THEORY OF UNCONSTRAINED OPTIMISATION

MARKUS GRASMAIR

In this note, we will discuss the basic theory required for the solution of finite dimensional, continuous optimisation problems. In the first section, we will discuss the basic definitions and show conditions that guarantee existence of solutions to optimisation problems, both in the constrained and unconstrained case. Then we restrict ourselves to unconstrained problems, where we derive different necessary and sufficient optimality conditions. Here, the constrained case is significantly more difficult and will be discussed later in the lecture. Finally, we consider in particular convex functions and their properties as relates to optimisation. It turns out that convex functions are in a sense among the most simple functions to minimise, both in a theoretical setting (as discussed here) and in practice with numerical algorithms (as we will show throughout the course).

1. Solutions of Optimisation Problems

We will start with discussing the existence of solutions of minimisation problems of the form

$$\min_{x\in\Omega} f(x),$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a real valued function (the *cost function* or *objective function*) and $\Omega \subset \mathbb{R}^d$ is some set (the *feasible set*).

1.1. Notions of Minimisers.

First we have to clarify what we mean by a solution of an optimisation problem.

Definition 1.1 (Global minimiser)**.** A point $x^* \in \Omega$ is called a *global minimiser* (or *global minimum*, or *global solution*) of the optimisation problem $\min_{x\in\Omega} f(x)$, if

$$f(x^*) \le f(x)$$

for all $x \in \Omega$.

The point x^* is *strict global minimiser*, if $f(x^*) < f(x)$ for all $x \in \Omega$, $x \neq x^*$.

Example 1.2. The point $x = -\pi/2$ is a global minimiser of the function $f(x) = \sin(x)$, since $\sin(-\pi/2) = -1$ and $\sin(x) \ge -1$ for every $x \in \mathbb{R}$. However, we also have that $\sin(3\pi/2) = -1 = \sin(-\pi/2)$. Thus $x = -\pi/2$ is a global minimiser, but not a strict global minimiser of f.

One problem of global minimisers is that they are incredibly hard to recognise in general. In order to verify that a point x^* is a global minimiser, one would have to compare $f(x^*)$ with every other value $f(x)$, no matter how large the distance between x and x^* is. In actual applications, however, one usually may only obtain the value of f (and, possibly, some of its derivatives) at a small number of selected points. With only this information available, only in very special cases is it possible to prove that a given point x^* is really a global minimiser.

As an alternative, we therefore consider local minimisers:

Date: 29th February 2024.

Definition 1.3 (Local minimiser)**.** A point $x^* \in \Omega$ is called a *local minimiser* (or *local minimum*, or *local solution*) of the optimisation problem $\min_{x\in\Omega} f(x)$, if there exists $\varepsilon > 0$ such that $f(x^*) \leq f(x)$ whenever $x \in \Omega$ satisfies $\|x - x^*\| \leq \varepsilon$.

Slightly strengthening this notion, we obtain:

Definition 1.4 (Strict local minimiser)**.** A point x^* is called a *strict local minimiser* of $\min_{x\in\Omega} f(x)$, if there exists $\varepsilon > 0$ such that

$$f(x^*) < f(x)$$

whenever $x \in \Omega$, $x \neq x^*$ satisfies $\|x - x^*\| \leq \varepsilon$.

That is, we replace the inequality $\leq$ by the strict inequality $<$ in the definition of the local minimiser.

In addition, it makes sometimes sense to strengthen this notion further:

Definition 1.5 (Isolated local minimiser)**.** A point $x^* \in \Omega$ is called an *isolated local minimiser* of the problem $\min_{x\in\Omega} f(x)$, if there exists $\varepsilon > 0$ such that x^* is the only local minimiser of f in Ω an ε-ball around x^*. That is, if $y^* \neq x^*$ is another local minimiser of f in Ω, then $\|x^* - y^*\| > \varepsilon$.

Example 1.6. The point $x = -\pi/2$ is a local minimiser of the function $f(x) = \sin(x)$. In fact, it is an isolated local minimiser, because it is the only local minimiser of f within the open interval $(-5\pi/2, 3\pi/2)$.

Example 1.7. Define $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = (x^2 - y)^2$. Then $f(x,y) \geq 0$ for all $(x,y) \in \mathbb{R}^2$, and $f(x,y) = 0$ if $x^2 = y$. Thus all points of the form (x, x^2) for $x \in \mathbb{R}$ are local and global minimisers of f, but none of them is an isolated minimiser.

Example 1.8. Assume that $A \in \mathbb{R}^{d\times d}$ is a symmetric matrix[1] and define the function $f\colon \mathbb{R}^d \to \mathbb{R}$ by

$$f(x) = \langle x, Ax\rangle = x^T Ax.$$

Then there are three possibilities:

- If the matrix A is positive definite, then $x^* = 0$ is the unique strict global and local minimiser of f.
- If the matrix A is positive semi-definite, but not positive definite, then $x^* = 0$ is a global minimiser, but it is not strict and not isolated. In fact, if $v \in \mathbb{R}^d$ is any eigenvector for the eigenvalue 0 of A, then $f(v) = 0$ and v is another global minimiser.
- If the matrix A is not positive semi-definite (that is, either indefinite or negative definite), then the function f does not admit global or local minimisers.

Example 1.9. From the definition, it is easy to see that every isolated local minimiser is necessarily a strict local minimiser. The converse, however, does not necessarily hold as seen by the (rather pathological) function

$$f(x) = \begin{cases} 2x^2 + x^2\sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

This function has a strict local minimiser at $x = 0$ (which is at the same time the unique global minimiser of f), but there exists a sequence of (isolated!) local minimisers converging to 0. Thus the minimiser at 0 is not isolated. See also Figure 1.

[1]For more information on symmetric matrices, see Appendix A.

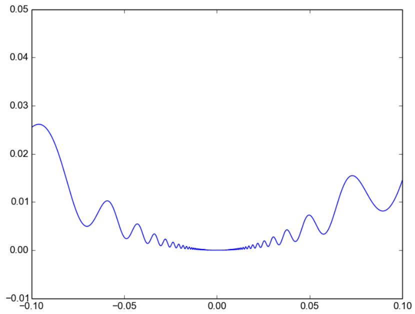

FIGURE 1. A close-up view of the function $f(x) = 2x^2 + x^2\sin(1/x)$ near 0. The point $x = 0$ is the unique global minimum, but is also an accumulation point of isolated local minima.

1.2. **Existence of Minimisers.** In general, an optimisation problem need not attain local or global solutions. Consider for instance the following optimisation problems:

- The problem $\min_{x\in\mathbb{R}} f(x)$ with $f(x) = e^{-x^2}$.
  The function f satisfies $f(x) > 0$ for all $x \in \mathbb{R}$, but $\lim_{x\to\pm\infty} f(x) = 0$. That is, the function values can be arbitrarily close to 0, but there is no point $x \in \mathbb{R}$ at which the value 0 is actually attained.
- The problem $\min_{x>0} x$.
  Again, the objective function is strictly positive on the feasible set $\Omega = \mathbb{R}_{>0}$, but the values can be arbitrarily close to 0.
- The problem $\min_{x\in\mathbb{R}} f(x)$ with $f(x) = x^2$ for $x \neq 0$ and $f(0) = 1$.
  Again, the function attains only positive values that can be arbitrarily close to 0.

It turns out, however, that these three examples are in a sense the typical counter-examples to the existence of minimisers: In the case of the function e^{-x^2}, the problem is that the function to be minimised becomes smaller as the argument increases; in the case of the second counter-example, we are minimising over a feasible set that is not close; in the last example, the problem is a discontinuity at the point where we would "naturally" expect the minimum. By excluding these two possibilities, that is, by requiring the function f to be continuous and to grow as its argument tends to infinity, we can indeed guarantee the existence of a minimiser. Because discontinuous functions can be important in some applications, it makes sense to try to obtain results for this type of functions as well, though.

As a preparation, we introduce the notion of level sets of functions, which play an important role in optimisation.

Definition 1.10 (Level sets)**.** Assume that $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is a function and that $\alpha \in \bar{\mathbb{R}}$. By

$$L_f(\alpha) := \{x \in \mathbb{R}^d : f(x) \leq \alpha\}$$

we denote the *(lower) level set of f* for the level α.

Lemma 1.11. *Assume that $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is a function and that $\alpha < \beta \in \bar{\mathbb{R}}$. Then $L_f(\alpha) \subset L_f(\beta)$. Moreover, $L_f(+\infty) = \mathbb{R}^d$.*

Proof. Exercise! ☐

Lemma 1.12. *Assume that f, $g\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ are functions and that $f(x) \ge g(x)$ for all $x \in \mathbb{R}^d$. Then*
$$L_f(\alpha) \subset L_g(\alpha)$$
for all $\alpha \in \bar{\mathbb{R}}$.

Proof. Exercise! □

Lemma 1.13. *Assume that f, $g\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ are functions and define $h\colon \mathbb{R}^d \to \bar{\mathbb{R}}$,*
$$h(x) = \max\bigl\{f(x), g(x)\bigr\}.$$
Then
$$L_h(\alpha) = L_f(\alpha) \cap L_g(\alpha)$$
for all $\alpha \in \bar{\mathbb{R}}$.

Proof. Exercise! □

The previous result can easily be extended to arbitrary collections of functions:

Lemma 1.14. *Assume that $f_i\colon \mathbb{R}^d \to \bar{\mathbb{R}}$, $i \in I$, is any family of functions, and define $h\colon \mathbb{R}^d \to \bar{\mathbb{R}}$,*
$$h(x) := \sup_{i \in I} f_i(x).$$
Then
$$L_h(\alpha) = \bigcap_{i \in I} L_{f_i}(\alpha)$$
for all $\alpha \in \bar{\mathbb{R}}$.

Proof. Exercise! □

The following result shows the connection between strict local and global solutions of an optimisation problem, and properties of the level sets of f.

Lemma 1.15. *Assume that $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is a function and that $x^* \in \mathbb{R}^d$. Denote $\alpha^* := f(x^*)$. Then x^* is a strict global solution of the problem* $\min_{x\in\mathbb{R}^d} f(x)$*, if and only if $L_f(\alpha^*) = \{x^*\}$.*

Similarly, x^* is a strict local solution of the problem $\min_{x\in\mathbb{R}^d} f(x)$*, if and only if there exists $\varepsilon > 0$ such that $L_f(\alpha^*) \cap B_\varepsilon(x^*) = \{x^*\}$.*

Proof. This is an immediate consequence of the definitions of level sets and strict minima: The point x^* is a strict global minimum, if and only if $f(x) > f(x^*) = \alpha^*$ for all $x \in \mathbb{R}^d$, $x \neq x^*$. This is equivalent to the statement that $L_f(\alpha^*) = \{x^*\}$.

For local minima, the argumentation is similar. □

We now introduce the main notions that allow us to prove existence of solutions of optimisation problems.

Definition 1.16 (Coercivity)**.** A function $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is *coercive*, if for all $\alpha \in \mathbb{R}$ the level set $L_f(\alpha)$ is bounded.

