Optimization Theory
Lower semi-continuity, compactness, and existence of solutions

Anton Evgrafov

Department of Mathematical Sciences, NTNU anton.evgrafov@math.ntnu.no

1 Reading

Chapter 1 in Nocedal and Wright, "Numerical optimization."

2 What is optimization?

Optimum (the neuter form of optimus) originates from the Latin, and translates to English as "the best." Therefore, "to optimize something (system/process/acitivity etc)" is normally understood as "to bring something to its best possible state." There are three important terms in this interpretation, which need further clarification:

- "To bring": a modeller needs to identify the parameters of the system (process, activity), which can be varied. These may be discrete or real-valued parameters, or even more general objects such as functions, geometric surfaces, or similar. In this course we will mostly deal with parameters assuming real values. It will be convenient to collect all such parameters in a vector $x \in \mathbb{R}^n$ of *optimization* or *decision variables.*
- "The best": in order to compare the states corresponding to various parameter values we need to introduce a total ordering on the set of parameters. Typically this is done by employing a real-valued *objective* or *cost function* $x \mapsto f(x) \in \mathbb{R}$ (sometimes $x \mapsto f(x) \in \mathbb{R} \cup \{+\infty\}$), with the convention that the "better" values of the parameters correspond to the smaller values of f. Thus to choose the best parameter values we need to find x corresponding to the smallest value of f.
- "Possible": not all combinations or values of the parameters are valid. Limited availability of physical resources (time, money, raw materials, labour, etc) or demand requirements may introduce upper and lower limits on the parameters. There might be technical/logical restrictions on the values or the relationships between various parameters. We will abstractly collect all the admissible values of the parameters for the problem under the consideration into a *feasible set* Ω.
  In most applications, the set Ω is defined as a solution set to a system of inequalities and equalities, which results from the list of all the restrictions on the parameters:
  $$\Omega = \{\, x \in \mathbb{R}^n \mid g_i(x) \geq 0, i \in \mathcal{I},\, g_j(x) = 0, j \in \mathcal{E} \,\}, \tag{1}$$

and the function g_i, $i \in \mathcal{I}$ and g_j, $j \in \mathcal{E}$ will be referred to as the *inequality* and *equality constraints*[1]. Depending on whether the constraints are present, we classify the problem (2) as *constrained* or *unconstrained*.

To summarize, we will be concerned with solving problems of the type

$$\begin{aligned} &\text{minimize } f(x), \\ &\text{subject to } x \in \Omega, \end{aligned} \tag{2}$$

where Ω may be further described with inequality and equality constraints. One may generalize this framework in many ways; for example, instead of parameters in $\mathbb{R}^n$ we may consider other spaces with different topological and/or algebraic structures, such as for example spaces of matrices, functions, curves and surfaces, etc. Instead of inequality (equality) constraints of the type $g_i(x) \geq 0$ ($g_i(x) = 0$) one may instead demand $g(x) \in \mathcal{K}$, where $\mathcal{K}$ is a *cone*[2] (satisfying some technical requirements) in a suitable vector space. We will keep the problem formulation (2) as it provides plenty of the room for modelling, development of the theory, and efficient algorithms. Furthermore, this is the formulation considered in the textbook of the course.

Please note that people often use the expression *mathematical programming* interchangeably with optimization. The program refers to a "decision program" and not a computer program, as optimization/mathematical programming has a much longer history than computer programming.

3 What does it mean to solve the problem (2)?

One distinguishes between two most important types of solutions to (2).

Definition 1 (Global minimum). *A point $x^* \in \Omega$ is called the point of* global minimum, *if for every $x \in \Omega$ we have the inequality $f(x^*) \leq f(x)$.*

Geometrically, x^* is a point of global minimum if the graph $\{ (x, f(x)) \mid x \in \Omega \}$ lies "above" the horizontal plane $\{ (x, f(x^*)) \mid x \in \Omega, \alpha = f(x^*) \}$ *and* touches it at the point $(x^*, f(x^*))$ (but possibly at other points, too).

