) \\ \text{s.t.} \quad & x_v \in \{-1, +1\} \quad \forall v \in V. \end{aligned}$$

This problem can be readily written as a continuous, nonlinear optimization problem by noting that

$$x_v \in \{-1, +1\} \iff x_v^2 = 1.$$

This leads to the nonlinear optimization problem

⁴NP-completeness of MAX-CUT was shown by Garey, M. R., Johnson, D. S., & Stockmeyer, L. [GJS76].

$$\begin{aligned} \max_x \quad & \sum_{(u,v) \in E} \frac{1 - x_u \cdot x_v}{2} \\ \text{s.t.} \quad & x_v^2 = 1 \quad \forall v \in V, \end{aligned} \tag{2}$$

which is a quadratically-constrained quadratic optimization problem. \square

Despite the overall generally negative message, I want to end with a message of encouragement. Sure, we cannot hope to construct a machine that can solve all nonlinear optimization problems. But nonlinear optimization theory is still incredibly powerful, both because *many useful/powerful optimization problems can be solved efficiently*, and because *even for the hard problems, nonlinear optimization might have some tricks up its sleeve*. As a case in point, it turns out that the optimization problem in (2), which is hard, can be *relaxed* to a pretty non-obvious optimization problem that *can* be efficiently solved.

For those interested, this is obtained by introducing the relaxation

$$\begin{aligned} \max_y \quad & \sum_{(u,v) \in E} \frac{1 - y_{uv}}{2} \\ \text{s.t.} \quad & y_{vv} = 1 \quad \forall v \in V \\ & [y_{uv}] \succeq 0, \end{aligned}$$

where $[y_{uv}]$ denotes the matrix where the entry in row $u \in V$ and column $v \in V$ is y_{uv} and $\succeq 0$ means that the matrix just defined must be positive semidefinite. The important fact here is that this relaxation *can* be efficiently solved (we will see how later on in this course). With some clever rounding scheme, one can then recover a solution to the original MAX-CUT instance that is guaranteed to be 0.878-approximate. This algorithm, due to [Goemans, M. X., & Williamson, D. P. \[GW95\]](#) makes deep use of nonlinear optimization.

Bibliography for this lecture

- [GW95] Goemans, M. X., & Williamson, D. P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6), 1115–1145. <https://doi.org/10.1145/227683.227684>
- [Gül10] Güler, O. (2010). *Foundations of Optimization*. Springer. <https://link.springer.com/book/10.1007/978-0-387-68407-9>
- [Ter76] Terkelsen, F. (1976). The Fundamental Theorem of Algebra. *Am. Math. Mon.* <https://www.tandfonline.com/doi/abs/10.1080/00029890.1976.11994202>
- [Fef67] Fefferman, C. (1967). An Easy Proof of the Fundamental Theorem of Algebra. *The American Mathematical Monthly*, 74(7), 854–855. <http://www.jstor.org/stable/2315823>
- [Blå05] Blåsjö, V. (2005). The Isoperimetric Problem. *Am. Math. Mon.* <https://www.tandfonline.com/doi/pdf/10.1080/00029890.2005.11920227>
- [GJS76] Garey, M. R., Johnson, D. S., & Stockmeyer, L. (1976). Some simplified NP-complete graph problems. *Theoret. Comput. Sci.*, 1(3), 237–267. [https://doi.org/10.1016/0304-3975\(76\)90059-1](https://doi.org/10.1016/0304-3975(76)90059-1)

Changelog

- 02/04/2025: Added note on bounded but not closed domains.
- 02/07/2025: Removed footnote in L1.2, which was meant under the hypothesis of S closed (thanks Charis).