Recall here that a set $S \subset \mathbb{R}^d$ is bounded, if there exists $r > 0$ such that $S \subset B_r(0)$. In particular, the empty set $S = \emptyset$ is by definition bounded.

Remark 1.17. Since $L_f(+\infty) = \mathbb{R}^d$ for all functions f, we only consider the level sets $L_f(\alpha)$ for *finite* $\alpha \in \mathbb{R}$ in the definition of coercivity.

Remark 1.18. Alternatively, we can say that a function $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is coercive, if we have for every sequence $\{x_k\}_{k\in\mathbb{N}}$ with $\|x_k\| \to \infty$ that $f(x_k) \to \infty$.

Example 1.19. The function $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2 + y^2$ is coercive: For $\alpha \ge 0$ we have that
$$L_f(\alpha) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le \alpha\} = \overline{B_{\sqrt{\alpha}}}(0),$$
and for $\alpha < 0$ we have that $L_f(\alpha) = \emptyset$.

Example 1.20. The function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$ is *not* coercive: For instance, the level set $L_f(1)$ of f is the negative half-line $L_f(1) = \mathbb{R}_{\le 0}$ which is unbounded.

Lemma 1.21. *Assume that f, $g\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ are functions and that $f(x) \ge g(x)$ for all $x \in \mathbb{R}^d$. Assume moreover that g is coercive. Then f is coercive as well.*

Proof. Let $\alpha \in \mathbb{R}$. Since $f \ge g$, it follows that $L_f(\alpha) \subset L_g(\alpha)$. Moreover, the coercivity of g implies that $L_g(\alpha)$ is bounded. Thus $L_f(\alpha)$ is bounded as well, being the subset of the bounded set $L_g(\alpha)$. □

Example 1.22. Assume that $A \in \mathbb{R}^{d\times d}$ is symmetric and define $f\colon \mathbb{R}^d \to \mathbb{R}$, $f(x) = \langle x, Ax\rangle$. Then f is coercive, if and only if A is positive definite.

In order to see this, we assume first that A is not positive definite. Then A has an eigenvalue $\lambda \le 0$ with corresponding eigenvalue $v \in \mathbb{R}^d \setminus \{0\}$. For all $c \in \mathbb{R}$ we then have that $f(cv) = \langle cv, A(cv)\rangle = c^2\lambda\|v\|^2 \le 0$, which shows that $\mathbb{R}v \subset L_f(0)$. Thus $L_f(0)$ is unbounded, and therefore f is not coercive.

Conversely, assume that A is positive definite. Denote by $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_d > 0$ the ordered eigenvalues of A and let $v_1, \ldots, v_d$ be a corresponding orthonormal eigenbasis of A. Then
$$f(x) = \langle x, Ax\rangle = \sum_{i=1}^d \lambda_i \langle x, v_i\rangle^2 \ge \lambda_d \sum_{i=1}^d \langle x, v_i\rangle^2 = \lambda_d\|x\|^2 := g(x)$$
for all $x \in \mathbb{R}^d$. Thus we can bound f from below by the coercive function $g(x) = \lambda_d\|x\|^2$, which shows that f is coercive as well.

The next notion we need is a generalisation of continuity. For the following definition recall that a subset $K \subset \mathbb{R}^d$ is closed if all convergent sequences in K have their limit in K. That is, if $\{x_k\}_{k\in\mathbb{N}}$ is a sequence such that $x_k \in K$ for all $k \in \mathbb{N}$ and that $x := \lim_{k\to\infty} x_k \in \mathbb{R}^d$ exists, then necessarily $x \in K$.

Definition 1.23 (Lower semi-continuity)**.** A function $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is *lower semi-continuous* (lsc), if for all $\alpha \in \mathbb{R}$ the level set $L_f(\alpha)$ of f is closed.

Example 1.24. The function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ for $x \ne 0$ and $f(0) = 1$ is *not* lower semi-continuous: For instance, we have that
$$L_f(1/4) = \{x \ne 0 : x^2 \le 1/4\} = [-1/2, 0) \cup (0, 1/2],$$
which is not closed.

On the other hand, the function $g\colon \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$ for $x \ne 0$ and $g(0) = -1$ is lower semi-continuous: For $\alpha \ge 0$ we have that $L_g(\alpha) = [-\sqrt{\alpha}, +\sqrt{\alpha}]$, which is a closed interval. For $-1 \le \alpha \le 0$ we have that $L_g(\alpha) = \{0\}$, which is closed. Finally, for $\alpha < -1$ we have that $L_g(\alpha) = \emptyset$, which is closed as well.

Lemma 1.25. *Every continuous function $f\colon \mathbb{R}^d \to \mathbb{R}$ is also lower semi-continuous.*

Proof. Let $\alpha \in \mathbb{R}$ and assume that $\{x_k\}_{k\in\mathbb{N}}$ is a sequence such that $x_k \in L_f(\alpha)$ for all $k \in \mathbb{N}$ and that $x := \lim_{k\to\infty} x_k \in \mathbb{R}^d$ exists. We have to show that $x \in L_f(\alpha)$ as well, that is, that $f(x) \le \alpha$. Since $x_k \in L_f(\alpha)$ for all k, it follows that $f(x_k) \le \alpha$ for all k. Thus the continuity of f implies that
$$f(x) = f\Bigl(\lim_{k\to\infty} x_k\Bigr) = \lim_{k\to\infty} f(x_k) \le \alpha,$$
which proves the claim. □

Lower semi-continuous functions play a prominent role in optimisation: On the one hand, they are amenable to minimisation, as the existence results in the following section show. On the other hand, they appear naturally in certain branches of optimisation in the form of min-max (or inf-sup) problems, that is, problems of the form

$$\inf_{x\in\Omega}\sup_{y\in W} g(x,y).$$

The following result shows that the function $h(x) := \sup_{y\in W} g(x,y)$ is lower semi-continuous (but not necessarily continuous), provided the function g is continuous in x for each y.

Lemma 1.26. *If $f_i \colon \mathbb{R}^d \to \bar{\mathbb{R}}$, $i \in I$, is* any *family of lower semi-continuous functions, then the function $h\colon \mathbb{R}^d \to \bar{\mathbb{R}}$,*

$$h(x) := \sup_{i\in I} f_i(x)$$

is lower semi-continuous as well.

Proof. Let $\alpha \in \mathbb{R}$. Since f_i is lower semi-continuous, it follows that $L_{f_i}(\alpha)$ is closed for all $i \in I$. Thus

$$L_h(\alpha) = \bigcap_{i\in I} L_{f_i}(\alpha)$$

is an intersection of closed sets and thus closed itself. □

Lemma 1.27. *Assume that $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is a function. The following are equivalent:*

(1) f is lower semi-continuous.

(2) For all $x \in \mathbb{R}^d$ and all sequences $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^d$ such that $x = \lim_{k\to\infty} x_k$ and $\lim_{k\to\infty} f(x_k) \in \bar{\mathbb{R}}$ exists we have that

$$f(x) \le \lim_{k\to\infty} f(x_k). \tag{1}$$

Proof. Assume first that f is lower semi-continuous. Let moreover $x \in \mathbb{R}^d$, and let $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^d$ be a sequence such that $x = \lim_{k\to\infty} x_k$ and $\lim_{k\to\infty} f(x_k) \in \bar{\mathbb{R}}$ exists. We have to show that $f(x) \le \lim_{k\to\infty} f(x_k)$. Let therefore $\alpha > \lim_{k\to\infty} f(x_k)$ be arbitrary. Then $f(x_k) < \alpha$ for all sufficiently large k, and thus $x_k \in L_f(\alpha)$ for all sufficiently large k. Now the lower semi-continuity of f implies that $L_f(\alpha)$ is a closed set. Since the sequence $\{x_k\}_{k\in\mathbb{N}}$ is eventually contained in $L_f(\alpha)$, it follows therefore that $x = \lim_{k\to\infty} x_k$ is contained in $L_f(\alpha)$ as well. Thus $f(x) \le \alpha$. Since this holds for every $\alpha > \lim_{k\to\infty} f(x_k)$, we obtain that $f(x) \le \lim_{k\to\infty} f(x_k)$ as desired.

Now assume that (2) holds. Let moreover $\alpha \in \mathbb{R}$. We have to show that $L_f(\alpha)$ is closed. Let therefore $\{x_k\}_{k\in\mathbb{N}}$ be some sequence in $L_f(\alpha)$ such that $x := \lim_{k\to\infty} x_k \in \mathbb{R}^d$ exists. We have to show that $x \in L_f(\alpha)$, that is, that $f(x) \le \alpha$. By construction, we have that $x_k \in L_f(\alpha)$ for all $k \in \mathbb{N}$, and thus $f(x_k) \le \alpha$ for all $k \in \mathbb{N}$. We now have two possibilities:

- The sequence $\{f(x_k)\}_{k\in\mathbb{N}}$ is bounded. Then there exists a subsequence $f(x_{k'})$ and some $y \in \mathbb{R}$ such that $\lim_{k'} f(x_{k'}) = y$. Thus the sub-sequence $\{x_{k'}\}_{k'}$ satisfies the assumption of (2), and thus $f(x) \le \lim_{k'} f(x_{k'})$. Since $f(x_{k'}) \le \alpha$ for all k', it follows that also $f(x) \le \alpha$ and thus $x \in L_f(\alpha)$.
- The sequence $\{f(x_k)\}_{k\in\mathbb{N}}$ is unbounded below. Then there exists a subsequence $f(x_{k'})$ such that $\lim_{k'} f(x_{k'}) = -\infty$. Again we can then use the assumption (2) and conclude that $f(x) \le \lim_{k'} f(x_{k'}) = -\infty < \alpha$. Thus, again, $x \in L_f(\alpha)$.

In both situations we have arrived at the conclusion that $x \in L_f(\alpha)$. Since the sequence $\{x_k\}_{k\in\mathbb{N}}$ was arbitrary, it follows that $L_f(\alpha)$ is closed, which proves the assertion. □

In view of the previous result, we can say that, in a sense, the difference between continuity and lower semi-continuity is that the equal-sign in the definition of continuity is replaced by a lower-than-or-equal sign. Be careful with this characterisation, though, as it is not sufficient that (1) holds for sequences $(x_k)_{k\in\mathbb{N}}$ where $f(x_k)$ converge to a *finite* limit, but also for those where $f(x_k)$ converges to an *infinite* limit value (specifically to $-\infty$). As an example, the function $f(x) = 1/x$ for $x \neq 0$ with $f(0) = 0$ is *not* lower semi-continuous.