Points of global minimum may not exist even when the function is bounded from below:

Example 1. Consider a positive function $f(x) = \exp(-x^2)$. This function approaches zero arbitrarily close: for every $\epsilon > 0$ it suffices to take $|x| > [\log(\epsilon^{-1})]^{1/2}$ to get $0 < f(x) < \epsilon$. Therefore, the global minimum, if existed, must satisfy the inequality $f(x^*) < \epsilon$, for any $\epsilon > 0$. However, there is no $x^* \in \mathbb{R}$ such that $f(x^*) = 0$.

---

[1] In this course we will assume that both $\mathcal{I}$ and $\mathcal{E}$ are *finite* index sets.

[2] A cone C in a vector space is a set, which is invariant under multiplication with positive scalars; that is $\lambda C = C$, for every $\lambda > 0$. Examples include the zero cone $\{0\}$; the cone of vectors with non-negative components $\mathbb{R}^n_+$; or the cone of symmetric positive semi-definite matrices $\mathcal{S}^n_+$.

Unless we have information about the global behaviour of the function over the feasible set, global solutions, even when exist, are incredibly difficult to recognize. Indeed, assume that an oracle provides us with a globally optimal solution $x^* \in \Omega$, and our task is to verify her/his guess. Then, in accordance with the definition 1, we should compare $f(x^*)$ with the value $f(x)$, evaluated at *every other point* $x \in \Omega$, which is most often practically impossible. Instead, we will look for points, which can be characterized with the knowledge of the function only in the vicinity of a given point. For differentiable functions such an information will be available from the local Taylor series expansion of f and the constraints.

Definition 2 (Local minimum). *A point $x^* \in \Omega$ is called the point of* local minimum*, if it is a point of global minimum in the feasible set restricted to some neighbourhood of x^*. That is, if there is $\epsilon > 0$ such that for every $x \in \{\, y \in \Omega \mid \|y - x^*\| < \epsilon \,\}$ we have the inequality $f(x^*) \leq f(x)$. If the latter inequality is strict whenever $x \neq x^*$ in the vicinity of x^*, we say that x^* is the point of* strict *local minimum.*

4 Very briefly: "standard tricks" in optimization modelling

4.1 Auxiliary optimization variables

It is often convenient to introduce additional variables, which are not associated with the parameters of the system we are trying to model. One standard type of such auxiliary variables is a slack variable, which allows us to switch from inequality to equality constraints (and simple bounds):

$$g(x) \geq 0 \qquad \iff \qquad g(x) - s = 0,\ s \geq 0.$$

Note that one may, in principle, replace s with s^2 and drop the restriction on the slack variable, but most often this is not such a good idea.

Another type of auxiliary variables appears when we move the objective function f into constraints instead:

$$\begin{cases} \min\limits_{x} f(x), \\ \text{s.t. } x \in \Omega, \end{cases} \qquad \iff \qquad \begin{cases} \min\limits_{(x,z)} z, \\ \text{s.t. } (x,z) \in \{\, (\tilde{x},\tilde{z}) \in \Omega \times \mathbb{R} \mid \tilde{z} - f(\tilde{x}) \geq 0 \,\}, \end{cases}$$

This trick allows one to transform a problem of minimizing a piece-wise smooth objective function $f(x) = \max\{f_1(x), f_2(x), \dots, f_k(x)\}$, where $f_1, \dots, f_k$ are smooth functions, into a problem with smooth objective and constraints:

$$\begin{aligned} &\min_{(x,z)} z, \\ &\text{s.t. } x \in \Omega, \\ &\qquad z - f_1(x) \geq 0, \\ &\qquad \vdots \\ &\qquad z - f_k(x) \geq 0. \end{aligned}$$

Similar tricks allow one to deal with minimizing a variety of non-smooth functions such as $\|\cdot\|_1$ and $\|\cdot\|_\infty$-norms of vectors (provide the details utilizing the fact that $|x| = \max\{x, -x\}$).

4.2 Soft and hard constraints

In some applications, most notably financial, certain constraints may be violated at a cost. Such constraints are typically known as "soft" constrains (as opposed to the "hard" constraints, which must be satisfied no matter what). We can turn a "hard" inequality constraint $g(x) \geq 0$ into a soft constraint as follows. First, we introduce an artificial variable $s \geq 0$, which will measure how much the constraint g is violated, that is, we consider the constraints $g(x) + s \geq 0$, $s \geq 0$ instead. Second, we need to add the cost of violation, say $h(s)$[3], to the objective function. That is, instead of $f(x)$ we minimize $f(x) + h(s)$.

