

Lecture 3

The special case of convex functions

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

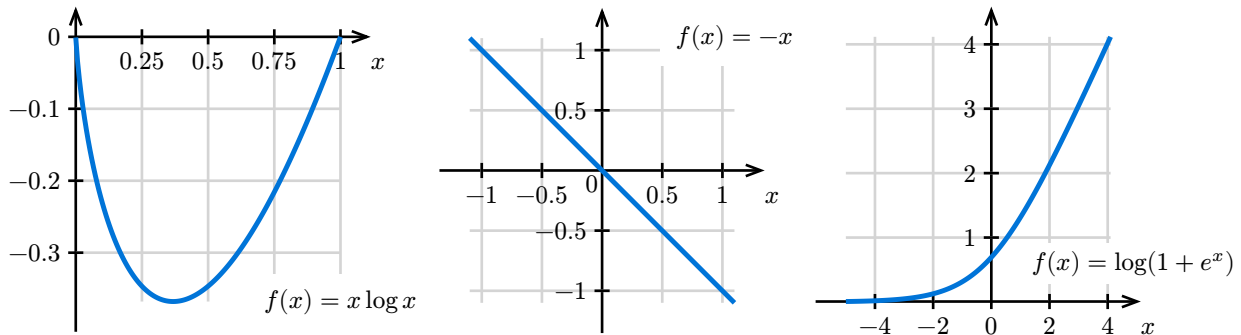
In the previous lecture, we have seen how any solution x to a nonlinear optimization problem defined on a convex feasible set $\Omega \subseteq \mathbb{R}^n$ must necessarily satisfy the first-order optimality condition

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

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.

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:



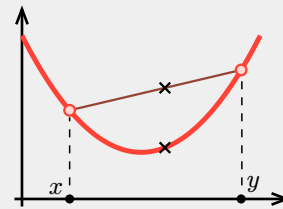
In particular, due to their curvature, local optima of these functions are also global optima, and the first-order optimality condition completely characterizes optimal points.

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.

Definition 1.1 (Convex function). Let $\Omega \subseteq \mathbb{R}^n$ be convex.

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

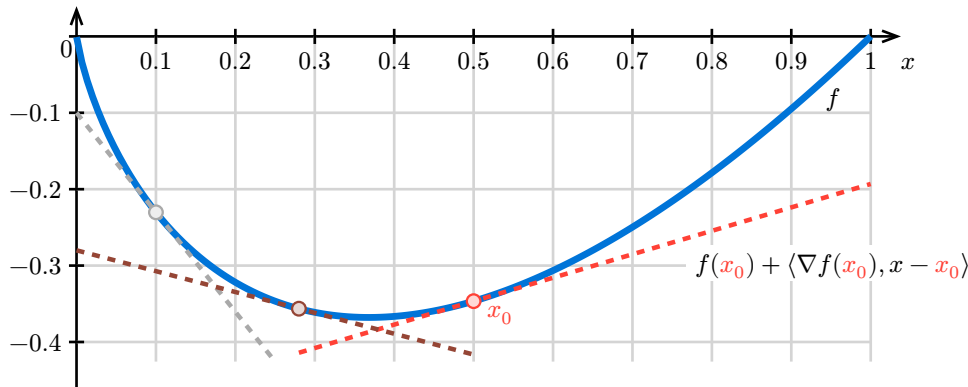
$$f((1-t) \cdot x + t \cdot y) \leq (1-t) \cdot f(x) + t \cdot f(y).$$



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

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

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

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

$$f(x + t \cdot (y - x)) \leq f(x) + t \cdot (f(y) - f(x)) \quad \forall t \in [0, 1].$$

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

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

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

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

Rearranging yields the result. \square

1.2 Sufficiency of first-order optimality conditions

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

Theorem 1.2. Let $\Omega \subseteq \mathbb{R}^n$ be convex and $f : \Omega \rightarrow \mathbb{R}$ be a convex differentiable function. Then,

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

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 1.1. \square

2 Equivalent definitions of convexity

Theorem 2.1. Let $\Omega \subseteq \mathbb{R}^n$ be a convex set, and $f : \Omega \rightarrow \mathbb{R}$ be a function. The following are equivalent definitions of convexity:

1. $f((1-t) \cdot x + t \cdot y) \leq (1-t) \cdot f(x) + t \cdot f(y)$ 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 twice differentiable and Ω is open) $\nabla^2 f(x) \succcurlyeq 0$ for all $x \in \Omega$. ↓ Often easiest to check

Proof. We have already seen how (1) \implies (2) in Theorem 1.1. To conclude the proof, it suffices to show that under differentiability (2) \implies (1), and that under twice differentiability and openness of Ω (2) \iff (3). We break the proof into separate steps.

- *Proof that (2) \implies (1).* Pick any $a, b \in \Omega$ and $t \in (0, 1)$, and consider the point

$$\Omega \ni z := t \cdot a + (1-t) \cdot b.$$

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

$$\begin{aligned} f(a) &\geq f(z) + \langle \nabla f(z), a - z \rangle, \\ f(b) &\geq f(z) + \langle \nabla f(z), b - z \rangle. \end{aligned}$$

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

$$\begin{aligned} t \cdot f(a) + (1-t) \cdot f(b) &\geq f(z) + \langle \nabla f(z), t \cdot a + (1-t) \cdot b - z \rangle \\ &= f(z), \end{aligned}$$

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

- *Proof that (2) \implies (3).* Pick any two $x, y \in \Omega$, and $t \in (0, 1]$. Define $x_t := x + t \cdot (y - x)$; note that $x_t \in \Omega$ by convexity of Ω . From (2), we can write

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

Summing the inequalities, we therefore conclude that

$$\begin{aligned} 0 &\geq \langle \nabla f(x) - \nabla f(x_t), x_t - x \rangle \\ &= \langle \nabla f(x) - \nabla f(x_t), t \cdot (y - x) \rangle \\ &= t \cdot \langle \nabla f(x) - \nabla f(x_t), y - x \rangle. \end{aligned}$$

Rearranging and dividing by t^2 , we have

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

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

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

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

- *Proof that (3) \implies (2).* By hypothesis, for any $x, y \in \Omega$ and $\tau \in [0, 1]$,

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

Hence, taking a double integral,

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

Rearranging yields (2). □

3 Examples of convex functions

The third criterion of Theorem 2.1 is usually the easiest to check in practice. For example, from that criterion it follows immediately that these functions are convex:

- $f(x) = a^\top x + b$ for any $a, b \in \mathbb{R}^n$;
- $f(x) = x^\top A x$ for any $A \succcurlyeq 0$, 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 $f(x) = \log(1 + e^x)$.

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 3.1. 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 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* $f \downarrow g$ of two convex functions $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$, defined as

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

In all the cases above, it is straightforward to verify the preservation of convexity starting from the definition of convexity given in Definition 1.1.

Further readings

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

- [BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004. [Online]. Available: <https://web.stanford.edu/~boyd/cvxbook/>
- [HL01] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer, 2001. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-642-56468-0>
- [Nes18] Y. Nesterov, *Lectures on Convex Optimization*. Springer International Publishing, 2018. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-319-91578-4>