2

Convexity

As we have seen, convexity is a key property for the solution of optimization problems. In this chapter, we will define the basic concepts and properties of convex objects, and derive the first results concerning mathematical optimization.

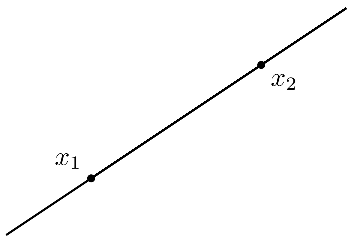
2.1 Affine Combinations

Definition 2.1. *Given two points $x_1, x_2 \in \mathbb{R}^n$, we call* affine combination *the point obtained as $y = \theta_1 x_1 + \theta_2 x_2$ for any two multipliers $\theta_1, \theta_2 \in \mathbb{R}$ satisfying the condition $\theta_1 + \theta_2 = 1$.*

The geometric interpretation of an affine combination is quite simple. By substituting $\theta_2 = 1 - \theta_1$ in the equation defining y, we obtain

$$y = \theta_1 x_1 + \theta_2 x_2 = \theta_1 x_1 + (1 - \theta_1) x_2 = x_2 + \theta_1 (x_1 - x_2)$$

so the affine combinations of x_1 and x_2 lie on the line passing through them.



In general, given m points $x_1, \ldots, x_m \in \mathbb{R}^n$ and m multipliers $\theta_1, \ldots, \theta_m \in \mathbb{R}$ such that $\sum_{i=1}^m \theta_i = 1$, we call $y = \sum_{i=1}^m \theta_i x_i$ their affine combination.

Definition 2.2. *A set is* affine *iff it contains any affine combination of its points.*

This is a very common approach in mathematics that will occur over and over in this course: given an operation A (on some mathematical objects), we say that a set has the property if it is closed w.r.t. said operation.

What is the structure of an affine set? The geometrical intuition with the affine combination of two points suggests that an affine set is related to linear spaces (i.e., vector subspaces), but without the restriction of having to pass through the origin. This is explained and formalized by the following proposition:

Proposition 2.1. *Any affine set C can be expressed as $C = V + x_0$, where $x_0 \in C$ and V is a vector subspace.*

Any $x_0 \in C$ will do.

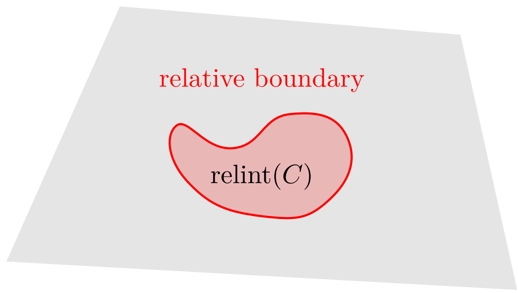


Figure 2.1: Relative interior of a non full-dimensional set.

Since an affine set is just a translation of a linear space, it is also very natural to define the dimension of the affine set as the dimension of the corresponding subspace, i.e., $\dim(C) = \dim(V)$.

This is also a recurring construct.

Definition 2.3. *Given an arbitrary set C, we define its* affine hull *as the set $\mathit{aff}(C)$ of all points that can be obtained as affine combinations of elements of C.*

It is easy to show that the affine hull of a set C is the smallest affine set containing C. Now that we can "attach" an affine set to arbitrary sets, we can also define the *affine dimension* of a set C as the dimension of its affine hull.

The concept of affine dimension is quite important in convex analysis, as it allows to define the very useful concept of *relative interior*.

At least for proofs...

We denote with $B(x, r)$ a ball centered at x with radius r.

