1   27th of September 2018 — A. Frangioni

Definition 1.1 (Minimizing sequence)**.** *Le us consider the **minimum problem** defined in* **??***:*

$$(P) \qquad f_* = \min\{f(x) \ : x \in X\}$$

*we term **minimizing sequence** a sequence that gets closer to the optimal value:* $\{x_i\}$ *s.t.* $\{f(x_i)\} \to f_*$.

It goes without saying that we would like to avoid minimizing sequences that do not lead to an optimum.

As an example, consider

$\min\{x \ : \ x \in \mathbb{R} \wedge x > 0\}$ $\{x_i = 1/i\}$, where $\{f(x_i)\} \to 0$, but it is not an optimum or

$\min\{1/x \ : \ x \in \mathbb{R} \wedge x > 0\}$ $\{x_i = i\}$, where $\{f(x_i)\} \to 0$, but it is not an optimum.

Definition 1.2 (Interior and border of a set)**.** *Given* $S \subseteq \mathbb{R}^n$, *we say that* $x \in int(S)$ *(x is an **interior point**) if* $\exists r > 0$ *s.t.* $\mathcal{B}(x, \ r) \subseteq S$.

*On the other hand, we term **border points** those* $x \in \partial(S)$ *such that* $\forall r > 0 \ \exists y, z \in \mathcal{B}(x, \ r)$, *where* $y \in S \wedge z \notin S$. *Intuitively, a point* x *lies on the border if, all the points in the ball centered in* x *lie for a half inside the set* S *and for the other half outside.*

Notice that a point on the boundary is not necessarily inside the ball, in that case we talk about **open set** (a set which is identical to its interior: $S = int(S)$).

Definition 1.3 (Closure of a set)**.** *Let* $S \subset \mathbb{R}^n$ *we term **closure** of* S *the set* $cl(S) = int(S) \cup \partial S$.

Definition 1.4 (Closed set)**.** *We say that a set* $S \in \mathbb{R}^n$ *is **closed** if it coincides with his closure:* $S = cl(S)$.

*Equivalently, a set is termed **closed** if its complementary is open:* $\mathbb{R}^n \backslash S$.

It is interesting to notice that the cases of functions that lead to minimizing sequences that do not bring to an optimum are all defined in open sets.

Notice that there are sets that are both open and closed, for example R^n.

Definition 1.5 (Bounded set)**.** *Let* $S \subseteq \mathbb{R}^n$. *We say that* S *is **bounded** if* $\exists r > 0$ *such that* $S \subseteq \mathcal{B}(0, r)$.

Intuitively, a bounded set does not go to ∞.

Definition 1.6 (Compact set)**.** *Let* $S \subseteq \mathbb{R}^n$. *We term* S ***compact*** *if it is both closed and bounded.*

Definition 1.7 (Accumulation point)**.** *Let* $\{x_i\}$ *be a sequence.* x *is an **accumulation point** if* $\exists \ \{x_{n_i}\}$ *subsequence of* $\{x_i\}$ *such that* $\{x_{n_i}\} \to x \equiv \ \liminf_{i \to \infty} d(x_i, \ x) = 0$.

Fact 1.1. *Let S be a compact set and let $\{x_i\} \subseteq S$ minimizing sequence. (Bolzano-Weierstrass) $\{x_i\}$ has an **accumulation point** $x \in S$, thus the limit of the sequence has to be a feasible solution.*

Why did we say *feasible* but not *optimal*? If the function is not continuous (cfr. Figure 1.4) it may happen that the sequence is minimizing, but the limit is not the optimum.
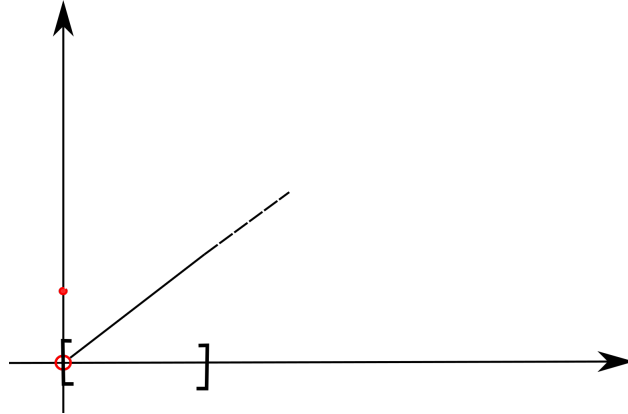

FIGURE 1.1: Case of non-continuity of the objective function in the border point $(0,0)$.