Lemma 1.28. *Assume that $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is lower semi-continuous and that $\lambda > 0$. Then the function $g\colon \mathbb{R}^d \to \bar{\mathbb{R}}$, $g(x) := \lambda f(x)$ is lower semi-continuous as well.*

Proof. This immediately follows from the relation

$$L_{\lambda f}(\alpha) = \bigl\{x \in \mathbb{R}^d : \lambda f(x) \le \alpha\bigr\} = \bigl\{x \in \mathbb{R}^d : f(x) \le \alpha/\lambda\bigr\} = L_f(\alpha/\lambda)$$

and the fact that the level sets of f are closed. □

Remark 1.29. Assume that $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is lower semi-continuous and that $\lambda < 0$. Then it is not necessarily true that λf is lower semi-continuous as well. A simple counterexample here is the function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = -1$ if $x \le 0$ and $f(x) = +1$ is $x > 0$. This function is lower semi-continuous, as its level sets are

$$L_f(\alpha) = \begin{cases} \emptyset & \text{if } \alpha < -1, \\ (-\infty, 0] & \text{if } -1 \le \alpha < 1, \\ \mathbb{R} & \text{if } \alpha \ge 1. \end{cases}$$

However, the level sets of $-f$ are

$$L_{-f}(\alpha) = \begin{cases} \emptyset & \text{if } \alpha < -1, \\ (0, +\infty) & \text{if } -1 \le \alpha < 1, \\ \mathbb{R} & \text{if } \alpha \ge 1. \end{cases}$$

In particular, $L_{-f}(0)$ is not closed, and thus $-f$ is not lower semi-continuous.

Lemma 1.30. *Assume that f, $g\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ are lower semi-continuous and define $h\colon \mathbb{R}^d \to \bar{\mathbb{R}}$, $h(x) = f(x) + g(x)$. Then h is lower semi-continuous as well.*

Proof. Let $x \in \mathbb{R}^d$, and let $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^d$ be a sequence with $x = \lim_{k\to\infty} x_k$ such that $v := \lim_{k\to\infty} h(x_k)$ exists in $\bar{\mathbb{R}}$. Then there exists a subsequence $\{x_{k'}\}_{k'}$ such that both $u := \lim_{k'} f(x_{k'})$ and $w := \lim_{k'} g(x_{k'})$ exist in $\bar{\mathbb{R}}$. Moreover, by the lower semi-continuity of f and g we have that $f(x) \le u$ and $g(x) \le w$, which in particular implies that $u, w > -\infty$. Thus $u + w$ is well-defined and we have that

$$\begin{aligned} v = \lim_{k'} h(x_{k'}) = \lim_{k'}(f(x_{k'}) + g(x_{k'})) \\ = \lim_{k'} f(x_{k'}) + \lim_{k'} g(x_{k'}) = u + w \ge f(x) + g(x) \ge h(x). \end{aligned}$$

This proves the lower semi-continuity of h. □

Remark 1.31. In Lemma 1.30 we have to restrict ourselves to functions that do not take the value $-\infty$ so that their sum is well-defined. If we allow for f and g to be arbitrary extended real valued functions, then there might exist some $x \in \mathbb{R}^d$ such that, for instance, $f(x) = -\infty$ and $g(x) = +\infty$, and thus the sum $f(x) + g(x)$ has no well-defined meaning.

We now discuss the existence of solutions of optimisation problems. The main additional ingredient for proving existence is the following result from multivariate calculus:

Theorem 1.32 (Heine–Borel). *Let $K \subset \mathbb{R}^d$ be bounded and assume that $\{x_k\}_{k\in\mathbb{N}} \subset K$ is a sequence in K. Then there exist $x^* \in \mathbb{R}^d$ and a sub-sequence $\{x_{k'}\}_{k'}$ such that $x^* = \lim_{k'} x_{k'}$.*

Theorem 1.33. *Assume that $f \colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is coercive and lower semi-continuous. Then the optimisation problem $\min_{x\in\mathbb{R}^d} f(x)$ has a global solution.*

Proof. Without loss of generality, assume that $f(x) < +\infty$ for some $x \in \mathbb{R}^d$. Else $f(x) = +\infty$ for all $x \in \mathbb{R}^d$ and the claim is trivial.

Denote $f^* := \inf_{x\in\mathbb{R}^d} f(x) < +\infty$. That is, $f^* \le f(x)$ for all $x \in \mathbb{R}^d$ and there exists a sequence $\{x_k\}_{k\in\mathbb{N}}$ such that $f^* = \lim_{k\to\infty} f(x_k)$. Since $\lim_{k\to\infty} f(x_k) = f^* < +\infty$, it follows that there exists $\alpha \in \mathbb{R}$ such that $f(x_k) \le \alpha$ for all $k \in \mathbb{N}$. In other words, $x_k \in L_f(\alpha)$ for all $k \in \mathbb{N}$. Now the coercivity of f implies that $L_f(\alpha)$ is bounded. Thus the Heine–Borel Theorem implies the existence of a subsequence $\{x_{k'}\}_{k'}$ that converges to some $x^* \in \mathbb{R}^d$. Since $x^* = \lim_{k'} x_{k'}$ and $f^* = \lim_{k'} f(x_{k'})$ and f is lower semi-continuous, we can conclude from Lemma 1.27 that $f(x^*) \le f^*$. Thus $f(x^*) \le f(x)$ for all $x \in \mathbb{R}^d$, which shows that x^* is a global solution of the optimisation problem $\min_{x\in\mathbb{R}^d} f(x)$. In particular, a global solution exists. □

In the constrained case, we can get rid of the coercivity of f if the feasible set Ω is bounded:

Proposition 1.34. *Assume that $f \colon \mathbb{R}^d \to \bar{\mathbb{R}}$ is lower semi-continuous and that $\Omega \subset \mathbb{R}^d$ is non-empty, closed and bounded. Then the optimisation problem $\min_{x\in\Omega} f(x)$ has a global solution.*

Proof. The proof is almost identical to that of Theorem 1.33: We define again $f^* := \inf_{x\in\Omega} f(x) < +\infty$. Then there exists a sequence $\{x_k\}_{k\in\mathbb{N}} \subset \Omega$ with $f^* = \lim_{k\to\infty} f(x_k)$. Because of the boundedness of Ω, there exists a subsequence $\{x_{k'}\}_{k'}$ and some $x^* \in \mathbb{R}^d$ such that $x^* = \lim_{k'} x_{k'}$. The lower semi-continuity of f then again implies that $f(x^*) \le f^* \le f(x)$ for all $x \in \Omega$. Finally, the set Ω is closed and thus x^* actually is an element of Ω. This shows that x^* is a global solution of $\min_{x\in\Omega} f(x)$. □

While these existence results cover a large family of optimisation problems, they still might require some adaptation to concrete cases. Thus it is more important to understand the idea of the proof than the results themselves.

A general approach to proving existence of solutions of optimisation problems $\min_{x\in\Omega} f(x)$ with lower semi-continuous f is the following (also known as *the direct method in the calculus of variation*):

- Choose a minimising sequence, that is, a sequence $(x_k)_{k\in\mathbb{N}}$ such that $f(x_k)$ converges to $f^* := \inf_{x\in\Omega} f(x)$. Such a sequence always exists (provided that Ω is non-empty).
- Show by whatever means that you can choose the minimising sequence in such a way that it is bounded.
- Use the Heine–Borel Theorem to extract a convergent subsequence $(x_{k'})$.
- Show that the limit $x^* := \lim_{k'} x_{k'}$ satisfies $x^* \in \Omega$.
- Conclude from the lower semi-continuity of f that x^* solves the optimisation problem at hand.

Remark 1.35. It is important to note here that the coercivity and lower semi-continuity of f are sufficient, but not necessary conditions for the existence of a global solution of the problem $\min_{x\in\mathbb{R}^d} f(x)$. That is, from the fact that f is not coercive or not lower semi-continuous, we cannot immediately conclude that the problem $\min_{x\in\mathbb{R}^d}$ has no solution. For instance, the function $f(x) = 1 - e^{x^2}$ if $x \neq 1$ and $f(1) = 17$ is neither coercive nor lower semi-continuous, but it still has a unique global minimum, namely $x^* = 0$.

2. Characterisation of Solutions

We now discuss how we can characterise local solutions of unconstrained optimisation problems by the properties of their derivatives. Later in the course, we will look at the much harder problem of characterising solutions of *constrained* problems as well. The basis of everything that follows are first and second order Taylor expansions of sufficiently regular functions.

To that end, recall that $\nabla f(x) \in \mathbb{R}^d$ denotes the gradient of the differentiable function $f: \mathbb{R}^d \to \mathbb{R}$ (that is, the vector of its partial derivatives), and $H_f(x) \in \mathbb{R}^{d\times d}$ denotes the Hessian of the twice differentiable function $f: \mathbb{R}^d \to \mathbb{R}$, that is, the matrix of its second order partial derivatives. For a refresher in differentiation of multi-variate functions, it can be a good idea to have a look at Section B in the appendix, where you can find the most important notation and results. Moreover, many of the results will invoke the positive definiteness or positive semi-definiteness of the Hessian of f. In order to refresh your knowledge in these notions, you might want to skim through Section A in the appendix, which is concerned with symmetric matrices. There you can also find different methods for checking whether some given matrix is positive definite.

In particular, it is worth remembering Taylor's theorem, which will be used in various settings quite a few times throughout this course:

Theorem 2.1 (Taylor)**.** *Assume that $f \in C^1(\mathbb{R}^d)$ and that $x^* \in \mathbb{R}^d$. Then*

$$f(x) = f(x^*) + \langle \nabla f(x^*), x - x^*\rangle + o(\|x - x^*\|) \qquad \text{as } x \to x^*.$$

If $f \in C^2(\mathbb{R}^d)$, then

$$f(x) = f(x^*) + \langle \nabla f(x^*), x - x^*\rangle + \frac{1}{2}\langle x - x^*, H_f(x^*)(x - x^*)\rangle + o(\|x - x^*\|^2) \quad \text{as } x \to x^*.$$

Moreover there exists for every $x \in \mathbb{R}^d$ some $0 < t_x < 1$ such that

$$f(x) = f(x^*) + \langle \nabla f(x^*), x - x^*\rangle + \frac{1}{2}\langle x - x^*, H_f(tx + (1-t)x^*)(x - x^*)\rangle.$$

Proof. See Calculus 2. □

2.1. Necessary Conditions.

Theorem 2.2 (First order necessary condition)**.** *Assume that $f \in C^1(\mathbb{R}^d)$ and that x^* is a local solution of the optimisation problem*

$$\min_{x\in\mathbb{R}^d} f(x).$$

Then

$$\nabla f(x^*) = 0.$$

Proof. Assume to the contrary that $\nabla f(x^*) =: p \neq 0$. Then a first order Taylor series expansion of $f(x^* - tp)$ with $t > 0$ implies that

$$\begin{aligned}\lim_{t\to 0^+} \frac{f(x^* - tp) - f(x^*)}{t} &= \lim_{t\to 0^+} \frac{\langle \nabla f(x^*), -tp\rangle + o(\|tp\|)}{t} \\ &= -\|p\|^2 + \lim_{t\to 0^+} \frac{o(\|tp\|)}{t} = -\|p\|^2 < 0.\end{aligned}$$

This, however, means that $f(x^* - tp) < f(x^*)$ for all sufficiently small $t > 0$, which in turn implies that x^* cannot be a local minimiser of f. □

Theorem 2.3 (Second order necessary conditions)**.** *Assume that $f \in C^2(\mathbb{R}^d)$ and that x^* is a local solution of the optimisation problem*

$$\min_{x \in \mathbb{R}^d} f(x).$$

Then

$$\nabla f(x^*) = 0$$

and $H_f(x^*) \in \mathbb{R}^{d \times d}$ is positive semi-definite.