Definition 2.4. *The relative interior of a set C is defined as the interior of C w.r.t its affine hull $\mathit{aff}(C)$. More formally:*

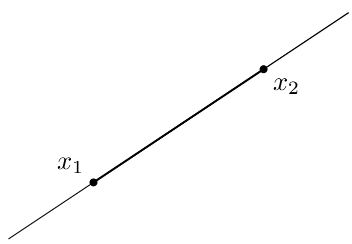
$$\mathit{relint}(C) = \{x \in C \mid B(x, r) \cap \mathit{aff}(C) \subseteq C \text{ for some } r > 0\}$$

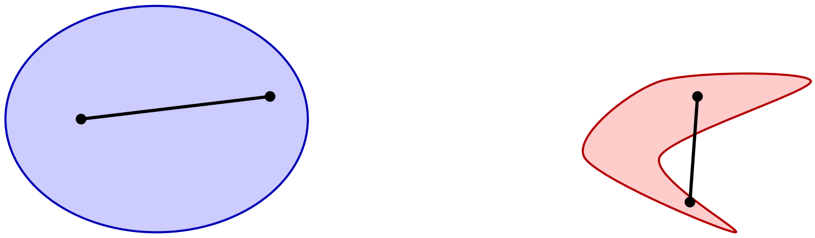
The relative interior of C is the most convenient definition of interior as it is "natural" even when the set is not full dimensional, i.e., $\dim(C) < n$: the regular interior would always be empty in this case. Analogously, we define the relative boundary as $\mathrm{relbd}(C) = \mathrm{cl}(C) \setminus \mathrm{relint}(C)$.

2.2 Convex Sets

Definition 2.5. *Given two points $x_1, x_2 \in \mathbb{R}^n$, we call* convex combination *the point obtained as $y = \theta_1 x_1 + \theta_2 x_2$ for any two* non-negative *multipliers $\theta_1, \theta_2 \in \mathbb{R}_+$ satisfying the condition $\theta_1 + \theta_2 = 1$.*

We notice that the difference w.r.t. an affine combination is just that the multipliers are non-negative: from the geometrical point of view, this means that a convex combination is restricted to lie between x_1 and x_2 or, in other words, that the set of convex combinations is no longer described by the line passing through x_1 and x_2, but rather the segment having x_1 and x_2 as endpoints.





(a) Convex set.

(b) Non-convex set.

In general, given m points $x_1, \dots, x_m \in \mathbb{R}^n$ and m non-negative multipliers $\theta_1, \dots, \theta_m \in \mathbb{R}_+$ such that $\sum_{i=1}^m \theta_i = 1$, we call $y = \sum_{i=1}^m \theta_i x_i$ their convex combination.

Definition 2.6. *A set is* convex *iff it contains any convex combination of its points.*

Definition 2.7. *Given an arbitrary set C, we define its* convex hull *as the set conv(C) of all points that can be obtained as convex combinations of elements of C.*

Again, it is easy to show that the convex hull of a set C is the smallest convex set containing C.

The following sets are easily shown to be convex: *Prove it!*

- $\emptyset$ and $\mathbb{R}^n$
- any affine set (clearly!)
- hyperplanes $\{x \mid a^\top x = b\}$ for any $a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$ *Is this affine?*
- halfspaces $\{x \mid a^\top x \leq b\}$ for any $a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$ *And this one?*
- Euclidian balls

  $$B(x_c, r) = \{x \mid \|x - x_c\| \leq r\}$$

  for any center $x_c \in \mathbb{R}^n$ and radius $r > 0$.
- ellipsoids

  $$\mathcal{E}(x_c, P, r) = \{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leq r\}$$

  for any symmetric positive definite matrix $P \succ 0$, center $x_c \in \mathbb{R}^n$, and radius $r > 0$.
- polyhedra $P = \{x \mid Ax \leq b, Cx = d\}$

2.2.1 *Calculus of convex sets*

Proving the convexity of a given set starting from the definition can sometimes be very tricky or tedious. A different approach consists in showing that the set can be obtained from a simpler one, already known to be convex, through a sequence of operations that preserve convexity. Here is a list of the most common convexity-preserving operators: *Like the ones in the previous list.*

1. *Intersection.* Given an arbitrary family $\mathcal{A}$ of convex sets, their intersection $C = \bigcap_{a \in \mathcal{A}} C_a$ is a convex set.