Definition 1.8 (Domain). *Let $f : D \to \mathbb{R}$. We term D **domain**: $D = dom(f)$.*

In this course we will not take into account the domain of functions, because we can make functions defined in the whole space as follows:
$f : \mathbb{R}^n \to \bar{\mathbb{R}}$, where $f(x) = \infty$ for $x \notin D$.

Definition 1.9 (Graph and epigraph). *Let $f : \mathbb{R}^n \to \mathbb{R}$. We term **graph** $gr(f) = \{(f(x), \, x) \, : \, x \in dom(f)\}$.*
*On the other hand, we term **epigraph** $epi(f) = \{(v, \, x) \, : \, x \in dom(f) \ \wedge \ v \geq f(x)\}$.*

Definition 1.10 (Level and sublevel set). *Let $f : \mathbb{R}^n \to \mathbb{R}$. We term **level set** $L(f, v) = \{x \in dom(f) \, : \, f(x) = v\}$.*
*On the other hand, we term **sublevel set** $S(f, v) = \{x \in dom(f) \, : \, f(x) \leq v\}$.*

Intuitively, the level sets are projections of the points of the domain and they are used in order to "look" at the function even if we are working in $\mathbb{R}^n$.
A graphic example of level set is displayed in Figure 1.6.

1.1 Continuity

Definition 1.11 (Continuity). *Let $f : \mathbb{R}^n \to \mathbb{R}$. We say that f is **continuous** if $\forall \varepsilon > 0 \ \exists \delta > 0$ such that $|f(y) - f(x)| < \varepsilon \ \forall y \in \mathcal{B}(x, \ \delta)$.*
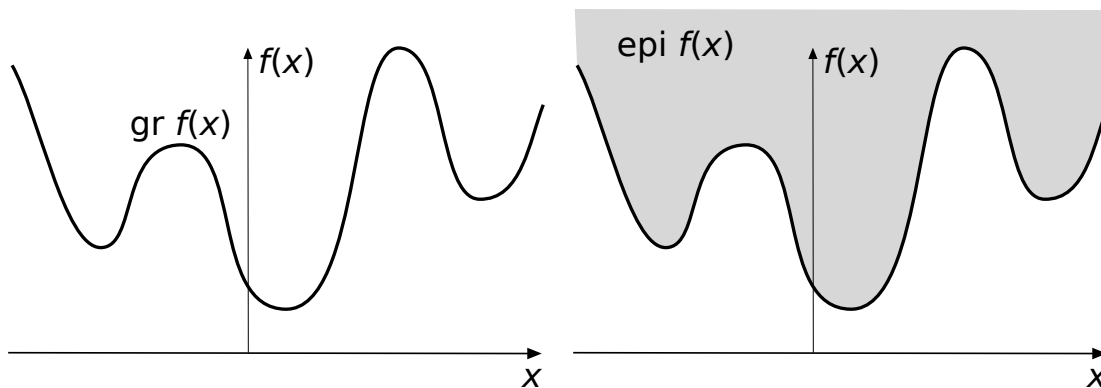
FIGURE 1.2: On the left-hand side the graph of the function $f(x)$, while on the right-handed side the epigraph of such function.
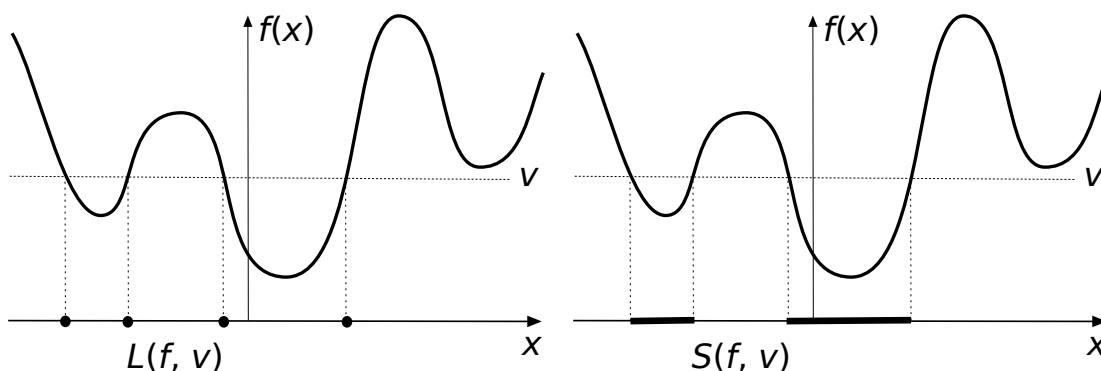