Proof. From Theorem 2.2 we already know that $\nabla f(x^*) = 0$. Assume therefore that $\nabla f(x^*) = 0$ but $H_f(x^*)$ is not positive semi-definite. Then $H_f(x^*)$ has a negative eigenvalue $\lambda < 0$ with associated eigenvector $p \in \mathbb{R}^d$. Now a second order Taylor series expansion implies that

$$\begin{aligned}\lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(x^* + tp) - f(x^*)}{t^2} &= \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{\langle \nabla f(x^*), tp\rangle + \langle tp, H_f(x^*)(tp)\rangle/2 + o(\|tp\|^2)}{t^2} \\ &= \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{t^2 \langle p, H_f(x^*)p\rangle + o(\|tp\|^2)}{t^2} = \lim_{\substack{t \to 0 \\ t \neq 0}} \frac{t^2 \lambda \|p\|^2 + o(\|tp\|^2)}{t^2} = \lambda \|p\|^2 < 0.\end{aligned}$$

Thus $f(x^* + tp) < f(x^*)$ for all $t \neq 0$ sufficiently close to 0, which again implies that x^* cannot be a local minimiser of f. □

Note that the previous conditions are necessary, but not sufficient, for x^* to be a local minimiser of f:

Example 2.4. Consider the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = -x^4$. Then $f'(0) = 0$ and $f''(0) = 0$. Thus the second order necessary optimality conditions are satisfied for $x^* = 0$, but 0 is no local minimser of f (but rather a global maximiser).

2.2. **Sufficient Conditions.** In order to obtain a sufficient optimality condition, one has to strengthen the positive semi-definiteness of the Hessian to positive definiteness:

Theorem 2.5. *Assume that $f \in C^2(\mathbb{R}^d)$ and that $x^* \in \mathbb{R}^d$ is such that $\nabla f(x^*) = 0$ and $H_f(x^*)$ is positive definite. Then x^* is an isolated and strict local minimiser of f.*

Proof. Since $H_f(x^*)$ is positive definite, there exists $\sigma > 0$ (the smallest singular value of $H_f(x^*)$) such that

$$\langle p, H_f(x^*)\rangle \geq \sigma \|p\|^2$$

for all $p \in \mathbb{R}^d$. Thus we obtain from a second order Taylor series expansion of f around x^* that

$$\begin{aligned} f(x^* + p) &= f(x^*) + \langle \nabla f(x^*), p\rangle + \frac{1}{2}\langle p, H_f(x^*)p\rangle + o(\|p\|^2) \\ &\geq f(x^*) + \frac{\sigma}{2}\|p\|^2 + o(\|p\|^2) \text{ as } p \to 0.\end{aligned}$$

For sufficiently small $\|p\|$, the term $o(\|p\|^2)$ is dominated by $\frac{\sigma}{4}\|p\|^2$. Thus we have that

$$f(x^* + p) \geq f(x^*) + \frac{\sigma}{4}\|p\|^2$$

for all p with $\|p\|$ sufficiently small. This shows that x^* is a strict local minimiser of f.

We still have to show that x^* is an isolated local minimiser of f. To that end, we perform a first order Taylor series expansion of ∇f around x^* and obtain that

$$\nabla f(x^* + p) = \nabla f(x^*) + H_f(x^*)p + o(\|p\|) = H_f(x^*)p + o(\|p\|).$$

Taking the inner product with p, this implies that

$$\langle p, \nabla f(x^* + p)\rangle = \langle p, H_f(x^*)\rangle + \langle p, o(\|p\|)\rangle \geq \sigma\|p\|^2 + o(\|p\|^2).$$

Again, the term $\sigma\|p\|^2/2$ dominates the term $o(\|p\|^2)$ for sufficiently small $\|p\|$, which implies that

$$\langle p, \nabla f(x^* + p)\rangle \geq \frac{\sigma}{2}\|p\|^2$$

for all p with $\|p\|$ sufficiently small. In particular, this shows that $\nabla f(x^* + p) \neq 0$ if $p \neq 0$ is sufficiently close to 0. In view of the first order necessary optimality condition, this further implies that no point $x^* + p$ with $p \neq 0$ sufficiently close to 0 can be a local minimiser of f. In other words, x^* is an isolated local minimiser. □

Definition 2.6. Let $f \in C^1(\mathbb{R}^d)$. A point $x^* \in \mathbb{R}^d$ with $\nabla f(x^*) = 0$ is called a *critical point* of f.

Remark 2.7. From the first order optimality condition, it follows that every local minimiser (or maximiser) of a C^1 function f is also a critical point. Conversely, every critical point may be a local minimiser, a local maximiser, or neither of these. Sometimes, the points in the latter category, that is, critical points that are neither local minimisers nor local maximisers, are called saddle points of f. We will, however, not adopt this notation and rather use the term “saddle point” in a more restrictive setting later in the course.

Example 2.8. Define $f \colon \mathbb{R}^2 \to \mathbb{R}$,

$$f(x, y) = 3x^4 + 4x^3 + 12y^2 - 24xy.$$

We want to compute the local and global minimisers of f on $\mathbb{R}^2$. To that end, we first compute the gradient and Hessian of f, which are

$$\nabla f(x, y) = \begin{pmatrix} 12x^3 + 12x^2 - 24y \\ 24y - 24x \end{pmatrix}$$

and

$$H_f(x, y) = \begin{pmatrix} 36x^2 + 24x & -24 \\ -24 & 24 \end{pmatrix}.$$

For the computation of the critical points of f, we have to solve the system of equations

$$\nabla f(x, y) = \begin{pmatrix} 12x^3 + 12x^2 - 24y \\ 24y - 24x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

From the second equation, we immediately get that $x = y$. Inserting this result in the first equation, we obtain the condition

$$12x^3 + 12x^2 - 24x = 0.$$

This equation has three solutions $x = -2$, $x = 0$, and $x = 1$. We thus have the three critical points $(-2, -2)$, $(0, 0)$, and $(1, 1)$.

- The point $(-2, -2)$. Here the Hessian of f is

  $$H_f(-2, -2) = \begin{pmatrix} 96 & -24 \\ -24 & 24 \end{pmatrix},$$

  which is positive definite.[2] Thus this point is a strict local minimiser.
- The point $(0, 0)$. Here the Hessian of f is

  $$H_f(0, 0) = \begin{pmatrix} 0 & -24 \\ -24 & 24 \end{pmatrix},$$

  which is indefinite. Thus this point is no local minimiser (and neither a maximiser).

[2] If you disagree in that being obvious, please have a look at Lemma A.9.

- The point $(1,1)$. Here the Hessian of f is
$$H_f(1,1) = \begin{pmatrix} 60 & -24 \\ -24 & 24 \end{pmatrix},$$
which is positive definite. Thus this point is a strict local minimiser.

Finally, we want to determine whether any of the points $(-2,-2)$ and $(1,1)$ is a global minimiser. Computing the function values at these points, we obtain that $f(-2,-2) = -8$ and $f(1,1) = -5$. Thus $(-2,-2)$ is the only possible candidate for a global minimiser. Note, however, that we cannot yet conclude that $(-2,-2)$ actually *is* a global minimiser: From all what we know about the function, it could still be possible that f has no global minimisers at all and, for instance, is unbounded below. We thus try to investigate whether the function f is coercive. To that end, we note that the inequality
$$4xy \le 4x^2 + y^2$$
holds for all x, $y \in \mathbb{R}$: This follows from the fact that
$$4x^2 - 4xy + y^2 = (2x-y)^2 \ge 0.$$
Thus we can estimate
$$f(x,y) = 3x^4 + 4x^3 + 12y^2 - 24xy \ge 3x^4 + 4x^3 - 24x^2 + 6y^2 \ge x^4 + 6y^2 - C$$
for some $C > 0$ and all x, $y \in \mathbb{R}$.[3] This shows that f is coercive. Since it is obviously continuous, we can conclude that f admits a global minimum, and, as seen above, the only candidate for this global minimum is the point $(-2,-2)$. This shows that $(-2,-2)$ is actually the (unique) strict global minimum of f.

3. Optimisation and Convex Functions

We now give a very brief introduction into the notion of convexity and into its importance within the field of optimisation.

3.1. Convex Functions and Sets.

Definition 3.1 (Convex Function)**.** A function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *convex*, if for every x, $y \in \mathbb{R}^d$ and $0 \le \lambda \le 1$ the inequality
$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda) f(y)$$
holds.

The function f is *strictly convex*, if for every $x \ne y \in \mathbb{R}^d$ and $0 < \lambda < 1$ the inequality
$$f\big(\lambda x + (1-\lambda)y\big) < \lambda f(x) + (1-\lambda) f(y)$$
holds. That is, the inequality defining the convexity of a function is strict whenever possible.

More graphically, this means that for each pair of points $(x, f(x))$ and $(y, f(y))$ lying on the graph of f, the connecting line segment remains above (or rather: not below) the graph. It is strictly convex if the connecting line segment stays strictly above the graph. See Figure 2.

Similarly, we can also define convex sets:

Definition 3.2. A set $C \subset \mathbb{R}^d$ is convex, if for all points x, $y \in C$ and $0 \le \lambda \le 1$ we have
$$\lambda x + (1-\lambda) y \in C.$$

[3] If we really want, we can, for instance, argue here that $2x^4 + 4x^3 - 24x^2 \ge 0$ whenever $|x| \ge 5$ and that $2x^4 + 4x^3 - 24x^2 \ge -4 \cdot 5^3 - 24 \cdot 5^2 = -1100$ whenever $|x| \le 5$. Thus $C = -1100$ would work.

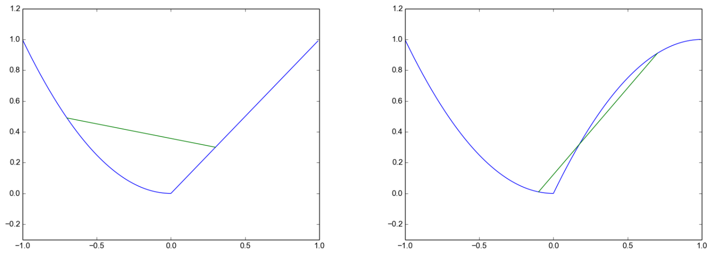

FIGURE 2. *Left:* Typical example of a convex (but not strictly convex) function. Note that no differentiability is assumed. *Right:* Typical example of a non-convex function. There exist points on the graph such that the connecting line segment does not lie completely above the graph.

That is, a set is convex, if whenever we are given two points x and y in C the whole line segment connecting these two points is also contained in C.

Remark 3.3. It is also possible to define convexity for functions $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ that may take the values $\pm\infty$. However, in this case the definition is a bit more intricate in particular in the case where we simultaneously have function values $f(x) = +\infty$ and $f(y) = -\infty$. The reason for this is that our definition of convexity would lead to sums of the form $+\infty + (-\infty)$, which are undefined as they stand. Because of these complications, we will restrict ourselves to finite valued convex functions.

