MIT 6.7220/15.084 — Nonlinear Optimization (Spring '25) Thu, Feb 13th 2025

Lecture 4
The special case of convex functions

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

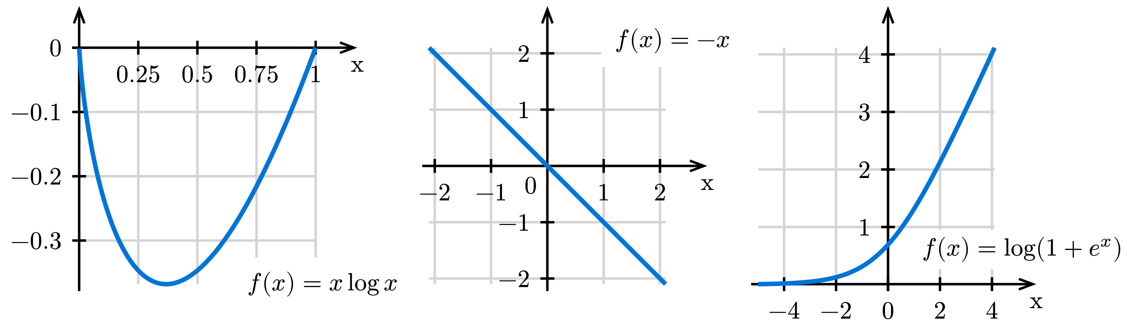
In the previous lecture, we have seen how any solution x to a nonlinear optimization problem defined on a convex feasible set $\Omega \subseteq \mathbb{R}^n$ must necessarily satisfy the first-order optimality condition

$$\langle \nabla f(x), y - x \rangle \geq 0 \qquad \forall y \in \Omega.$$

In general, this optimality condition is only *necessary* but *not sufficient*. However, there exists a notable class of functions for which such a condition *is* sufficient. These are called *convex functions*, and are the topic of today's lecture.

L4.1 Convex functions

Intuitively, a good mental picture for convex functions is as functions that "curve upward" (think of a bowl for example). All the following functions are convex:



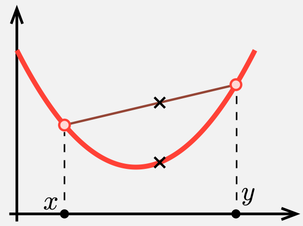
In particular, due to their curvature, local optima of these functions are also global optima, and the first-order optimality condition completely characterizes optimal points. To capture the condition on the curvature in the most general terms (that is, without even assuming differentiability of the function), the following definition is used.

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

Definition L4.1 (Convex function)**.** Let $\Omega \subseteq \mathbb{R}^n$ be convex.

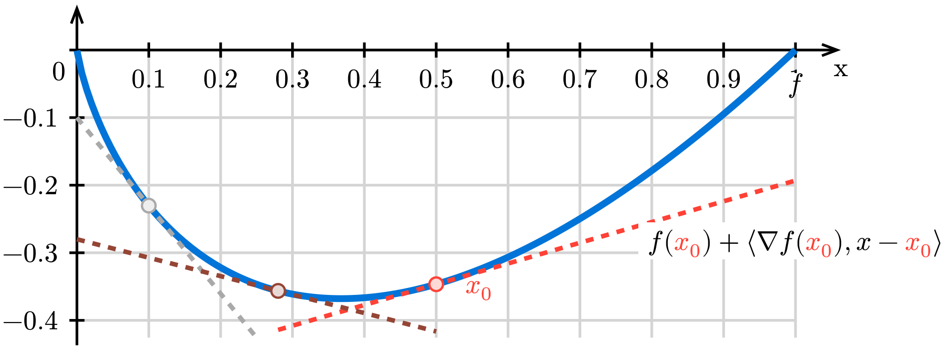
A function $f : \Omega \to \mathbb{R}$ is *convex* if, for any two points $x, y \in \Omega$ and $t \in [0, 1]$,

$$f\big((1-t)\cdot x + t \cdot y\big) \le (1-t)\cdot f(x) + t \cdot f(y).$$



L4.1.1 Convexity implies bounding by linearization

Assuming that f is not only convex but also differentiable, a very important property of convex functions is that they lie above their linearization at any point.