FIGURE 1.3: On the left-handed side of the figure, the level set of the function f, while on the right-hand side the sublevel set of such function.

Notice that if f, g continuous at x then:

1. $f + g$, $f \cdot g$ continuous at x

2. $\max\{f, \ g\}$ and $\min\{f, \ g\}$ continuous at x

3. $f \circ g \equiv f(g(\cdot))$ continuous at x

Theorem 1.2 (Intermediate value)**.** *Let* $f : \mathbb{R} \to \mathbb{R}$. *$f$ is continuous on* $[a, \ b]$ *if* $\forall v$ $\min\{f(a), \ f(b)\} \leq v \leq \max\{f(a), \ f(b)\} \ \exists \ c \in [a, \ b]$ *such that* $f(c) = v$.

Theorem 1.3 (Weierstrass extreme value theorem)**.** *Let* $X \subseteq \mathbb{R}^n$ *be a compact set and let* f *continuous on* X. *Then* (P) *has an optimal solution.*

Equivalently, let $X \subseteq \mathbb{R}^n$ *compact and let* f *continuous on* X. *Then all accumulation points of any minimizing sequence are optima and* there is at least one.

3

Definition 1.12 (Lipschitz continuity). *Let $f : \mathbb{R}^n \to \mathbb{R}$. We term f **Lipschitz continuous** (L.c.) on $S \in \mathbb{R}^n$ if $\exists L > 0$ such that*

$$|f(x) - f(y)| \leq L \|x - y\| \quad \forall x, y \in S$$

*More generally, f is **globally Lipshitz continuous** if $S = \mathbb{R}^n$ and it is **locally Lipshitz continuous** if at $x \; \exists \; \varepsilon > 0 \; S = \mathcal{B}(x, \; \varepsilon)$.*

Notice that the L constant value depends on S. The wider the set the smaller L.

Fact 1.4. *Let $f : \mathbb{R}^n \to R$. On a compact set $S \in \mathbb{R}^n$ if f is continuous then it is Lipshitz continuous.*

Definition 1.13 (Lower-upper semi-continuous). *Let $f : \mathbb{R}^n \to \mathbb{R}$ f is **lower[upper] semi-continuous** (l.[u.]s.c.) at x if*

$$\{x_i\} \to x \implies f(x) \leq \liminf_{i \to \infty} f(x_i) \;\; [f(x) \geq \limsup \ldots]$$

Equivalently, $\liminf_{y \to x} f(y) \geq f(x)$, $\limsup_{y \to x} f(y) \leq f(x)$.

1.2 Derivatives

In this section we address the problem of inferring information on a complicated function arround a certain value $\bar{x}$ using very simple functions, that provide information which is reliable close to the point we are studying. Those functions are called "derivatives".

It goes without saying that it is not possible to compute such simple representation in any kind of point.