Example 2.1. *The positive semidefinite cone S_+^n can be expressed as:*

$$\bigcap_{z \neq 0} \{X \in S^n \mid z^\top X z \geq 0\}$$

2. *Affine functions.* Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be an affine function, i.e., $f(x) = Ax + b$ for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then if S is a convex set, then so are:

$$f(S) = \{f(x) \mid x \in S\}$$
$$f^{-1}(S) = \{x \mid f(x) \in S\}$$

   Simple examples include scaling and translations.

3. *Product.* Given $C_1 \subseteq \mathbb{R}^{n_1}$ and $C_2 \subseteq \mathbb{R}^{n_2}$ convex sets, their (cartesian) product $C_1 \times C_2 \subseteq \mathbb{R}^{n_1+n_2}$ is also convex.

4. *Linear combinations.* Given $M_1, \ldots, M_k$ convex sets in $\mathbb{R}^n$ and arbitrary multipliers $\lambda_1, \ldots, \lambda_k$, the set $\sum_{i=1}^k \lambda_i M_i$ is a convex set.

5. *Projection.* Let $S \in R^m \times \mathbb{R}^n$ be a convex set. Then its projection

$$T = \text{proj}_{\mathbb{R}^m}(S) = \{x \in \mathbb{R}^m \mid (x, y) \in S \text{ for some } y \in \mathbb{R}^n\}$$

   is a convex set.

2.2.2 *Topological properties*

For any subset $M \subseteq \mathbb{R}^n$ we have the natural relations:

$$\text{relint}(M) \subseteq M \subseteq \text{cl}(M)$$

However, the inclusions can be very non tight on pathological sets.

Example 2.2. *Consider as set M the set of all rational numbers in the interval $[0, 1]$. Here we have relint(M) = int(M) = $\emptyset$ and cl(M) = $[0, 1]$.*

A very intuitive explanation is that convex sets, by definition, cannot have holes.

For convex sets, things are much nicer. In particular, if M is a convex set, then:

- $\text{relint}(M)$, $\text{int}(M)$ and $\text{cl}(M)$ are all convex sets;
- $M \neq \emptyset \Rightarrow \text{relint}(M) \neq \emptyset$
- $\text{cl}(M) = \text{cl}(\text{relint}(M))$
- $\text{relint}(M) = \text{relint}(\text{cl}(M))$

A nice consequence of those facts is the following proposition:

Proposition 2.2. *If M is convex set, $x \in$ relint(M) and $y \in$ cl(M), then we have that the whole segment $[x, y) \subseteq$ relint(M).*

2.3 Separation Theorem

One of most fundamental properties of convex sets, which has far reaching consequences in optimization, is the so-called *separation* theorem:

Theorem 2.1. *Let C and D be non-empty convex sets that do not intersect ($C \cap D = \emptyset$). Then there exists a* separating hyperplane*, i.e., there exist $a \in \mathbb{R}^n, a \neq 0$ and $b \in \mathbb{R}$, such that:*

$$a^\top x \leq b \quad \forall x \in C$$
$$a^\top x \geq b \quad \forall x \in D$$

In order to prove this fundamental result, we need a few intermediate lemmas.

Lemma 2.1. *Let $A \subseteq \mathbb{R}^n$ be a non-empty set. Define the function $f : \mathbb{R}^n \to \mathbb{R}$:*

$$f(x) = d(x, A) = \inf_{y \in A} \|x - y\|$$

as the distance between a point x and the set A. Then f is continuous.