An important property of convex functions, which we will need later in the course when dealing with constrained optimisation problems, is the fact that lower level sets of convex functions are convex sets:

Lemma 3.4. *Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is a convex function. Then $L_f(\alpha) \subset \mathbb{R}^d$ is a convex set for every $\alpha \in \mathbb{R}$.*

Proof. Let $\alpha \in \mathbb{R}$ be fixed and assume that $x, y \in L_f(\alpha)$. That is, $f(x) \le \alpha$ and $f(y) \le \alpha$. Let moreover $0 < \lambda < 1$. Then the convexity of f implies that

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda)f(y) \le \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

(Note here that we have used that $\lambda \ge 0$ and $(1-\lambda) \ge 0$.) This shows that $\lambda x + (1-\lambda)y \in L_f(\alpha)$ which in turn shows the convexity of $L_f(\alpha)$. □

Corollary 3.5. *Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is convex. Then the set of global minimisers of f is convex.*

Proof. This is a direct consequence of Lemma 3.4 with $\alpha := \inf_{x\in\mathbb{R}^d} f(x)$. □

Remark 3.6. Note that the converse of Lemma 3.4 does not hold. That is, from the fact that all level sets $L_f(\alpha)$ of a function $f\colon \mathbb{R}^d$ are convex, we cannot conclude that f is convex. A simple counterexample is the function $f(x) = x^3$. The level sets of f are simply the half-lines $L_f(\alpha) = (-\infty, \alpha^{1/3}]$ which are obviously convex. However, the function f is not convex, as, for instance, $f(-1) = -1 > \frac{1}{2}(f(-2) + f(0)) = \frac{1}{2}(-8+0) = -4$.

Remark 3.7. There is a very close connection between convex sets and convex functions: One can show that a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is convex, if and only if the so-called *epigraph* of f, which is the subset of $\mathbb{R}^d \times \mathbb{R}$ consisting of all points (x,t) with $t \ge f(x)$, is a convex set.

It is easy to show the following properties of convex functions:

- If the functions f, $g\colon \mathbb{R}^d \to \mathbb{R}$ are convex, then so is the function $f+g$.
- If $f\colon \mathbb{R}^d \to \mathbb{R}$ is convex and $\lambda \geq 0$, then also the function λf is convex.
- Every linear (or affine) function is convex.
- If both f and $-f$ are convex, then the function f is affine (that is, $f(x) = \langle a, x\rangle + b$ for some $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$).
- If f and g are convex functions, then the function h defined by $h(x) := \max\{f(x), g(x)\}$ is also convex.

From the point of view of optimisation, one of the many good properties of convex functions is the fact that there is no difference between global and local minima; instead, every local minimum is automatically global, as the following result shows.

Lemma 3.8. *Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is convex. Then every local minimiser of f is already a global minimiser.*

Proof. Assume to the contrary that $x \in \mathbb{R}^d$ is no global minimiser of f. Then there exists $y \neq x$ with $f(y) < f(x)$. However, because of the convexity of f we have for every $0 < \lambda < 1$ that

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) < \lambda f(x) + (1-\lambda)f(x) = f(x).$$

That is, setting $x_\lambda = \lambda x + (1-\lambda)y$ we have $f(x_\lambda) < f(x)$ for all $0 < \lambda < 1$. Since $x_\lambda \to x$ as $\lambda \to 1$, this shows that x cannot be a local minimiser of f. □

3.2. **Differentiable Convex Functions.** In the definition of convex functions above, we have not assumed any regularity of f (apart from f only taking finite values). Indeed, one of the main advantages of the (rather extensive) theory of convex functions is that it allows to deal with non-differentiable functions using almost the same methods as we would use for differentiable functions. In particular, it is possible to introduce generalised notions of derivatives that in turn can be used for the characterisation and computation of solutions of optimisation problems. However, we will consider in the following *differentiable* convex functions, and we will study what the convexity of a function implies for its derivative.[4]

Proposition 3.9. *Assume that the function $f\colon \mathbb{R}^d \to \mathbb{R}$ is differentiable. Then f is convex, if and only if for every x, $y \in \mathbb{R}^d$ the inequality*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle \tag{2}$$

is satisfied.

Proof. Assume first that f is convex and let $x \neq y \in \mathbb{R}^d$. We note that $\langle \nabla f(x), y-x\rangle$ is precisely the directional derivative of f at the point x in direction $(y-x)$, that is,

$$\langle \nabla f(x), y - x\rangle = \lim_{t\to 0} \frac{1}{t}\bigl(f(x + t(y-x)) - f(x)\bigr).$$

Using the convexity of f and that

$$x + t(y-x) = ty + (1-t)x,$$

[4]We will not discuss non-continuous convex functions—the main reason being that they do not exist in the setting used here: It can be shown that every function $f\colon \mathbb{R}^d \to \mathbb{R}$ is not only continuous, but actually locally Lipschitz continuous. Moreover, this implies that convex functions are actually almost everywhere differentiable (in the sense that the set of points where the derivative does not exists has Lebesgue measure zero).

It *is* possible to construct non-continuous convex functions, but only if one either restricts the domain of the function to some (non-open) convex subset C of $\mathbb{R}^d$ (that is, we have a function $f\colon C \to \mathbb{R}$), or one allows the function to take the value $+\infty$. For instance, the function $f(x) = x$ for $x > 0$, and $f(0) = 1$ is convex on the half line $\mathbb{R}_{\geq 0}$, but discontinuous at 0.

we can estimate the right hand side by

$$\frac{1}{t}\big(f(x+t(y-x))-f(x)\big) \le \frac{1}{t}\big(tf(y)+(1-t)f(x)-f(x)\big) = f(y)-f(x)$$

for $0 < t < 1$, implying that

$$\langle \nabla f(x), y-x\rangle \le f(y)-f(x),$$

which is precisely (2).

Assume now that the inequality (2) holds for all x, $y \in \mathbb{R}^d$. Let w, $z \in \mathbb{R}^d$ and $0 \le \lambda \le 1$. Denote moreover

$$x := \lambda w + (1-\lambda)z.$$

Then the inequality (2) implies that

$$\begin{aligned} f(w) &\ge f(x) + \langle \nabla f(x), w-x\rangle, \\ f(z) &\ge f(x) + \langle \nabla f(x), z-x\rangle. \end{aligned} \tag{3}$$

Note moreover that

$$w-x = (1-\lambda)(w-z) \qquad \text{and} \qquad z-x = \lambda(z-w).$$

Thus, if we multiply the first line in (3) with λ, the second line with $1-\lambda$, and then add the two inequalities, we obtain

$$\begin{aligned} \lambda f(w) + (1-\lambda)f(z) \ge f(x) + \lambda\langle \nabla f(x), (1-\lambda)(w-z)\rangle + (1-\lambda)\langle \nabla f(x), \lambda(z-w)\rangle \\ = f\big(\lambda w + (1-\lambda)z\big). \end{aligned}$$

Since w and z were arbitrary, this proves the convexity of f. □

Remark 3.10. Following basically the same proof as above and strategically replacing inequalities by strict inequalities, one can show that a differentiable function f is strictly convex, if and only if

$$f(y) > f(x) + \langle \nabla f(x), y-x\rangle$$

whenever $x \ne y \in \mathbb{R}^d$.

Remark 3.11. Originally, we have introduced the convexity of functions by requiring that the graph of the function lies below all the line segments connecting two points on the graph. From the previous result, however, we obtain an alternative characterisation of convexity that requires that the graph of the function lies *above* all the tangents to the function. While the first definition of convexity may seem to be more natural and more general, a definition of convexity of functions by means of the properties of the "tangents" to the function turns out to be very useful in the context of optimisation and leads to interesting generalisations of the notion of convexity. In this course we, unfortunately, don't have the time to discuss these ideas further, though.

As an immediate consequence of Proposition 3.9 one obtains the result that the first order necessary condition for a minimiser is, in the case of convex functions, also a sufficient condition. More precisely, the following holds:

Corollary 3.12. *Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable. Then x^* is a global minimiser of f, if and only if $\nabla f(x^*) = 0$.*

Proof. First recall that the condition $\nabla f(x^*) = 0$ is, independent of the convexity of f, a necessary condition for x^* to be a global (and indeed already local) minimiser. Thus we only need to show that this condition actually implies that x^* *is* a global minimiser. Assume therefore that $\nabla f(x^*) = 0$ and let $y \in \mathbb{R}^d$. Then Proposition 3.9 implies that

$$f(y) \geq f(x^*) + \langle \nabla f(x^*), y - x^* \rangle = f(x^*).$$

Thus x^* is a global minimiser. $\square$

Proposition 3.13. *Assume that the function $f\colon \mathbb{R}^d \to \mathbb{R}$ is differentiable. Then f is convex, if and only if for every x, $y \in \mathbb{R}^d$ the inequality*

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \tag{4}$$

is satisfied.

Proof. Assume first that f is convex and that x, $y \in \mathbb{R}^d$. Then Proposition 3.9 implies the two inequalities

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

Adding these inequalities, we obtain that

$$f(y) + f(x) \geq f(x) + f(y) + \langle \nabla f(x) - \nabla f(y), y - x \rangle,$$

which simplifies to

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0.$$

Conversely, assume that (4) holds for all x, $y \in \mathbb{R}^d$. Let moreover z, $x \in \mathbb{R}^d$. Then the mean value theorem implies that there exists $0 < \lambda < 1$ such that

$$f(z) - f(x) = \langle \nabla f\big(x + \lambda(z - x)\big), z - x \rangle.$$

Now we can write

$$\begin{aligned} \langle \nabla f\big(x + \lambda(z - x)\big), z - x \rangle &= \langle \nabla f\big(x + \lambda(z - x)\big) - \nabla f(x), z - x \rangle + \langle \nabla f(x), z - x \rangle \\ &= \frac{1}{\lambda} \langle \nabla f\big(x + \lambda(z - x)\big) - \nabla f(x), x + \lambda(z - x) - x \rangle + \langle \nabla f(x), z - x \rangle. \end{aligned}$$

Applying (4) with $y = x + \lambda(z - x)$ (and recalling that $\lambda > 0$), it follows that

$$\langle \nabla f\big(x + \lambda(z - x)\big), z - x \rangle \geq \langle \nabla f(x), z - x \rangle.$$

Thus

$$f(z) - f(x) \geq \langle \nabla f(x), z - x \rangle,$$

which is just a reformulation of (2). Since this holds for every z, $x \in \mathbb{R}^d$, we obtain the convexity of f from Proposition (3.9). $\square$

Remark 3.14. Again, one can show with essentially the same proof that f is strictly convex, if and only if

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle > 0$$

whenever $x \neq y \in \mathbb{R}^d$.

