

A first look at convex analysis

Crash course in convexity to have just enough material to study the gradient descent.

Samuel Vaiter

2024-02-07

Table of contents

Convex set	2
Convex functions	3
Definition	3
Local and global minima of a convex function	4
Differentiable properties of a convex function	5
Strongly convex functions	6
Definition	6
Differentiability, strong convexity and quadratic lower bound	6
Lipschitz continuity of the gradient	7
Definition	7
Quadratic upper bound	7
Co-coercivity of the gradient of a L -smooth function	9

DEFINITIONS.

1. There are in a plane certain terminated bent lines (*καμπύλαι γραμμαὶ πεπερασμέναι*)†, which either lie wholly on the same side of the straight lines joining their extremities, or have no part of them on the other side.

2. I apply the term **concave in the same direction** to a line such that, if any two points on it are taken, either all the straight lines connecting the points fall on the same side of the line, or some fall on one and the same side while others fall on the line itself, but none on the other side.

Figure 1: Definition of concavity by Archimede

Despite the fact that many problems encountered in statistics and machine learning are not convex, convexity is an important to study in order to understand what is the generic behaviour

of an optimization procedure.

Convex set

A set $C \subseteq \mathcal{X}$ is convex if every segment of C is contained in C . It can be formalized as the following definition and illustrated in TODO.

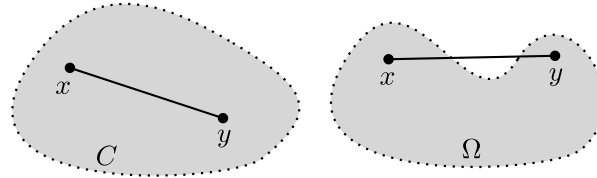


Figure 2: C is a convex set, Ω is not.

Definition 0.1 (Convex set). A set $C \subseteq \mathcal{X}$ is **convex** if

$$\forall(x, y) \in C^2, \forall \lambda \in [0, 1], \quad \lambda x + (1 - \lambda)y \in C.$$

One can observe that it is possible to replace $\lambda \in [0, 1]$ by $\lambda \in (0, 1)$ without any difference (prove it!) In most of these lecture notes, the set \mathcal{X} will be the Euclidean space \mathbb{R}^p , but remark that this notion only requires that \mathcal{X} is a vector space over the real numbers \mathbb{R} , possibly of infinite dimension.

Example 0.1 (Basic convex sets).

- The convex sets of \mathbb{R} are *exactly* the intervals of \mathbb{R} .
- Any affine hyperplane $H = \{x \in \mathcal{X} : \varphi(x) = \alpha\}$ where \mathcal{X} is a real vector space, φ a linear functional and $\alpha \in \mathbb{R}$ is convex.
- Any half-space $H_- = \{x \in \mathcal{X} : \varphi(x) \leq \alpha\}$ is also convex.
- The unit simplex of \mathbb{R}^d defined as

$$\Delta^{d-1} = \left\{ x \in \mathbb{R}^d : x_i \geq 0 \text{ and } \sum_i x_i = 1 \right\}$$

is convex.

- The nonnegative orthant $\mathbb{R}_{\geq 0}^d = \{x \in \mathbb{R}^d : x_i \geq 0\}$ is convex.

The following proposition reviews some important algebraic properties of convex sets.

Proposition 0.1 (Operations on convex sets).

- Arbitrary intersection. Let $(C_i)_{i \in I}$ convex sets where I is a set. Then $\bigcap_{i \in I} C_i$ is convex.
- Cartesian product. Let C_1, \dots, C_n be sets of $\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_n}$. Then $C_1 \times \dots \times C_n$ is convex if, and only if, every C_i is convex.
- Sum. Let C_1, C_2 two convex sets. Then the (Minkowski) sum $C_1 + C_2$ is convex.
- Affine map. Let $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be an affine map and $C \subseteq \mathbb{R}^n$ a convex set. Then the image $A(C)$ is convex.

Convex functions

Definition

The intuitive definition of a convex function is a function such that “the line joining $f(x)$ and $f(y)$ lies above the graph of f between x and y ”. Formally, this definition makes sense using the definition of the epigraph of f .

Definition 0.2 (Epigraph). Let $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ and $f \neq +\infty$. The **epigraph** of f is the subset of $\mathbb{R}^n \times \mathbb{R}$ defined by

$$\text{epi } f := \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : t \geq f(x)\}.$$

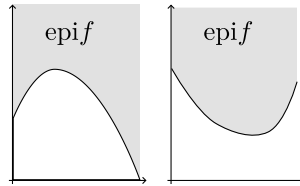


Figure 3: (Left) Epigraph of a concave function. (Right) Epigraph of a convex function

Equipped with this notion, we can define convex functions using the definition of convex sets (Definition 0.1) applied to the epigraph (Definition 0.2).