The idea of soft constraints is also utilized in penalty methods for constrained optimization, allowing one to transform the constrained problem into an unconstrained one, or the one with very simple constraints.

5 Very briefly: classification

- *Unconstrained optimization* refers to the situations when $\Omega = \mathbb{R}^n$ in (2); *constrained optimization* otherwise.
- *Linear programming/optimization*: the objective function and all the constraints are first order polynomials; *non-linear optimization* otherwise.
- *Quadratic programming*: the objective function is a second order polynomial and all the constraints are first order polynomials.
- *Convex programming/optimization*: the objective function and the feasible set Ω is convex; if the constraints are given explicitly, then all the inequality constraints are concave functions and the equality constraints are affine (first order polynomials).
- *Non-smooth/non-differentiable optimization*: normally refers to the situation, when the objective function $f(x)$ (or some of the constraints, though problems in this class are often unconstrained or involve only simple constraints, such as bounds on the variables) is not differentiable at least at some points. If all the functions involved in the problem are at least once differentiable, we deal with *differentiable* (sometimes *smooth*) *optimization.*
- *Semi-infinite programming*: the number of decision variables is finite, but the number of constraints is infinite.
- *Semi-definite programming*: optimization over spaces of symmetric matrices, restricted to be positive semi-definite.
- *Calculus of variations*: optimization over spaces of functions.

[3] A typical example of $h(s)$ is Ms, where $M > 0$ is the cost of violating the constraint $g(x) \geq 0$ "per unit of violation."

6 Basic existence of solutions results

One of the weakest forms of continuity under which one may expect the problem (2) to admit solutions is as follows.

Definition 3 (Lower semi-continuity). *Consider a function $f : \Omega \to \mathbb{R} \cup \{+\infty\}$ on a set $\Omega \subseteq \mathbb{R}^n$. We say that the function f is lower semi-continuous (l.s.c.) on Ω, if the lower level set $\Omega_\alpha = \{\, x \in \Omega \mid f(x) \leq \alpha \,\}$ is relatively closed*[4] *in Ω for any $\alpha \in \mathbb{R}$.*

Example 2. Lower semi-continuous functions appear naturally in the context of so-called min-max (inf-sup) problems, where we try to find a minimum of a function, which is defined through a maximization problem. For example, let

$$f(x) = \sup_{t \in \mathbb{R}} \{1 - \exp(-x^2 t^2)\}.$$

Then

$$f(x) = \begin{cases} 0, & \text{if } x = 0, \\ 1, & \text{otherwise,} \end{cases}$$

which is clearly discontinuous at $x = 0$, despite the fact that each individual function $1 - \exp(-x^2 t^2)$ is continuous in (x, t). Nevertheless, the function f is still l.s.c. on $\mathbb{R}$, because for any $\alpha \in \mathbb{R}$ we have

$$\Omega_\alpha = \begin{cases} \emptyset, & \alpha < 0, \\ \{0\}, & 0 \leq \alpha < 1, \\ \mathbb{R}, & 1 \leq \alpha, \end{cases}$$

all of which are closed sets in $\mathbb{R}$.

We now define the concept of a minimizing sequence. One can establish the existence of solutions to the problem (2) without appealing to minimizing sequences (as indeed we will). Nevertheless, minimizing sequences constitute an important and often utilized concept in their own right.

Definition 4 (Minimizing sequence). *Assume that $\Omega \neq \emptyset$ and let $f^* = \inf_{x \in \Omega} f(x)$.*[5] *By the definition of the infimum, there is a sequence of numbers $\{f_k\}_{k=1}^\infty$ in the set $\{\, f(x) \mid x \in \Omega \,\}$, such that $\lim_{k \to \infty} f_k = f^*$. By the definition of f_k, there is $x_k \in \Omega$ such that $f_k = f(x_k)$. We say that the sequence $\{x_k\}_{k=1}^\infty$ is a* minimizing sequence *for the problem* (2).

[4] Relative closedness in this notation means that if a sequence of points $\{x_k\}_{k=1}^\infty$ in Ω_α converges to a limit $x^* \in \Omega$, then $x^* \in \Omega_\alpha$. In other words, the set Ω_α contains all its limit points, which are also in Ω. Therefore closedness, relative to the whole space $\mathbb{R}^n$, coincides with the regular definition of closedness.