Proof. Let $x, y \in \mathbb{R}^n$ and $z \in A$. By definition of distance we have:

$$d(x, A) \leq d(x, z) \leq d(x, y) + d(y, z)$$

By taking the infimum w.r.t. $z \in A$ on both sides we get:

$$d(x, A) \leq d(x, y) + d(y, A)$$

By a symmetric argument, we can also derive the analogous:

$$d(y, A) \leq d(x, y) + d(x, A)$$

Now, combining the two we obtain:

$$\|d(x, A) - d(y, A)\| \leq d(x, y)$$

As $d(x, y) \to 0$, so does $\|f(x) - f(y)\|$, and thus f is continuous. □

Lemma 2.2. *Let $A \subseteq \mathbb{R}^n$ be a non-empty* closed *set, and let $y \notin A$. Then there exists $\overline{x} \in A$ at minimum distance, i.e.,*

$$d(y, A) = d(y, \overline{x}) \leq d(y, x) \quad \forall x \in A$$

Proof. Since $A \neq \emptyset$, there exists $\hat{x} \in A$ s.t. $d(y, A) \leq d(y, \hat{x})$. Now, let's define the set

$$A' = A \cap \{x \mid d(x, y) \leq d(y, \hat{x})\}$$

Note that by construction $d(y, A') = d(y, A)$. By definition A' is closed and bounded, while by the previous lemma $f(x) = d(x, A')$ is continuous: by the Weierstrass theorem it attains its minimum on A'. In other words, $\exists \overline{x} \in A' \subseteq A$ such that $d(y, \overline{x}) = d(y, A') = d(y, A)$. □

Lemma 2.3. *If C is a closed set and D is closed and bounded (i.e.,* compact*), and both are non-empty, then there exist $x \in C$ and $y \in D$ such that $d(x, y) = D(C, D)$.*

Proof. Let $f : D \to \mathbb{R}$ be defined as $f(z) = d(z, C)$. Since D is compact, and f is continuous, then again by Weierstrass there exists $y \in D$ such that $d(D, C) = d(y, C)$. By the previous lemma, since C is closed, there exists $x \in C$ such that:

$$d(x, y) = d(y, C) = d(C, D)$$

□

Note that all the assumptions in the lemma above are needed:

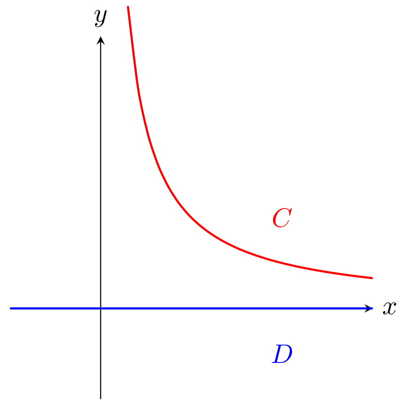


Figure 2.3: Two convex sets not satisfying the assumptions of Lemma 2.3.

- if any of the two sets is open, the claim is obviously false;
- if they are both closed, but none of them is compact, then again the claim is false. See Figure 2.3 for an example.

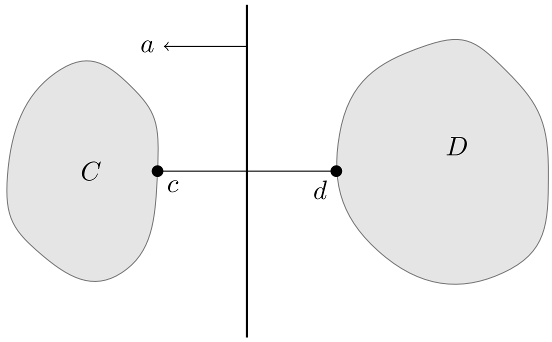
We are now ready to prove a special case of the separation theorem.

Proposition 2.3. *Let C and D be non-empty* closed *convex sets that do not intersect, with one one of them being* compact. *Then there exists a* separating hyperplane, *i.e., there exists $a \in \mathbb{R}^n, a \neq 0$ and $b \in \mathbb{R}$, such that:*

$$\begin{aligned} a^\top x &\le b \quad \forall x \in C \\ a^\top x &\ge b \quad \forall x \in D \end{aligned}$$

Proof. By lemma 2.3, there exist $c \in C$ and $d \in D$ such that $d(c,d) = d(C,D)$, and clearly $d(c,d) > 0$, as otherwise we would have $c = d$, contradicting the assumption than $C \cap D = \emptyset$. Now let's just define $a = d - c$ and $b = \frac{(d-c)^\top (d+c)}{2} = \frac{\|d\|^2 - \|c\|^2}{2}$. and the results follows A geometrical interpretation of the choice of a and b is given below.