Definition 0.3 (Convex function). A function $f : \mathbb{R}^p \rightarrow \bar{\mathbb{R}}$ such that $f \neq +\infty$ is **convex function** iff its epigraph $\text{epi } f$ is convex.

It is possible to make explicit this definition as follow:

$$\forall x, y \in \text{dom } f, \forall \lambda \in [0, 1], \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (1)$$

Following the convention introduced in (Hiriart-Urruty and Lemarechal 1996), we will denote the set of all convex functions of \mathbb{R}^p as $\text{Conv } \mathbb{R}^p$, and the set of closed – that is, with closed epigraph – convex functions of \mathbb{R}^p as $\overline{\text{Conv}} \mathbb{R}^p$.

Local and global minima of a convex function

For a (un)constrained minimization problem on a convex set $C \subseteq \text{dom } f$ with a function $f : C \rightarrow \mathbb{R}$,

$$\min_{x \in C} f(x), \tag{2}$$

we say that x^* is

- a local solution of (Equation 2), if there exists a neighborhood \mathcal{O} of x^* such that for all $x \in C \cap \mathcal{O}$, $f(x) \geq f(x^*)$;
- a (global) solution if for all $x \in C$, $f(x) \geq f(x^*)$.

A fundamental result is that, **when f is convex**, every local solution of (Equation 2) is global solution, and the set of (global) solutions is a convex set.

Theorem 0.1 (Local minima of convex functions are global). *Assume that f is a convex function and C is a convex **closed** set. Then,*

- *Any local solution of (Equation 2) is a global solution;*
- *The set of global solutions of (Equation 2) is convex.*

Proof. Part 1. By contradiction. Consider a local solution x^* and assume that x^* is not a global solution. In particular, there exists $z \in C$ such that $f(z) < f(x^*)$. By convexity,

$$f((1-t)x^* + tz) \leq (1-t)f(x^*) + tf(z) < f(x^*) \quad \forall t \in [0, 1]$$

Given any open neighborhood \mathcal{O} of x^* , there exists $t > 0$ such that $(1-t)f(x^*) + tf(z) < f(x^*)$ (contradiction).

Part 2. Take two global solutions, and study the segment joining it: let x^* and z two solutions. We have $f(x^*) = f(z)$, and in turn,

$$f(x^*) \leq f((1-t)x^* + tz) \leq (1-t)f(x^*) + tf(z) = f(x^*),$$

where the first inequality comes from the equality of the function at x^* and z , and the second by convexity. \square

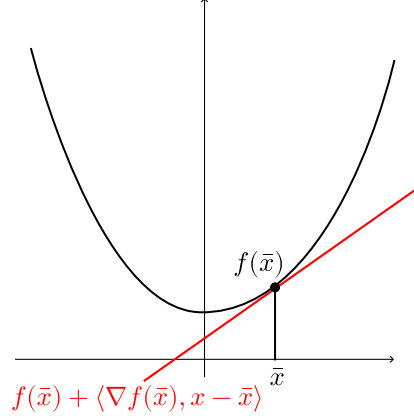


Figure 4: A differentiable convex function and its tangent. The curve $y = f(x)$ is always above its tangent in red.

Differentiable properties of a convex function

A differentiable function is convex if, and only if, it is above all its tangents, see Figure 4

Proposition 0.2. *Let f be a differentiable function on an open set $\Omega \subseteq \mathbb{R}^p$, and $C \subseteq \Omega$ a convex subset of Ω . Then f is convex on C if, and only if,*

$$\forall x, \bar{x} \in C, \quad f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle. \quad (3)$$

Proof. Assume f is convex on C . Let $x, \bar{x} \in C$ and $\lambda \in (0, 1)$. By Definition 0.3, we have

$$f(\lambda x + (1 - \lambda)\bar{x}) - f(\bar{x}) \leq \lambda(f(x) - f(\bar{x})).$$

Dividing by λ , and letting λ goes to 0, we obtain Equation 3.

Reciprocally, let $x, y \in C$, $\lambda \in (0, 1)$ and define $z = \alpha x + (1 - \alpha)y \in C$. Applying Equation 3 twice, we have

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

Using a convex combination of these two lines, we get

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

which is exactly Equation 1. □

A similar criterion can be derived for strictly convex function. Remark that Theorem 0.1 is a direct consequence of Proposition 0.2 for the class of differentiable convex functions: a local minima x^* satisfies $\nabla f(x^*) = 0$ and thus (Equation 3) gives $f(x) \geq f(x^*)$ for all $x \in C$.

Strongly convex functions

Definition

A strongly convex function satisfies a stronger statement of the Jensen's inequality.