[5] We adopt the convention that every subset of the real line has an infimum, or the greatest lower bound, by letting the infimum of the set unbounded from below be equal to $-\infty$ and the infimum of the empty set be equal $+\infty$. We apply a similar convention to sup.

We know that the function is continuous iff the convergence of sequences is preserved by the function, that is, $x_k \to x^* \implies f(x_k) \to f(x^*)$. One can provide a similar characterization of lower semi-continuity as well, which will be useful for our purposes when applied to minimizing sequences.

Proposition 1. *A function $f : \Omega \to \mathbb{R} \cup \{+\infty\}$ is l.s.c. on Ω iff for every sequence $\{x_k\}_{k=1}^\infty$ in Ω converging to $x^* \in \Omega$ it holds that $f(x^*) \leq \liminf_{k\to\infty} f(x_k)$*[6]

Proof. Suppose that f is l.s.c. on Ω. For the sake of contradiction assume that for some convergent sequence $\{x_k\}_{k=1}^\infty$ in Ω it holds that $\liminf_{k\to\infty} f(x_k) < f(x^*)$, where $\Omega \ni x^* = \lim_{k\to\infty} x_k$. In particular, we get that $\liminf_{k\to\infty} f(x_k) < +\infty$ and as a result there is $\alpha \in \mathbb{R}$, such that $\liminf_{k\to\infty} f(x_k) < \alpha < f(x^*)$. We will demonstrate that there there is a subsequence of x_k, which belongs to Ω_α, thus showing that Ω_α is not relatively closed (it does not contain one of its limit points, namely x^*); this is our desired contradiction with the lower semi-continuity of f. We proceed with yet another proof by contradiction. Indeed, suppose that Ω_α contains only finitely many elements x_k, that is, there is an index N such that for all $k \geq N$ it holds that $x_k \notin \Omega_\alpha$, or equivalently, $f(x_k) > \alpha$. Then $\inf_{k\geq n} f(x_k) \geq \alpha$ for every $n \geq N$ and as a result also $\lim_{n\to\infty} \inf_{k\geq n} f(x_k) \geq \alpha$. This contradicts with our choice of α: $\liminf_{k\to\infty} f(x_k) < \alpha < f(x^*)$.

We now prove the implication in the opposite direction. Suppose that f is *not* l.s.c. on Ω. Then there is $\alpha \in \mathbb{R}$ such that Ω_α is not relatively closed in Ω. That is, there is a sequence $\{x_k\}_{k=1}^\infty$ in Ω_α, with a limit $x^* \in \Omega \setminus \Omega_\alpha$. Clearly, we then have $\liminf_{k\to\infty} f(x_k) \leq \liminf_{k\to\infty} \alpha = \alpha < f(x^*)$, which concludes the proof. $\square$

We now establish existence of solutions to the problem (2) without requiring any algebraic/geometric properties of the problem (such as convexity). The result is normally attributed to Weierstrass.

Theorem 1 (Existence of solutions). *Let Ω be a non-empty compact*[7] *set in $\mathbb{R}^n$ (or any other metric space, for that matter) and $f : \Omega \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous on Ω. Then the problem* (2) *admits at least one global minimum.*

Proof. Let $\{x_k\}_{k=1}^\infty$ be a minimizing sequence for the problem (2) (see Definition 4). Owing to the compactness of Ω it holds that for $\{x_k\}_{k=1}^\infty$ contains a subsequence, say $\{x'_k\}_{k=1}^\infty$, converging to some point $x^* \in \Omega$. Utilizing the definition of the infimum, Proposition 1, the fact that $\{f(x'_k)\}_{k=1}^\infty$ is a subsequence of the converging sequence $\{f(x_k)\}_{k=1}^\infty$, and the definition of the minimizing sequence we obtain the following string of inequalities:

$$\inf_{x\in\Omega} f(x) \leq f(x^*) \leq \liminf_{k\to\infty} f(x'_k) = \lim_{k\to\infty} f(x'_k) = \lim_{k\to\infty} f(x_k) = \inf_{x\in\Omega} f(x),$$

[6] Recall that for any sequence $\{\alpha_k\}_{k=1}^\infty$ of real numbers, $\liminf_{k\to\infty} \alpha_k = \lim_{n\to\infty} \inf_{k\geq n} \alpha_k$. $\liminf$ (finite or infinite) exists for an arbitrary sequence, as the sequence $\beta_n = \inf_{k\geq n} \alpha_k$ is monotonically non-decreasing.