Note that this is strict *separation. Also, given the points of minimum distance between the two sets, the proof is constructive.*



In details, let us show that $a^\top x \ge b \, \forall x \in D$ (a symmetric argument holds for the other side). Consider the function $f(x) = a^\top x - b$:

$$\begin{aligned} f(x) &= a^\top x - b \\ &= (d-c)^\top (x - \frac{d+c}{2}) \\ &= (d-c)^\top (x - d + d - \frac{d+c}{2}) \\ &= (d-c)^\top (x - d) + \frac{1}{2} \underbrace{\|d-c\|^2}_{>0} \end{aligned}$$

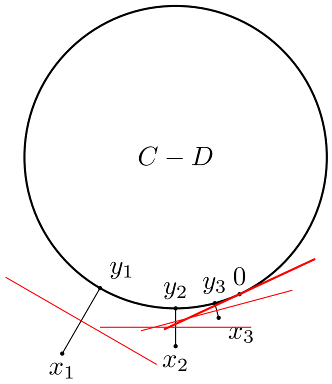


Figure 2.4: Geometric representation of construction for the general case.

We want to show that $f(x) \geq 0 \; \forall x \in D$. By contradiction, let us assume $\exists u \in D$ such that $f(u) < 0$. Since the last term in $f(u)$ is positive, this implies that $(d - c)^\top (u - d) < 0$. Notice that

$$\frac{d}{dt} \|d + t(u - d) - c\|^2 |_{t=0} = 2(d - c)^\top (u - d) < 0$$

So, for same small $t > 0$, we have $\|d + t(u - d) - c\| < \|d - c\|$, and this implies that $y = d + t(u - d)$ is closer to c than d. This is however a contradiction, because y is a convex combination and d and u, and thus belongs to the convex set D, so it cannot be closer to c than d. □

We are finally ready to prove the general case of the separation theorem.

Proof. Consider the set $C - D$: this is a linear combination of two convex sets, and thus convex itself. Also, clearly $0 \notin C - D$, as the sets do not intersect. Now, there are two cases:

1. $C - D$ is closed. Thus, we can separate by Proposition 2.3 $C - D$ from 0: let a be the corresponding hyperplane (clearly, we can pick $b = 0$).

$$a^\top (x - y) \leq 0 \quad \forall x \in C \; \forall y \in D$$

But then we can easily construct a separating hyperplane for C and D as:

$$a^\top x \leq \underbrace{\max_{x \in C} a^\top x \leq \min_{y \in D} a^\top y}_{b \text{ in here}} \leq a^\top y$$

2. $C - D$ is *not* closed. If $0 \notin \mathrm{cl}(C - D)$, then we can repeat the construction from the previous point and separate 0 from $\mathrm{cl}(C - D)$, and thus C from D. The last case to handle is thus if $0 \in \mathrm{cl}(C - D)$, i.e., $0 \in \mathrm{bd}(C - D)$. Consider a sequence of points $\{x_k\} \to 0$ converging to 0 from the outside of $\mathrm{cl}(C - D)$. Let then y_k be the corresponding points of minimum distance of x_k w.r.t. $\mathrm{cl}(C - D)$ (always well defined by Lemma 2.3). Each x_k can be separated by $\mathrm{cl}(C - D)$ by Proposition 2.3. See Figure 2.4 for a geometrical representation of the construction. Define the normalized separating hyperplane vector as

This sequence always exists as 0 is on the boundary

$$a_k = \frac{y_k - x_k}{\|y_k - x_k\|}$$

Because this is, because of the normalization, a bounded sequence of vectors in $\mathbb{R}^n$, then we have that, taking a subsequence if necessary,

And $\mathbb{R}^n$ is closed.