Definition 0.4 (Strongly convex function). A function $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is *strongly convex* of modulus $\mu > 0$ if $\text{dom } f$ is convex and

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{1}{2}\mu\lambda(1 - \lambda)\|x - y\|^2,$$

for all $x, y \in \text{dom } f$ and $\lambda \in [0, 1]$.

Another way to define strongly convex function is to say that f is μ -strongly convex if the function

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

is a convex function.

A strongly convex function defined on the whole space \mathbb{R}^d has a **unique** minimizer.

Differentiability, strong convexity and quadratic lower bound

A function f is strongly convex if, and only if, it is *strictly above* its tangent, and you can “fit” a quadratic form between the tangent and the function itself. This is a key property to defined lower bounds on f .

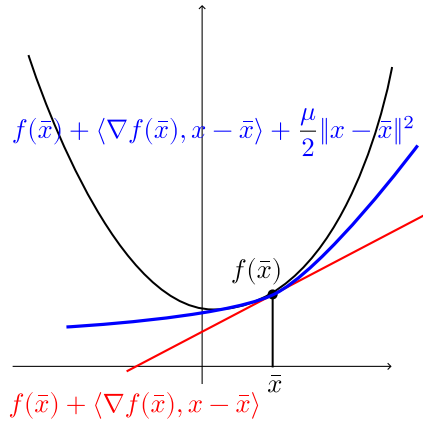


Figure 5: Strong convexity and differentiability. There is space for a quadratic form between the curve $y = f(x)$ and the tangent.

Proposition 0.3. *Let f be a differentiable and μ -strongly convex function. Then,*

$$\forall x, \bar{x} \in \text{dom } f, \quad f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{\mu}{2} \|x - \bar{x}\|^2.$$

Note that if f is C^2 , f is μ -strongly convex if, and only if, its Hessian $\nabla^2 f \geq \mu \text{Id}$, i.e., is definite positive.

An easy consequence is that it is possible to establish a lower bound on the values based on a lower bound on the distance between points:

$$f(x) \geq f(x^*) + \frac{\mu}{2} \|x - x^*\|^2, \quad (4)$$

where x^* is the minimizer of f .

Lipschitz continuity of the gradient

Definition

Definition 0.5 (L -smoothness). A differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be L -smooth if ∇f is L -Lipschitz, i.e.,

$$\forall x, y \in \text{dom } f, \quad \|\nabla f(x) - \nabla f(y)\|_* \leq L \|x - y\|.$$

Here $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Quadratic upper bound

L -smoothness is important in optimization because it ensures a quadratic upper bound of the function.

Proposition 0.4. *Let f a L -smooth function. Then, for all $x, y \in \text{dom } f$,*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L \|x - y\|^2. \quad (5)$$

Moreover, if $\text{dom } f$ is convex, then Equation 5 is equivalent to

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|x - y\|^2, \quad \forall x, y \in \text{dom } f. \quad (6)$$

Proof. The fact that L -smoothness implies (Equation 5) is a direct application of the (generalized) Cauchy-Schwarz inequality.

For the equivalence of (Equation 5) and (Equation 6), let's prove the two direction. Given $x, y \in \text{dom} f$, define the real function $g(t) = f(x + t(y - x))$.

Assuming Equation 5 holds, then

$$g'(t) - g'(0) = \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle \leq tL\|x - y\|^2.$$

Now,

$$\begin{aligned} f(y) = g(1) &= g(0) + \int_0^1 g'(t) \leq g(0) + g'(0) + \frac{L}{2}\|x - y\|^2 \\ &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|x - y\|^2. \end{aligned}$$

Reciprocally, if (Equation 5) holds, it is sufficient to apply it twice (to x, y and y, x) and combine the two equations. \square

This result has a strong implication on the minima of f .

Proposition 0.5. *Let f be a L -smooth function with full domain, and assume x^* is a minimizer of f . Then,*

$$\frac{1}{2L} \|\nabla f(z)\|_*^2 \leq f(z) - f(x^*) \leq \frac{L}{2} \|z - x^*\|^2 \quad \forall z \in \mathbb{R}^n.$$

Proof. The two side are proved separately.

rhs: Apply Equation 5 applied to $((x^*, z))$.

lhs: Minimize the quadratic upper bound for $x = z$:

$$\begin{aligned} f(x^*) = \min_{y \in \mathbb{R}^n} f(y) &\leq \inf_y \left(f(z) + \langle \nabla f(z), y - z \rangle + \frac{L}{2} \|y - z\|^2 \right) \\ &= \inf_{\|v\|=1} \inf_t \left(f(z) + t \langle \nabla f(z), v \rangle + \frac{Lt^2}{2} \right), \end{aligned}$$

where the last inequality is obtained by the change of variable $v = \frac{w}{\|w\|}$, $t = \|w\|$ and $w = y - z$. Factorizing the trinom in t leads to:

$$f(z) + t \langle \nabla f(z), v \rangle + \frac{Lt^2}{2} = \frac{L}{2} \left(t + \frac{1}{L} \langle \nabla f(z), v \rangle \right)^2 - \frac{1}{2L} \langle \nabla f(z), v \rangle^2 + f(z).$$