Remark 3.15. A function $G\colon \mathbb{R}^d \to \mathbb{R}^d$ is called *monotone*, if

$$\langle G(y) - G(x), y - x \rangle \geq 0$$

for every y, $x \in \mathbb{R}^d$. (In the particular case $d = 1$, this actually is the same as stating that G is monotonically non-decreasing; do check this equivalence!) With this notation, Proposition 3.13 can be reformulated as stating that f is convex if and only if ∇f is monotone.

3.3. **Hessians of Convex Functions.**

Proposition 3.16. *A twice differentiable function $f\colon \mathbb{R}^d \to \mathbb{R}$ is convex, if and only if the Hessian $H_f(x)$ is positive semi-definite for all $x \in \mathbb{R}^d$.*

Proof. Assume first that f is convex and let $x \in \mathbb{R}^d$. Define moreover the function $g\colon \mathbb{R}^d \to \mathbb{R}$ setting

$$g(y) := f(y) - \langle \nabla f(x), y - x\rangle.$$

Since the mapping $y \mapsto -\langle \nabla f(x), y - x\rangle$ is affine, it follows that g is convex. Moreover

$$\nabla g(y) = \nabla f(y) - \nabla f(x)$$

and

$$H_g(y) = H_f(y)$$

for all $y \in \mathbb{R}^d$. In particular, $\nabla g(x) = 0$. Thus Corollary 3.12 implies that x is a global minimiser of g. Now the second order necessary condition for a minimiser implies that $H_g(x)$ is positive semi-definite. Since $H_g(x) = H_f(x)$ and x was arbitrary, this proves that the Hessian of f is positive semi-definite for all $x \in \mathbb{R}^d$.

Now assume that the Hessian $H_f(x)$ of f is positive semi-definite for all $x \in \mathbb{R}^d$. Let moreover x, $y \in \mathbb{R}^d$. Then Taylor's theorem implies that

$$f(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle y - x, H_f(x + t(y - x))(y - x)\rangle$$

for some $0 \le t \le 1$. Since H_f is everywhere positive semi-definite, the quadratic term in this equation is always non-negative. Thus we can estimate

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle.$$

Proposition 3.9 proves now the convexity of f. □

Remark 3.17. There is *some* relation between the strict convexity of a function f and the positive definiteness of its Hessian. However, this relation is not completely straight-forward. It is possible to show (and actually pretty simple to show) that a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is strictly convex, if its Hessian $H_f(x)$ is positive definite for all x. However, the converse direction does not hold: The strict convexity of a function f does not imply that its Hessian is everywhere positive definite. As an example consider the function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^4$. This function is strictly convex, but $f''(0) = 0$. Still, it is possible to characterise the strict convexity of a univariate function $f\colon \mathbb{R} \to \mathbb{R}$ by the condition that the set of points $x \in \mathbb{R}$ with $f''(x) > 0$ is dense. Thus a twice differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ is strictly convex, if and only if the set $\{x \in \mathbb{R} : f''(x) > 0\}$ is dense in $\mathbb{R}$. To the best of my knowledge, there exists no (simple) generalisation of this characterisation to multivariate functions.

3.4. **Summary.** From the viewpoint of optimisation, the main results concerning convex functions (that we will need/refer to during this class) are:

- Convexity of a differentiable function can either characterised by the fact that all secants lie above the graph (Definition 3.1) or that all tangents lie below the graph (Proposition 3.9).
- If a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable, then the first order necessary condition for a minimum is actually sufficient. That is, the minimisation of f is equivalent to the solution of the equation
$$\nabla f(x) = 0.$$
- A function f is convex, if its Hessian is everywhere positive semi-definite. This allows us to test whether a given function is convex.
- If the Hessian of a function is everywhere positive definite, then the function is strictly convex. The converse does not hold.

Appendix A. Symmetric Matrices

A.1. Definitions.

Definition A.1. A matrix $A \in \mathbb{R}^{d\times d}$ is *symmetric* if $A^T = A$. It is *positive semi-definite* if

$$\langle x, Ax\rangle \geq 0 \text{ for all } x \in \mathbb{R}^d,$$

and *positive definite* if

$$\langle x, Ax\rangle > 0 \text{ for all } x \in \mathbb{R}^d \setminus \{0\}.$$

By $\mathbf{Sym}_d \subset \mathbb{R}^{d\times d}$ we denote the set of all symmetric matrices in $\mathbb{R}^{d\times d}$, by $\mathbf{Sym}_d^+$ the set of all symmetric and positive semi-definite matrices in $\mathbb{R}^{d\times d}$, and by $\mathbf{Sym}_d^{++}$ the set of all symmetric and positive definite matrices in $\mathbb{R}^{d\times d}$.

Remark A.2. Using the same definition as above, it is also possible to extend the notion of positive (semi-)definiteness to non-symmetric matrices. Since we are interested in optimisation, where the matrices will usually be either Hessians or approximations of Hessians, which are symmetric in all reasonable situations, we will not use this generalisation but instead always assume symmetry.

Before discussing the properties of symmetric and positive definite matrices, we will briefly mention the structure of the sets of these matrices.

Lemma A.3. *The set* $\mathbf{Sym}_d$ *is a vector space, the set* $\mathbf{Sym}_d^+$ *is a pointed convex cone, and the set* $\mathbf{Sym}_d^{++}$ *is a convex cone, that is:*

- *If* A, $B \in \mathbf{Sym}_n$ *and* λ, $\mu \in \mathbb{R}$, *then* $\lambda A + \mu B \in \mathbf{Sym}_n$.
- *If* A, $B \in \mathbf{Sym}_n^+$ *and* λ, $\mu \geq 0$, *then* $\lambda A + \mu B \in \mathbf{Sym}_n^+$.
- *If* A, $B \in \mathbf{Sym}_n^{++}$ *and* λ, $\mu > 0$, *then* $\lambda A + \mu B \in \mathbf{Sym}_n^{++}$.

Proof. This immediately follows from the definitions of symmetry and positive (semi-)definiteness. □

In other words, addition of symmetric positive definite matrices and multiplication with positive scalars preserves their structure. Note, however, that multiplication of such matrices does not: in particular, the product of two symmetric matrices cannot be expected to be symmetric again.

A.2. Eigenvalues and Eigenvectors.

We recall that eigenvalues $\lambda \in \mathbb{C}$ of a matrix $A \in \mathbb{R}^{d\times d}$ with corresponding eigenvectors $v \in \mathbb{C}^d \setminus \{0\}$ are defined by the equation

$$Av = \lambda v.$$

Note that in this definition both λ and v maybe complex, even if the matrix A itself is purely real. A typical example of this situation is the matrix

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

describing a rotation of 90 degree, which has the two eigenvalues $\pm i$. In the case of symmetric matrices, however, this situation cannot occur, and one can show quite easily that all eigenvalues of symmetric matrices have to be real:

Lemma A.4. *Assume that* $A \in \mathbf{Sym}_d$ *is a symmetric matrix. Then all eigenvalues of* A *are real. Additionally, every eigenvalue has a corresponding real eigenvector.*

Proof. Assume that λ is an eigenvalue of A with eigenvector $v \in \mathbb{C} \setminus \{0\}$. Then

$$\overline{v}^T Av = \overline{v}^T \lambda v = \lambda \|v\|^2.$$

On the other hand,

$$\overline{v}^T Av = \overline{Av}^T v = \overline{\lambda v}^T v = \overline{\lambda}\|v\|^2$$

showing that $\lambda = \overline{\lambda}$, which means that $\lambda \in \mathbb{R}$.

Now assume that $\lambda \in \mathbb{R}$ is a (real) eigenvalue of A. Then the corresponding eigenvectors are precisely the non-zero solutions of the equation $(A - \lambda I)v = 0$ (with $I \in \mathbb{R}^{n\times n}$ being the identity matrix). Because λ is real, so is $A - \lambda I$, and this matrix does not have full rank, as λ is an eigenvalue of A.[5] Therefore the equation $(A - \lambda I)v = 0$ has a non-zero real solution, which means that A has a real eigenvector corresponding to the eigenvalue λ. □

Additionally, one can show that the eigenvectors of symmetric matrices corresponding to different eigenvalues are always orthogonal:

Lemma A.5. *Assume that $\lambda \neq \mu$ are distinct eigenvalues of the symmetric matrix $A \in \mathbf{Sym}_d$ with corresponding (real) eigenvectors v, $w \in \mathbb{R}^d$. Then*

$$\langle v, w\rangle = 0.$$

Proof. We have

$$\langle v, Aw\rangle = \langle v, \mu w\rangle = \mu\langle v, w\rangle$$

and

$$\langle v, Aw\rangle = \langle A^T v, w\rangle = \langle Av, w\rangle = \langle \lambda v, w\rangle = \lambda\langle v, w\rangle . v^T Aw = (Av)^T w = \lambda v^T w.$$

This implies that

$$(\mu - \lambda)\langle v, w\rangle = 0.$$

Since $\mu \neq \lambda$, we obtain that $\langle v, w\rangle = 0$. □

In addition, one can show that symmetric matrices always have a full set of eigenvectors (that is, a set of n linearly independent eigenvectors):

Theorem A.6. *Every matrix $A \in \mathbf{Sym}_n$ can be decomposed as*

$$A = QDQ^T,$$

where $D \in \mathbb{R}^{n\times n}$ is a diagonal matrix with real diagonal entries (the eigenvalues of A) and $Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix, that is, $Q^T = Q^{-1}$. Moreover, the columns of Q form an (orthonormal) eigenbasis of A.

A different method of interpreting this previous result is by means of looking at the quadratic functional $x \mapsto \langle x, Ax\rangle$:

Lemma A.7. *Assume that $A \in \mathbf{Sym}_d$, denote by $\lambda_i \in \mathbb{R}$, $i = 1, \dots, d$ the eigenvalues of A, and let $q^{(i)} \in \mathbb{R}^d$, $i = 1, \dots, d$, be a corresponding orthonormal eigenbasis. Then*

$$\langle x, Ax\rangle = \sum_{i=1}^{d} \lambda_i \langle x, q^{(i)}\rangle^2.$$

Proof. Since $q^{(i)}$, $i = 1, \dots, d$, is an orthonormal basis of $\mathbb{R}^n$ we can write

$$x = \sum_{i=1}^{d} \langle x, q^{(i)}\rangle q^{(i)}.$$

Thus

$$\langle x, Ax\rangle = \sum_{i=1}^{d} \langle x, q^{(i)}\rangle\langle x, Aq^{(i)}\rangle = \sum_{i=1}^{d} \langle x, q^{(i)}\rangle\langle x, \lambda_i q^{(i)}\rangle = \sum_{i=1}^{d} \lambda_i \langle x, q^{(i)}\rangle^2.$$

□

[5] Note here that the rank of a real matrix stays the same if we regard it as a complex matrix instead.

Theorem A.8. *A matrix $A \in \mathbf{Sym}_d$ is positive semi-definite if and only if all eigenvalues of A are non-negative. It is positive definite, if and only if all eigenvalues are positive.*

Proof. Using the notation of Lemma A.7, we note that the term $\langle x, q_i \rangle^2$ is strictly positive for all $x \neq 0$. Thus the product $\langle x, Ax \rangle$ will be positive if all eigenvalues λ_i are positive, and non-negative if all eigenvalues λ_i are non-negative.