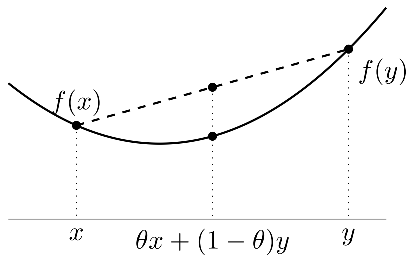


Figure 2.5: Jensen's inequality.

$\{a_k\} \to a$ for some a. Now, for every $z \in \text{cl}(C - D)$:

$$
\begin{aligned}
a^\top z &= \lim_{k \to +\infty} a_k^\top z && (\{a_k\} \to a) \\
&\leq \lim_{k \to +\infty} a_k^\top x_k && (a_k \text{ separating}) \\
&= 0 && (\{x_k\} \to 0)
\end{aligned}
$$

So a is the separating hyperplane we were looking for.

□

A nice corollary of the separation theorem is that, if a set C is non-empty and convex, then there exists a *supporting* hyperplane (a, b) for any point $x \in \text{bd}(C)$.

2.4 Convex Functions

This is called Jensen's inequality.

Definition 2.8. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain $domf$ is a convex set and for all $x, y \in domf$ and for all $0 \leq \theta \leq 1$ we have :*

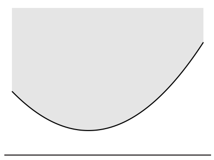
$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \tag{1}$$

A graphical representation of the convexity condition is depicted in Figure 2.5. Intuitively, a convex function always lies below the chord connecting any two points along the function curve.

It is called *strictly* convex if the inequality (1) is strict whenever $x \neq y$ and $0 < \theta < 1$. Note that convexity is, in some sense, a one-dimensional property: indeed it can be shown that a function f is convex iff all of its restrictions to a line that intersects the domain are convex.

There is a strong relation between convex sets and convex functions. An equivalent definition of convex function could indeed be: a function f is convex iff its epigraph

Intuitively, the set of points lying above the function curve.



$$\text{epi}f = \{(x, t) \in \mathbb{R}^{n+1} \mid x \in \text{dom}f,\ t \geq f(x)\}$$

is a convex set.

Definition 2.9. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is concave if $-f$ is convex.*

Note that any affine function is both convex and concave.

As with convex sets, convexity implicitly makes sure that convex functions can never be too pathological. For example, it is easy to show that a convex function must be continuous in the relative interior of its domain, so any discontinuity is only allowed on the boundary.

We can give a list of basic functions that are easily shown to be convex by just using the definition.

- e^{ax} is convex on $\mathbb{R}$ for any $a \in \mathbb{R}$;
- x^a is convex on $\mathbb{R}_{++}$ if $a \geq 1$ and $a \leq 0$, and concave if $0 \leq a \leq 1$;
- $|x|^p$ is convex on $\mathbb{R}$ for $p \geq 1$;
- $\log x$ is concave on $\mathbb{R}_{++}$;
- any norm on $\mathbb{R}^n$ is convex;
- the max function $f(x) = \max\{x_1, \ldots, x_n\}$ is convex.

2.4.1 *Calculus of convex functions*

As with convex sets, it is often more convenient to show a function to be convex by showing that it can be obtained from simpler convex functions through convexity-preserving operations.

Note that most of those operations correspond to some convexity-preserving operations on convex sets, via the epigraph relation.

1. *Non-negative weighted sum.* Given $f_1, \ldots, f_n$ convex functions and $w_1, \ldots, w_n \geq 0$ non-negative weights, the function $f = \sum_{i=1}^n w_i f_i$ is a convex function.