This follows directly from the definition, as we show next.

Theorem L4.1. Let $f : \Omega \to \mathbb{R}$ be a convex and differentiable function defined on a convex domain Ω. Then, at all $x \in \Omega$,

$$f(y) \ge \underbrace{f(x) + \langle \nabla f(x), y - x \rangle}_{\text{linearization of } f \text{ around } x} \qquad \forall y \in \Omega.$$

Proof. Pick any $x, y \in \Omega$. By definition of convexity, we have

$$f\big(x + t\cdot(y - x)\big) \le f(x) + t \cdot (f(y) - f(x)) \qquad \forall t \in [0, 1].$$

Moving the $f(x)$ from the right-hand side to the left-hand side, and dividing by t, we therefore get

$$\frac{f\big(x + t \cdot (y - x)\big) - f(x)}{t} \le f(y) - f(x) \qquad \forall t \in (0, 1].$$

Taking a limit as $t \downarrow 0$ and recognizing a directional derivative at x along direction $y - x$ on the left-hand side, we conclude that

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

Rearranging yields the result. □

L4.1.2 Sufficiency of first-order optimality conditions

The above result also immediately shows the *sufficiency* of first-order optimality conditions.

Theorem L4.2. Let $\Omega \subseteq \mathbb{R}^n$ be convex and $f : \Omega \to \mathbb{R}$ be a convex differentiable function. Then,

$$-\nabla f(x) \in \mathcal{N}_\Omega(x) \quad \iff \quad x \text{ is a minimizer of } f \text{ on } \Omega$$

Proof. We already know from Lecture 2 that $-\nabla f(x) \in \mathcal{N}_\Omega(x)$ is necessary for optimality. So, we just need to show sufficiency. Specifically, we need to show that if $\langle \nabla f(x), y - x\rangle \geq 0$ for all $y \in \Omega$, then surely $f(y) \geq f(x)$ for all $y \in \Omega$. This follows immediately from Theorem L4.1. □

L4.2 Equivalent definitions of convexity

Theorem L4.3. Let $\Omega \subseteq \mathbb{R}^n$ be a convex set, and $f : \Omega \to \mathbb{R}$ be a function. The following are equivalent definitions of convexity for f:

(1) $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$ for all $x, y \in \Omega, t \in [0, 1]$. — Most general

(2) [If f is differentiable] $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle$ for all $x, y \in \Omega$. — Most often used

(3) [If f is differentiable] $\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq 0$ for all $x, y \in \Omega$.

(4) [If f is twice differentiable and Ω is open] $\nabla^2 f(x) \succeq 0$ for all $x \in \Omega$. — Often easiest to check

The third criterion of Theorem L4.3 is usually the easiest to check in practice.

Example L4.1. For example, from that criterion it follows immediately that these functions are convex:

- $f(x) = a^\top x + b$ for any $a \in \mathbb{R}^n, b \in \mathbb{R}$;
- $f(x) = x^\top A x$ for any $A \succeq 0$, including $f(x) = \|x\|_2^2$;
- the *negative entropy* function $f(x) = \sum_{i=1}^n x_i \log x_i$ defined for $x_i > 0$;
- the function $f(x) = -\sum_{i=1}^n \log x_i$ defined for $x_i > 0$;
- the function $f(x) = \log(1 + e^x)$.

Remark L4.1. Condition (3) is also known as the *monotonicity* of the gradient ∇f. In dimension $n = 1$, the condition is equivalent to the statement that the derivative f' is *nondecreasing*.

Proof of Theorem L4.3. We have already seen how (1) $\Longrightarrow$ (2) in Theorem L4.1. To conclude the proof, we will show that under differentiability (3) $\Longleftrightarrow$ (2) $\Longrightarrow$ (1), and that under twice differentiability and openness of Ω, (3) $\Longleftrightarrow$ (4). We break the proof into separate steps.

▸ **Proof that (2) $\Longrightarrow$ (1).**

[*Intuition:* We sum the linear lower bounds centered in the point $z := t \cdot x + (1-t) \cdot y$ and looking in the directions $x - z$ and $y - z$.]