[7] Recall that a subset of $\mathbb{R}^n$ is compact iff it is closed and bounded (Heine–Borel Theorem). Further, from any sequence in a compact subset of a metric space we can extract a converging subsequence, which is utilized in the proof of Theorem 1.

which implies that $x^* \in \Omega$ is the point of global minimum for (2). $\square$

We now give an alternative proof of Theorem 1, which does not appeal to the convergence of sequences.

Proof (Alternative proof of Theorem 1). Let $f^* = \inf_{x\in\Omega} f(x)$, and let Ω_α denote the lower-level set of f for any real number α. Every Ω_α is a closed set owing to the lower semi-continuity of f, which is non-empty for any $\alpha > f^*$ by the definition of inf. Therefore, for any finitely many numbers $\alpha_1 > f^*$, $\alpha_2 > f^*$,…,$\alpha_N > f^*$ it holds that

$$\cap_{i=1}^N \Omega_{\alpha_i} = \Omega_{\min_{i=1,\dots,N} \alpha_i} \neq \emptyset,$$

since $\min_{i=1,\dots,N} \alpha_i > f^*$. This is precisely the condition for the family of closed sets $\{\, \Omega_\alpha \mid \alpha > f^* \,\}$ to have a *finite intersection property.* Owing to the compactness of Ω it holds that $\cap_{\alpha>f^*} \Omega_\alpha \neq \emptyset$. By construction, every point in $\cap_{\alpha>f^*} \Omega_\alpha = \Omega_{f^*}$ is a point of global minimum for (2); in particular, $f^* > -\infty$ as f cannot assume the value $-\infty$. $\square$

When establishing the existence of solutions to optimization problems, one can trade the compactness of Ω for the growth of f at infinity, which guarantees that the minimizing sequences stay bounded.

Definition 5. *A function $f : \Omega \to \mathbb{R} \cup \{+\infty\}$ is called* coercive *if $f(x) \to +\infty$ whenever $\|x\| \to +\infty$.*

Theorem 2. *Let Ω be a non-empty closed set in $\mathbb{R}^n$ and $f : \Omega \to \mathbb{R}$ be a coercive lower semi-continuous function on Ω. Then the problem* (2) *admits at least one global minimum.*

Exercises

1. Prove Theorem 2.
2. Show that all the assumptions in Theorems 1 and 2 are essential. Indeed, Example 1 shows that either compactness or coercivity are essential for the existence of solutions. Demonstrate that the lower semi-continuity is also needed by constructing an instance of the problem (2) with a not lower semi-continuous objective function f, which does not attain its infimum on some non-empty compact feasible set Ω.
3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic polynomial $f(x) = 0.5x^T Gx + x^T d$, where $G \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, and $d \in \mathbb{R}^n$ is arbitrary. Show that f is coercive on $\mathbb{R}^n$. (Hint: expand x in terms of the eigenvectors of G to estimate $x^T Gx$ from below.)
4. Provide details on how one can transform the non-smooth optimization problem of minimizing the 1- or ∞-norm of a vector to a smooth minimization problem by introducing auxiliary variables and additional constraints.

5. * Example 2 is not incidental! Indeed, consider a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined as $f(x) = \sup\{\, f_s(x) \mid s \in S \,\}$, where each function $f_s : \mathbb{R}^n \to \mathbb{R}$ is l.s.c., and S is an arbitrary index set (possibly uncountable). Show that f is l.s.c. on $\mathbb{R}^n$.
6. Show that the function $f : \Omega \to \mathbb{R}$ is l.s.c. on Ω if and only if it satisfies the following "ϵ/δ" definition of lower semi-continuity: for any x^* in Ω and any $\epsilon > 0$ there is $\delta > 0$ such that for any $x \in \Omega$, $\|x - x^*\| < \delta \implies f(x^*) < f(x) + \epsilon$. (You may use the equivalent characterization of lower semi-continuity given by Proposition 1).