Conversely, if $\lambda_i \leq 0$ for some i, we can choose $x = q^{(i)}$ and see that ${q^{(i)}}^T A q^{(i)} \leq 0$. Additionally, this last inequality is strict, if $\lambda_i < 0$. □

In the case $d = 2$, this result allows us to determine the positive definiteness of a matrix in a pretty straightforward way:

Lemma A.9. *A matrix $A \in \mathbf{Sym}_2$ is positive definite, if and only if* $\det A > 0$ *and* $\operatorname{trace} A > 0$. *It is positive semi-definite, if and only if* $\det A \geq 0$ *and* $\operatorname{trace} A \geq 0$.

Proof. We only show the (slightly more complicated) result for positive semi-definite matrices.

Denote by λ_1, $\lambda_2 \in \mathbb{R}$ the eigenvalues of A. We recall that the trace and determinant of A satisfy the relations

$$\det A = \lambda_1 \lambda_2 \qquad \text{and} \qquad \operatorname{trace} A = \lambda_1 + \lambda_2.$$

Assume now that A is positive semi-definite. Then we obtain from the previous result that λ_1, $\lambda_2 \geq 0$, and thus we immediately see that $\det A \geq 0$ and $\operatorname{trace} A \geq 0$ (as a product and sum of non-negative numbers, respectively).

Conversely, assume that $\det A \geq 0$ and $\operatorname{trace} A \geq 0$. Then, by the above formula for trace and determinant, we obtain that $\lambda_1 \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 \geq 0$. Now, if we actually have that $\lambda_1 \lambda_2 > 0$, then either both eigenvalues are strictly positive, or both eigenvalues are strictly negative. Since in addition $\lambda_1 + \lambda_2 \geq 0$, the latter situation is impossible, and thus λ_1, $\lambda_2 > 0$, which shows that A is actually positive definite.

On the other hand, if $\lambda_1 \lambda_2 = 0$, then necessarily one of the eigenvalues has to be equal to 0, say $\lambda_1 = 0$. Then $\lambda_2 = \lambda_1 + \lambda_2 = \operatorname{trace} A \geq 0$, which shows that $\lambda_1 = 0$ and $\lambda_2 \geq 0$ and thus A is positive semi-definite. □

Unfortunately, the situation in higher dimensions is more complicated. Therefore we will discuss some alternative characterisations of positive definiteness.

A.3. **Diagonal Dominance.** We recall first the Gershgorin disk theorem, which states that the eigenvalues of an arbitrary matrix can be found in certain disks centered at the diagonal elements.

Theorem A.10. *Assume that $\lambda \in \mathbb{C}$ is an eigenvalue of the matrix $A \in \mathbb{C}^{d \times d}$. Then there exists some index $1 \leq i \leq d$ such that*

$$|\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|.$$

In the particular case of a symmetric matrix (where the eigenvalues are real) we can use this theorem for restricting the eigenvalues of A to an easily computable interval.

Lemma A.11. *Assume that $A \in \mathbf{Sym}_d$ and define*

$$r^- := \min_i \Bigl(a_{ii} - \sum_{j \neq i} |a_{ij}| \Bigr), r^+ \quad := \max_i \Bigl(a_{ii} + \sum_{j \neq i} |a_{ij}| \Bigr).$$

Then all eigenvalues of A are contained in the intervall $[r^-, r^+]$.

Proof. This is an immediate consequence of Theorem A.10 together with the fact that the eigenvalues of symmetric matrices are real. □

In order to use this result for testing symmetric matrices for positive definiteness, we now introduce the notion of diagonally dominant matrices.

Definition A.12. A matrix $A \in \mathbb{C}^{d\times d}$ is diagonally dominant, if

$$|a_{ii}| \geq \sum_{j\neq i} |a_{ij}|$$

for all $1 \leq i \leq d$. It is strictly diagonally dominant, if this inequality is strict for all i.

Lemma A.13. *Assume that $A \in \mathbf{Sym}_d$ is diagonally dominant and $a_{ii} \geq 0$ for all i. Then A is positive semi-definite.*

Similarly, if $A \in \mathbf{Sym}_d$ is strictly diagonally dominant and $a_{ii} > 0$ for all i, then A is positive definite.

Proof. This immediately follows from Lemma A.11. □

Since (strict) diagonal dominance of a matrix can be tested very easily, this provides an efficient method for determining that a given matrix is positive definite. However, the result has the distinct drawback that it is not and "only if" statement: every strictly diagonally dominant, symmetric matrix with positive diagonal elements is positive definite, but not every positive definite matrix is strictly diagonally dominant. Therefore we need additional criteria for deciding on the positive definiteness of non-diagonally-dominant matrices.

A.4. **Cholesky Decomposition.**

Definition A.14. A *Cholesky decomposition* of a matrix $A \in \mathbb{R}^{d\times d}$ is a decomposition of the form

$$A = LL^T,$$

where $L \in \mathbb{R}^{d\times d}$ is a lower triangular matrix with positive diagonal entries.

It is easy to see that the existence of a Cholesky decomposition requires A to be symmetric and positive definite: The symmetry follows from the fact that $(LL^T) = LL^T$ is symmetric, while the positive definiteness follows from the equality

$$\langle x, Ax\rangle = \langle x, L^T Lx\rangle = \langle Lx, Lx\rangle = \|L^T x\|^2$$

and the fact that L, being a triangular matrix with non-zero diagonal elements, is invertible. Additionally, it is possible to show that the converse also holds:

Theorem A.15. *A matrix $A \in \mathbb{R}^{d\times d}$ has a Cholesky decomposition if and only if A is symmetric and positive definite.*

The criterion concerning the Cholesky decomposition is very appealing from a numerical point of view. It is possible to formulate an algorithm requiring $O(d^3)$ computations that either computes the Cholesky decomposition of a given symmetric matrix A or determines that A is not positive definite. Additionally, the Cholesky decomposition provides a numerical method for the solution of linear systems $Ax = b$ very similar to the standard Gauß-algorithm: First solve the lower triangular system $Ly = b$, and then upper triangular system $L^T x = y$; both of these systems can be solved in $O(d^2)$ time.

Now note that one of the main situations in numerical optimisation where it is necessary to test whether a given matrix $A \in \mathbf{Sym}_d$ is positive definite occurs when using Newton's method with line search. In case the Hessian of f at the current iterate x_k is positive definite, the Newton direction p_k, which is computed by solving the linear system $H_f(x_k)p_k = -\nabla f(x_k)$, is a descent direction, and line search methods actually make sense. Thus this situation requires not only a test of the positive definiteness of a matrix, but additionally the solution of a linear system involving the same matrix, both of which can be done using the Cholesky decomposition. Additional applications of the Cholesky decomposition for Newton's method can be found in Chapter 3.4 of Nocedal & Wright, *Numerical Optimization.*

A.5. **Principal Minors.** Finally, we discuss a method for deciding on the positive definiteness of a matrix, which can be helpful for theoretical considerations but also for small scale computations performed by hand.

Definition A.16. Given a matrix $A \in \mathbb{R}^{d\times d}$ and a non-empty subset of indices $I \subset \{1,\ldots,d\}$, we denote by

$$A_I := (a_{ij})_{i,j\in I}$$

the *principal submatrix of A corresponding to I.* In the particular case where $I = \{1,\ldots,k\}$ consists of the first k indices, we abbreviate A_I as A_k and denote by

$$A_k := (a_{ij})_{1\le i,j\le k}$$

the *k-th leading principal submatrix* of A.

Additionally, we call $\det A_I$ the *principal minor* corresponding to the I, and $\det A_k$ the *k-th leading principal minor* of A.

Theorem A.17. *A matrix $A \in \mathbb{R}^{d\times d}$ is positive definite, if and only if all its leading principal minors are positive, that is,* $\det A_k > 0$ *for all* $1 \le k \le n$.

As an example, a symmetric matrix $A \in \mathbb{R}^{3\times 3}$ is positive definite, if and only if $a_{11} > 0$, $a_{11}a_{22} > a_{21}^2$, and $\det A > 0$.

Theorem A.18. *A matrix $A \in \mathbb{R}^{d\times d}$ is positive semi-definite, if and only if all its principal minors are non-negative, that is,* $\det A_I \ge 0$ *for all* $I \subset \{1,\ldots,n\}$.

Remark A.19. A particular consequence of this last theorem is that positive semi-definite matrices have non-negative diagonal entries (as $a_{ii} = \det A_{\{i\}}$ is the determinant of the submatrix corresponding to the set $I = \{i\}$). Put differently, if a matrix $A \in \mathbf{Sym}_d$ has a single negative diagonal element, it cannot be positive semi-definite.

Appendix B. Differentiation

We now provide a brief review of differentiation in several variables, with a focus on what is needed for optimisation. For the following, recall that a set $U \subset \mathbb{R}^d$ is open, if there exists for every $x \in U$ some $\varepsilon > 0$ such that $B_\varepsilon(x) := \{y \in \mathbb{R}^d : \|x-y\| < \varepsilon\} \subset U$.

B.1. **Partial Derivatives.** Assume that $U \subset \mathbb{R}^d$ is open and that $f\colon U \to \mathbb{R}$ is a function. Let moreover $x \in U$. We define the partial derivatives

$$\begin{aligned}\partial_{x_i} f(x) &:= \lim_{h\to 0} \frac{f(x+he_i)-f(x)}{h}\\ &= \lim_{h\to 0} \frac{f(x_1,\ldots,x_{i-1},x_i+h,x_{i+1},\ldots,x_d)-f(x_1,\ldots,x_d)}{h}\end{aligned}$$

whenever the limit exists. Here e_i denotes the i-th standard basis vector in $\mathbb{R}^d$. Higher order partial derivatives can then defined inductively as

$$\partial_{x_{i_1}x_{i_2}\ldots x_{i_k}} f(x) = \partial_{x_{i_k}}\partial_{x_{i_1}\ldots x_{i_{k-1}}} f(x)$$

whenever all the limits involved in all the definitions exist. The number k is called the *order* of the derivative.

Note that the sequence in which the partial derivatives are taken *can* make a difference, although for typical functions it does not:

Example B.1. Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$,

$$f(x_1, x_2) = \frac{x_1^3 x_2 - x_1 x_2^3}{x_1^2 + x_2^2}$$

with $f(0,0) := 0$. Then one obtains that all second order partial derivatives exist everywhere. However, we have

$$\partial_{x_1 x_2} f(0,0) = -1 \qquad \text{whereas} \qquad \partial_{x_2 x_1} f(0,0) = +1.$$

(Exercise: Verify this result!)