Pick any $x, y \in \Omega$ and $t \in (0, 1)$, and consider the point

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

From the linearization bound (2) for the choices $(x, y) = (z, x), (z, y)$, we know that

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

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

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

where the equality follows since by definition $z = t \cdot x + (1-t) \cdot y$. Rearranging, we have (1).

▸ **Proof that (2) $\Longrightarrow$ (3).**

[*Intuition:* The idea here is to write condition (2) for the pair (x, y) and for the symmetric pair (y, x). Summing the inequalities leads to the statement.]

Pick any two $x, y \in \Omega$. From (2), we can write

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

Summing the inequalities, we therefore conclude that

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

which is the statement.

▸ **Proof that (3) $\Longrightarrow$ (4).**

[*Intuition:* Condition (4) uses a Hessian matrix (*i.e.*, second derivative), but (3) only contains a difference of gradients. Unsurprisingly, the idea is to consider (3) for two close-by points and take a limit to extract an additional derivative.]

Pick any $x, y \in \Omega$, and define the point $x_t := x + t \cdot (y - x)$. Using (3) we have

$$0 \leq \langle \nabla f(x_t) - \nabla f(x), x_t - x \rangle = t \cdot \langle \nabla f(x_t) - \nabla f(x), y - x \rangle.$$

Rearranging and dividing by t^2, we have

$$\frac{\left\langle \nabla f(x + t \cdot (y - x)) - \nabla f(x), y - x \right\rangle}{t} \geq 0.$$

Taking the limit as $t \downarrow 0$, we therefore have

$$\left\langle (y - x), \nabla^2 f(x)(y - x) \right\rangle \geq 0.$$

Since Ω is open by hypothesis, the direction of $y - x$ is arbitrary, and therefore we must have $\nabla^2 f(x) \succeq 0$, as we wanted to show.

▸ **Proof that (4) $\implies$ (3).**

[*Intuition:* To go from (3) to (4) we took a derivative in the direction $y - x$. To go back, we take an integral on the line $y - x$ instead.]

By hypothesis, for any $x, y \in \Omega$ and $\tau \in [0, 1]$,

$$0 \leq \left\langle y - x, \nabla^2 f(x + \tau \cdot (y - x)) \cdot (y - x) \right\rangle.$$

Hence, taking the integral,

$$\begin{aligned} 0 &\leq \int_0^1 \left\langle y - x, \nabla^2 f(x + t \cdot (y - x)) \cdot (y - x) \right\rangle \mathrm{d}t \\ &= \left\langle y - x, \int_0^1 \underbrace{\nabla^2 f(x + t \cdot (y - x)) \cdot (y - x)}_{= \frac{\mathrm{d}}{\mathrm{d}t} \nabla f(x + t \cdot (y - x))} \mathrm{d}t \right\rangle = \left\langle y - x, \nabla f(y) - \nabla f(x) \right\rangle. \end{aligned}$$

▸ **Proof that (3) $\implies$ (2).**

[*Intuition:* The idea here it to treat x as fixed, and integrate condition (3) on the line from x to y.]

Pick any $x, y \in \Omega$, and define the point $x_t := x + t \cdot (y - x)$ for $t \geq 0$. Using condition (3) we have

$$0 \leq \langle \nabla f(x_t) - \nabla f(x), x_t - x \rangle = t \cdot \langle \nabla f(x_t) - \nabla f(x), y - x \rangle,$$

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

Letting t range from 0 to 1 and integrating,

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

Rearranging yields $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$, which is (2). □

L4.3 Convexity-preserving operations

In addition to the criteria above, one can also recognize convex functions when they are obtained from simpler convex functions combined via *convexity-preserving operations*, such as the following.