2. *Composition with affine mappings.* If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$. Then the composite function $g : \mathbb{R}^m \to \mathbb{R} = f(Ax + b)$ is a convex function.

3. *Pointwise maximum and supremum.* If f_1 and f_2 are convex functions, their pointwise maximum $f(x) = \max\{f_1(x), f_2(x)\}$ is a convex function. This extends to any finite number of convex functions $f_1, \ldots, f_k$, and even to infinitely many by taking the supremum. In other words, given an arbitrary family $\mathcal{A}$ of convex functions, then the pointwise supremum:

$$g(x) = \sup_{a \in \mathcal{A}} f_a(x)$$

   is a convex function.

Which property of convex sets is this based on?

4. *Partial minimization.* If $f(x, y)$ is a convex function and C is a convex set, then $g(x) = \inf_{y \in C} f(x, y)$ is convex function.

2.4.2 *Firts-order conditions*

If the function is differentiable, i.e., the gradient $\nabla f(x)$ exists for any point $x \in \mathrm{dom} f$, then the following result gives a nice characterization of convex functions:

Proposition 2.4. *A function $f(x)$ is convex iff*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

for all $x, y \in \mathrm{dom} f$.

See Figure 2.6 for a geometrical interpretation of this result.

Proof. Let's start with the one dimensional ($n = 1$) case first, i.e.,

$$f(x) \text{ convex} \Leftrightarrow f(y) \geq f(x) + f'(x)(y - x) \quad \forall x, y \in \mathrm{dom} f$$

⇒) Take any two points $x, y \in \mathrm{dom} f$. Since f is convex, $(1-\theta)x+\theta y \in \mathrm{dom} f$ and we have

$$f(\underbrace{(1-\theta)x+\theta y}_{x+\theta(y-x)}) \leq (1-\theta)f(x)+\theta f(y)$$

Rearranging and dividing by θ we then obtain:

$$\frac{f(x+\theta(y-x))-f(x)}{\theta} \leq f(y)-f(x)$$

and finally:

$$\begin{aligned} f(y) &\geq f(x)+\frac{f(x+\theta(y-x))-f(x)}{\theta} \\ &= f(x)+f'(x)(y-x) && \text{(for } \theta \to 0\text{)} \end{aligned}$$

⇐) Take again any two points $x, y \in \mathrm{dom} f$ and consider the convex combination $z = \theta x+(1-\theta)y$. Let us apply the condition on the right twice, once with points x and z and once for the points y and z:

$$\begin{aligned} f(x) &\geq f(z)+f'(z)(x-z) \\ f(y) &\geq f(z)+f'(z)(y-z) \end{aligned}$$

If we multiply them by θ and $(1-\theta)$ (respectively), and add up, we obtain:

$$\theta f(x)+(1-\theta)f(y) \geq f(z)+f'(z)[\underbrace{\theta(x-z)+(1-\theta)(y-z)}_{\theta x+(1-\theta)y-z=0}]$$

and thus f is convex.

We can now consider the general n-dimensional case. Let $x, y \in \mathrm{dom} f$ and consider the restriction of f along the line passing through x and y, i.e., consider:

$$\begin{aligned} g(t) &= f(ty+(1-t)x) \\ g'(t) &= \nabla f(ty+(1-t)x)^\top (y-x) \end{aligned}$$

where the second is obtain by standard properties of multidimensional calculus.

⇒) Since f is convex so is g and we can apply the result just proved for the 1-dimensional case to obtain:

$$g(1) \geq g(0)+g'(0)(1-0)$$

which becomes, after substitutions:

$$f(y) \geq f(x)+\nabla f(x)^\top (y-x)$$

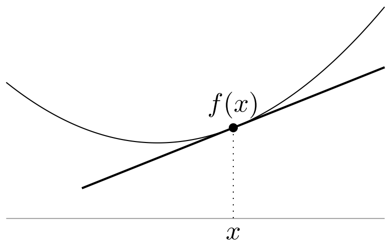


Figure 2.6: First order conditions.