The following result shows that the sequence of partial derivation does not matter provided that all the partial derivatives are continuous functions:

Theorem B.2. *Let $U \subset \mathbb{R}^d$ be open, let $f\colon U \to \mathbb{R}$ and assume that the partial derivatives $\partial_{x_i x_j} f$ and $\partial_{x_j x_i} f$ exist and are continuous in a neighbourhood of $x \in U$. Then*

$$\partial_{x_i x_j} f(x) = \partial_{x_j x_i} f(x).$$

This result immediately generalises to higher order partial derivatives: If all partial derivatives up to order k of a function f exist and are continuous in a neighbourhood of $x \in \mathbb{R}^n$, then the sequence of partial differentiation does not change the result. In such a case, one often uses a different notation for higher order derivatives using multi-indices, that is, tuples $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$. Given $\boldsymbol{\alpha} \in \mathbb{N}_0^n$, one defines

$$\partial^{\boldsymbol{\alpha}} f(x) := \underbrace{\partial_{x_1} \ldots \partial_{x_1}}_{\alpha_1 \text{ times}} \underbrace{\partial_{x_2} \ldots \partial_{x_2}}_{\alpha_2 \text{ times}} \cdots \underbrace{\partial_{x_d} \ldots \partial_{x_d}}_{\alpha_d \text{ times}} f(x).$$

The order of the derivative $\partial^{\boldsymbol{\alpha}}$ is then equal to

$$|\boldsymbol{\alpha}| := \alpha_1 + \ldots + \alpha_d.$$

Definition B.3. Let $U \subset \mathbb{R}^d$ be open and $k \in \mathbb{N}$. By $C^k(U)$ we denote the set of k-times continuously differentiable functions. That is, if $f\colon U \to \mathbb{R}$ is a function on U, then $f \in C^k(U)$ if and only if all partial derivatives of f up to order k exist for each $x \in U$, and all of the partial derivatives are continuous functions on U.

In addition, we denote by $C^0(U)$ or $C(U)$ the space of continuous functions on U.

Theorem B.4. *Let $U \subset \mathbb{R}^d$ be open and $k \in \mathbb{N}$. Then $C^k(U)$ is a vector space. That is, if f, $g \in C^k(U)$ and λ, $\mu \in \mathbb{R}$, then also $\lambda f + \mu g \in C^k(U)$.*

In addition, partial differentiation is a linear operation between different C^k-spaces:

Lemma B.5. *Let $U \subset \mathbb{R}^d$ be open, let $k \in \mathbb{N}$, and let $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ be a multi-index with $|\boldsymbol{\alpha}| := \ell \le k$. Then we have for all $f \in C^k(U)$ that $\partial^{\boldsymbol{\alpha}} f \in C^{k-\ell}(U)$. In addition, $\partial^{\boldsymbol{\alpha}}\colon C^k(U) \to C^{k-\ell}(U)$ is a linear operator.*

B.2. **Differential and Gradient.** Assume again that $f\colon \mathbb{R}^d \to \mathbb{R}$ and $x \in \mathbb{R}^d$. We say that f is *differentiable* in the point x, if there exists a linear function $Df(x)\colon \mathbb{R}^d \to \mathbb{R}$ such that

$$\lim_{y\to x} \frac{|f(y) - f(x) - Df(x)(y-x)|}{\|y-x\|} = 0.$$

The linear function $Df(x)$ is called the derivative of f at x. Put differently, the affine mapping $y \mapsto f(x) + Df(x)(y-x)$ is an approximation of the function $y \mapsto f(y)$ for which the error becomes much smaller than the distance $\|y-x\|$ of y from x, as y approaches x.

The next result shows (to some extent) the connection between derivatives and partial derivatives.

Lemma B.6. *Assume that f is differentiable in the point x. Then all partial derivatives of f at x exist. Define now the* gradient $\nabla f(x) \in \mathbb{R}^d$ *of f at x as the vector of all partial derivatives, that is,*

$$\nabla f(x) := \begin{pmatrix} \partial_{x_1} f(x) \\ \vdots \\ \partial_{x_n} f(x) \end{pmatrix}.$$

Then

$$Df(x)z = \langle \nabla f(x), z\rangle$$

for all $z \in \mathbb{R}^d$.

Conversely, if all partial derivatives of f exist in a neighbourhood of x and are continuous as a function of x, then f is differentiable at x.

Remark B.7. The existence of all partial derivatives does *not* imply that the multivariate function f is differentiable. Indeed, as a relatively simple consequence of the definition of differentiability of f, we obtain that every differentiable function is continuous: We can estimate

$$\begin{aligned} |f(y) - f(x)| &= |f(y) - f(x) - Df(x)(y-x) + Df(x)(y-x)| \\ &\le |f(y) - f(x) - Df(x)(y-x)| + |Df(x)(y-x)|. \end{aligned}$$

The last term tends to zero as $y \to x$, as linear functions (in finite dimensions) are continuous, and the first term tends to zero as $y \to x$ because of the differentiability of f. In total, this implies that $\lim_{y\to x} f(y) = f(x)$, that is, f is continuous at x if it is differentiable there.

Now consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined as

$$f(x_1, x_2) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & \text{if } (x_1,x_2) \neq (0,0), \\ 0 & \text{if } (x_1,x_2) = (0,0). \end{cases}$$

This function is not continuous in the point $(0,0)$: Indeed, assume that $c \in \mathbb{R}$ and consider the sequence $(x_1^{(k)}, x_2^{(k)}) := (1/k, c/k)$. Then obviously $\|(x_1^{(k)}, x_2^{(k)})\| \to 0$, but

$$f(x_1^{(k)}, x_2^{(k)}) = \frac{c}{1+c^2}$$

for all k. This shows that, unless $c = 0$, we have that $|f(x_1^{(k)}, x_2^{(k)})| \not\to 0$. In this case, we have a function that is constant along all rays starting at $(0,0)$, but the function value along those rays depends on its direction.

Because f is not continuous at $(0,0)$, it cannot be differentiable at that point. However, the *partial derivatives* of f at $(0,0)$ actually exist: We have that $f(0,x_2) =$

$0 = f(x_1, 0)$ for all x_1, $x_2 \in \mathbb{R}$, which immediately implies that

$$\partial_{x_1} f(0,0) = 0 \qquad \text{and} \qquad \partial_{x_2} f(0,0) = 0.$$

Thus, although the partial derivatives of f at $(0,0)$ exist, the function is not differentiable there in the sense introduced above.

B.3. **Directional Derivatives.** Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$, and that x, $z \in \mathbb{R}^d$. We define the *directional derivative* of f at x in direction z as

$$f'(x; z) := \lim_{t \to 0} \frac{f(x + tz) - f(x)}{t}$$

provided the limit exists.

Lemma B.8. *If f is differentiable at x, then f has directional derivatives $Df(x)z$ for all directions $z \in \mathbb{R}^d$, and we have that*

$$f'(x; z) = Df(x)z = \langle \nabla f(x), z \rangle$$

for all $z \in \mathbb{R}^d$.

Remark B.9. The partial derivatives of a function are the same as the directional derivatives in direction of the standard basis vectors e_i. Thus, partial derivatives are a special case of directional derivatives. In particular, we obtain from Lemma B.6 that a function is differentiable at x, if all directional derivatives $f'(y; z)$ exist in a neighbourhood of x and are, for each fixed $x \in \mathbb{R}^d$, continuous at x as a function of y.

Similar to the case of partial derivatives, however, the existence of all directional derivatives does not imply differentiability or even continuity of f. Consider for instance the function

$$f(x_1, x_2) := \begin{cases} 1 & \text{if } x_1^2 = x_2 \text{ and } x_1 \neq 0, \\ 0 & \text{else.} \end{cases}$$

Then (do check this!) all directional derivates of f at $(0,0)$ exist and are equal to 0. The function f, however, is obviously not continuous and thus not differentiable at $(0,0)$.

B.4. **Taylor's Theorem.** In the following, we will introduce the multi-variate generalisation of Taylor's theorem. To that end we define for a multi-index $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ its factorial as

$$\boldsymbol{\alpha}! := \alpha_1!\,\alpha_2!\,\ldots\alpha_d!.$$

Recall also that the order of a multi-index is defined as $|\boldsymbol{\alpha}| := \alpha_1 + \alpha_2 + \ldots + \alpha_d$.

Theorem B.10 (Taylor)**.** *Let $U \subset \mathbb{R}^d$ be open and let $k \in \mathbb{N}$. Assume that $f \in C^k(U)$ and $x \in U$. Then*

$$f(y) = f(x) + \sum_{|\boldsymbol{\alpha}|=1}^{k} \frac{1}{\boldsymbol{\alpha}!} \partial_{\boldsymbol{\alpha}} f(x) \prod_{i=1}^{d} (y_i - x_i)^{\alpha_i} + E_k(y - x)$$

for all $y \in U$, where the error term $E_k\colon \mathbb{R}^d \to \mathbb{R}$ satisfies

$$\lim_{z \to 0} \frac{E_k(z)}{\|z\|^k} = 0.$$

In the particular case $k = 2$, we can write Taylor's theorem in a relatively simple and compact form by introducing the *Hessian* of the (at least twice differentiable) function $f\colon \mathbb{R}^d \to \mathbb{R}$, defined as the matrix

$$H_f(x) := \begin{pmatrix} \partial_{x_1x_1} f(x) & \partial_{x_1x_2} f(x) & \dots & \partial_{x_1x_d} f(x) \\ \partial_{x_2x_1} f(x) & \partial_{x_2x_2} f(x) & \dots & \partial_{x_2x_d} f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{x_dx_1} f(x) & \partial_{x_dx_2} f(x) & \dots & \partial_{x_dx_d} f(x) \end{pmatrix} \in \mathbb{R}^{d\times d}.$$

If $f \in C^2(U)$ and $x \in U$, the Hessian is a symmetric matrix, as, in this case, $\partial_{x_ix_j} f(x) = \partial_{x_jx_i} f(x)$ for all i, j.

Lemma B.11. *Assume that $f \in C^2(U)$ and $x \in U$. Then*

$$f(y) = f(x) + \big\langle \nabla f(x), y - x\big\rangle + \frac{1}{2}\big\langle y - x, H_f(x)(y - x)\big\rangle + E_2(y - x)$$

for all $y \in U$. Here E_2 is the same error term as in Theorem B.10 with $k = 2$.

Remark B.12. Assume that $U \subset \mathbb{R}^d$ is open, $f \in C^2(U)$, and $x \in U$. Let moreover $z \in \mathbb{R}^d$ be fixed such that the whole line segment $[x, x+z] := \big\{x+tz : 0 \le t \le 1\big\}$ is contained in U, and consider the univariate function $t \mapsto g(t) := f(x + tz)$. Then, by the chain rule,

$$g'(t) = \big\langle \nabla f(x + tz), z\big\rangle$$

and

$$g''(t) = \big\langle z, H_f(x + tz)z\big\rangle.$$

Now we can use the univariate Taylor theorem for the function g and deduce the existence of $0 < t < 1$ such that

$$f(x + z) = g(1) = g(0) + g'(0) + \frac{1}{2}g''(t) = f(x) + \big\langle \nabla f(x), z\big\rangle + \frac{1}{2}\big\langle z, H_f(x + tz)z\big\rangle.$$

B.5. **Vector Valued Differentiable Functions.**

B.6. **The Chain Rule.**

Appendix C. Coercivity of Polynomial Functions

Trondheim, Norway
Email address: markus.grasmair@gmail.com