Hence,

$$\min_{y \in \mathbb{R}^n} f(y) \leq \inf_{\|v\|=1} \left(f(z) - \frac{1}{2L} \langle \nabla f(z), v \rangle^2 \right).$$

Minimizing the infimum boils down to maximizing $\langle \nabla f(z), v \rangle^2$, that is obtained with $v = \nabla f(z)$, hence the result. \square

Co-coercivity of the gradient of a L -smooth function

Convex L -smooth functions are co-coercive. This result is sometimes coined Baillon–Haddad theorem (Baillon and Haddad 1977), but one shall note that the original contribution is much more general than its application to the gradient.

Proposition 0.6. *Let f be a full-domain convex L -smooth function. Then, ∇f is $1/L$ -co-coercive, i.e.,*

$$\forall x, y \in \mathbb{R}^n, \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|_*^2.$$

Proof. Let $f_x(z) = f(z) - \langle \nabla f(x), z \rangle$ and $f_y(z) = f(z) - \langle \nabla f(y), z \rangle$. The two are convex.

Since f is L -smooth, f_x and f_y are also L -smooth.

Canceling the gradient of f_x shows that $z = x$ minimizes f_x . Hence, using the quadratic upper bound, we have

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), x - y \rangle &= f_x(y) - f_x(x) \\ &\geq \frac{1}{2L} \|\nabla f_x(y)\|_*^2 \\ &= \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|_*^2. \end{aligned}$$

Doing exactly the same thing for f_y leads a similar bound, and combining the two gives the result. \square

It turns out that smoothness, co-coercivity and upper quadratic boundness are equivalent properties for full-domain convex function.

Proposition 0.7. *Let f be a differentiable convex function with $\text{dom } f = \mathbb{R}^n$. Then the three following properties are equivalent:*

1. f is L -smooth.
2. The quadratic upper bound Equation 6 holds.
3. ∇f is $1/L$ -co-coercive.

Proof. The only one to prove is 2. \implies 3. (by Cauchy-Schwarz). \square

For the special case of the ℓ^2 norm, we have:

Proposition 0.8. *Let f be a differentiable convex function with $\text{dom } f = \mathbb{R}^n$, and L -smooth with respect to the Euclidean norm. Then,*

1. $L\text{Id} - \nabla f$ is monotone, i.e.

$$\forall x, y \in \mathbb{R}^n, \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L\|x - y\|_2^2.$$

2. $\frac{L}{2}\|x\|_2^2 - f(x)$ is convex.

3. If f is twice differentiable, then,

$$\lambda_{\max}(\nabla^2 f(x)) \leq L, \quad \forall x \in \mathbb{R}^n.$$

When f is both L -smooth and μ -strongly convex, we have the following “strengthened” co-coercivity.

Proposition 0.9. *Let f be a L -smooth and μ -strongly convex function. For any $x, y \in \mathbb{R}^n$, we have*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\mu L}{\mu + L}\|x - y\|^2 + \frac{1}{\mu + L}\|\nabla f(x) - \nabla f(y)\|^2.$$

Proof. Consider the convex function $\phi(x) = f(x) - \frac{\mu}{2}\|x\|^2$. Its gradient reads $\nabla\phi(x) = \nabla f(x) - \mu x$. Using the L -smoothness of f , we have that

$$\begin{aligned} \langle \nabla\phi(x) - \nabla\phi(y), x - y \rangle &= \langle \nabla f(x) - \nabla f(y), x - y \rangle - \mu\langle x - y, x - y \rangle \\ &\leq (L - \mu)\|x - y\|^2. \end{aligned}$$

Hence, ϕ is $(L - \mu)$ -smooth.

If $L = \mu$, then nothing is left to prove since f is μ -strongly convex. Otherwise, the $(L - \mu)$ -co-coercivity of ϕ (Proposition 0.6) gives us

$$\langle \nabla\phi(x) - \nabla\phi(y), x - y \rangle \geq \frac{1}{L - \mu}\|\nabla\phi(x) - \nabla\phi(y)\|^2.$$

Expliciting the expression of $\nabla\phi(x)$ allows us to conclude. □

Baillon, Jean-Bernard, and Georges Haddad. 1977. “Quelques Propriétés Des Opérateurs Angle-Bornés Etn-Cycliquement Monotones.” *Israel J. Math.* 26 (2): 137–50. <https://doi.org/10.1007/BF03007664>.

Hiriart-Urruty, Jean-Baptiste, and Claude Lemarechal. 1996. *Convex Analysis and Minimization Algorithms I: Fundamentals*. Corrected edition. Berlin: Springer.