$\Leftarrow$) Consider the two convex combinations $ty + (1-t)x$ and $\tilde{t}y + (1-\tilde{t})x$, and apply the condition:

$$f(ty + (1-t)x) \geq f(\tilde{t}y + (1-\tilde{t})x) \\ + \nabla f(\tilde{t}y + (1-\tilde{t})x)^\top \underbrace{(ty + (1-t)x - \tilde{t}y - (1-\tilde{t})x)}_{(y-x)(t-\tilde{t})}$$

This is nothing but:

$$g(t) \geq g(\tilde{t}) + g'(\tilde{t})(t - \tilde{t})$$

and again by the result proved in the 1-dimensional case we conclude that g is convex. Since the two points x and y were arbitrary, this is turn shows that f is convex as well.

□

Note how remarkable convexity is: it turns a *locally* valid piece of information (the gradient) into a *globally* valid one (a global *underestimator* of the function).

corollary 2.1. *If a function is convex and differentiable, then $\nabla f(x) = 0$ is a sufficient condition for x being a global minimizer of the function f, as it implies that $f(y) \geq f(x)$ for any $y \in domf$.*

2.4.3 *Second-order conditions*

If the function is twice differentiable, then we can show the following:

Proposition 2.5. *A function $f(x)$ is convex iff $\nabla^2 f(x) \succeq 0$ for all $x \in domf$.*

We remind that the matrix $\nabla^2 f(x)$ is called the *Hessian* of f at point x. For strictly convex functions, we can show the following (stronger) implication only in one direction:

Proposition 2.6. *If $\nabla^2 f(x) \succ 0$ for all $x \in domf$, then f is strictly convex.*

Note that the converse is false, take for example x^4.

2.5 Global optimality theorem

This is the most fundamental result about convex optimization.

Theorem 2.2. *Let C be a convex set and $f(x)$ be a convex function. Consider the convex optimization problem:*

$$\min_x \{f(x) \mid x\ inC\}$$

Note again that convexity turns local information into global information.

Every locally *optimal solution is also a* globally *optimal solution .*

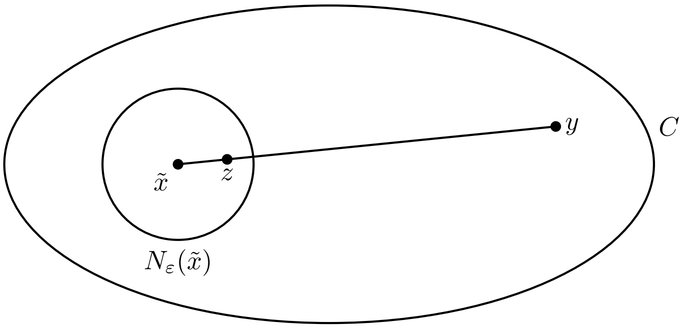


Figure 2.7: Graphical proof of Theorem 2.2.

Proof. Let $\tilde{x}$ be a locally optimal solution. By definition, there exists $\varepsilon > 0$ such that

$$f(\tilde{x}) \leq f(z) \quad \forall z \in B(\tilde{x}, \varepsilon) \cap C = \mathcal{N}_\varepsilon(\tilde{x})$$

Let us consider any $y \in C$ and choose $z = \lambda\tilde{x} + (1-\lambda)y$ on the segment between $\tilde{x}$ and y close enough to $\tilde{x}$ so that $z \in \mathcal{N}_\varepsilon(\tilde{x})$. Then we have:

$$f(\tilde{x}) \leq f(z) = f(\lambda\tilde{x} + (1-\lambda)y) \leq \lambda f(\tilde{x}) + (1-\lambda)f(y)$$

By rearranging we obtain $(1-\lambda)f(\tilde{x}) \leq (1-\lambda)f(y)$ and thus $f(\tilde{x}) \leq f(y)$, hence the claim. $\square$