Theorem L4.4. The following operations preserve convexity:

- Multiplication of a convex function $f(x)$ by a nonnegative scalar $c \geq 0$;
- Addition of two convex functions $f(x), g(x)$;
- Pointwise supremum of a collection J of convex functions $\{f_j(x) : j \in J\}$:

$$f_{\max}(x) := \max_{j \in J} f_j(x);$$

- Pre-composition $f(Ax + b)$ of a convex function f with an affine function $Ax + b$.
- Post-composition $g(f(x))$ of a convex function with an *increasing* convex function g;
- *Infimal convolution* $f \mathbin{\overset{\vee}{+}} g$ of two convex functions $f, g : \mathbb{R}^n \to \mathbb{R}$, defined as

$$(f \mathbin{\overset{\vee}{+}} g)(x) := \inf\{f(y) + g(x - y) : y \in \mathbb{R}^n\}.$$

In all the cases above, it is straightforward to verify the preservation of convexity starting from the definition of convexity given in Definition L4.1.

L4.4 Strict and strong convexity, and uniqueness of minimizers

Two stronger notions of convexity are known as *strict* and *strong* convexity, defined as follows.

Definition L4.2 (Strict and strong convexity)**.** Let $\Omega \subseteq \mathbb{R}^n$ be convex.

- A function $f : \Omega \to \mathbb{R}$ is *strictly convex* if, for any two distinct points $x, y \in \Omega$ and $t \in (0, 1)$,

$$f\Big((1-t) \cdot x + t \cdot y\Big) < (1-t) \cdot f(x) + t \cdot f(y).$$

- A function $f : \Omega \to \mathbb{R}$ is *strongly convex* with modulus $\mu > 0$ if the function

$$f(x) - \frac{\mu}{2}\|x\|_2^2$$

  is convex.

Note that strong convexity implies strict convexity, and strict convexity implies convexity. Neither of the reverse implications holds.

Exercise L4.1. What are equivalent characterizations of strict and strong convexity, in the spirit of Theorem L4.3? A word of caution: the condition $\nabla^2 f(x) \succ 0$ everywhere for Ω open and f twice differentiable is only *sufficient* for strict convexity but not *necessary* (Hint: consider the function x^4, which is strictly convex and yet has $f''(0) = 0$).

As we have seen above, convexity has the benefit of making first-order optimality conditions sufficient for optimality, and hence *equivalent* to optimality. Strict and strong convexity have the additional benefit of guaranteeing that minimizers, if they exist, must be *unique*.

Theorem L4.5. Let $\Omega \subseteq \mathbb{R}^n$ be convex, and $f : \Omega \to \mathbb{R}$ be a strictly convex function. Then, f has at most one minimizer.

Proof. Suppose that $x, y \in \Omega$ are both minimizers of f. This means that they must both attain the minimum value of f, *i.e.*, $f(x) = f(y) = f_*$. But then, for any $t \in (0, 1)$, we have

$$f\Big(\big(1-t\big) \cdot x + t \cdot y\Big) < (1-t) \cdot f(x) + t \cdot f(y) = (1-t)f_* + tf_* = f_*.$$

This shows that the point $(1-t) \cdot x + t \cdot y \in \Omega$ attains a lower value of f than f_*, which is absurd. □

Corollary L4.1. Since the function $\frac{1}{2}\|x - y\|_2^2$ is strongly convex, and hence strictly convex, it follows that any projection onto a convex set, if it exists, is unique.

Further readings

If you want to read more about convex functions, the following resources all contain an excellent treatment.

[HL01] Hiriart-Urruty, J.-B., & Lemaréchal, C. (2001). *Fundamentals of Convex Analysis*. Springer. `https://link.springer.com/book/10.1007/978-3-642-56468-0`

[Nes18] Nesterov, Y. (2018). *Lectures on Convex Optimization*. Springer International Publishing. `https://link.springer.com/book/10.1007/978-3-319-91578-4`

[BV04] Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press. `https://web.stanford.edu/~boyd/cvxbook/`

Changelog

- Feb 13, 2025: Adjusted intuition boxes in the proof of Thm L4.3.
- Feb 24, 2025: Disambiguated pre-composition in Theorem L4.4 (Thanks https://piazza.com/class/m6lg9aspoutda/post/45)
- Mar 6, 2025: Fixed a typo in Exercise L4.1 (thanks Kai Hung! https://piazza.com/class/m6lg9aspoutda/post/m7y7mz0k3ou3bk)
- Mar 6, 2025: Fixed a typo (thanks Khizer Shahid!)