Definition 1.14 (Left and right derivative). *Let $f : \mathbb{R} \to \mathbb{R}$. We term **left derivative** $f'_-(x) = \lim_{t \to 0_-} [f(x + t) - f(x)]/t$.*
*On the other hand, **right derivative** $f'_+(x) = \lim_{t \to 0_+} [f(x + t) - f(x)]/t$.*

Definition 1.15 (Differentiable). *Let $f : \mathbb{R} \to \mathbb{R}$. We say that f is **differentiable** at $x \in dom(f)$ if $f'_-(x) = f'_+(x) \Leftarrow$ they $\exists$.*

Fact 1.5. *Let $f : \mathbb{R} \to \mathbb{R}$ and differentiable in $x \in dom(f)$ then f is continuous in x.*

Theorem 1.6 (Mean value theorem). *$f : \mathbb{R} \to \mathbb{R}$ continuous on $[a, b]$ and differentiable on (a, b) then $\exists c \in (a, b)$ s.t. $f'(c) = (f(b) - f(a))/(b - a)$.*

Theorem 1.7 (Rolle's theorem). *Let $f : \mathbb{R} \to \mathbb{R}$. $f(a) = f(b)$ then $\exists \; c \in (a, b)$ s.t. $f'(c) = 0$*

Corollary 1.8. *In the same hypothesis of Rolle's theorem, let a and b consecutive roots of f. Then $f'(a)$ and $f'(b)$ have opposite sign.*

1.2.1   Multivariate differentiability

Definition 1.16 (Partial derivative). *Let $f : \mathbb{R}^n \to \mathbb{R}$. We term **partial derivative** of f w.r.t. x_i at $x \in \mathbb{R}^n$ as*

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0} \frac{f(x_1, \, ,\ldots, \, x_{i-1}, \, x_i + t, \, x_{i+1}, \, \ldots, \, x_n) - f(x)}{t}$$

In other words, it is just $f'(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)$, treating x_j for $j \neq i$ as constants.

Definition 1.17 (Gradient). *Given $f : \mathbb{R}^n \to \mathbb{R}$. We term **gradient** of f the vector of all the partial derivatives.*

Definition 1.18 (Directional derivative). *Let $f : \mathbb{R}^n \to \mathbb{R}$. We term **directional derivative** of f in point $x \in dom(f)$ along direction $d \in \mathbb{R}^n$*

$$\frac{\partial f}{\partial d}(x) := \lim_{t \to 0} \frac{f(x + td) - f(x)}{t}$$

In a multivariate space it is not possible to assure differentiability checking if the directional derivatives are equal, because there is an infinite number of different directions.

We are now ready to introduce the notion of multivariate differentiability.

Definition 1.19 (Differentiable). *Let $f : \mathbb{R}^n \to \mathbb{R}$. We say that f is **differentiable** at x if $\exists$ a linear function $\phi(h) =< c, h > +f(x)$ s.t.*

$$\lim_{\|h\| \to 0} \frac{\|f(x + h) - \phi(h)\|}{\|h\|} = 0$$

The intuition here is that the linear function should approximate f pretty well around x.

The gradient is the direction on which the function increases more rapidly, while the opposite of the gradient suggests the direction on which the function decreases most.

Fact 1.9. *Let $f : \mathbb{R}^n \to \mathbb{R}$ differentiable at x. Then f is locally Lipschitz continuous at x.*

Corollary 1.10. *Let $f : \mathbb{R}^n \to \mathbb{R}$ differentiable at x. Then f is continuous at x.*

Notice that the convers does not hold.

Fact 1.11. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let us assume $\exists \delta > 0$ s.t. $\forall i \, \frac{\partial f}{\partial x_i}(y)$ is continuous $\forall y \in \mathcal{B}(x, \delta)$. Then f differentiable at point x.*

Notice that the converse of Proposition 1.11 does not hold either.

We are now ready to introduce a class of functions that allows good results for optimization: $\mathcal{C}^1 := \nabla f(x)$ continuous.

Let $f \in \mathcal{C}^1$, then f is differentiable everywhere and also continuous everywhere.

Figure 1.4, Figure 1.5 and Figure 1.6 picture some functions which are not differentiable in the minimum.
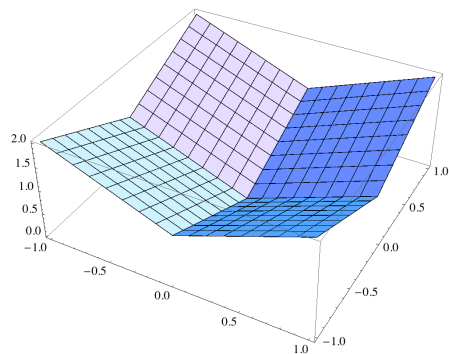
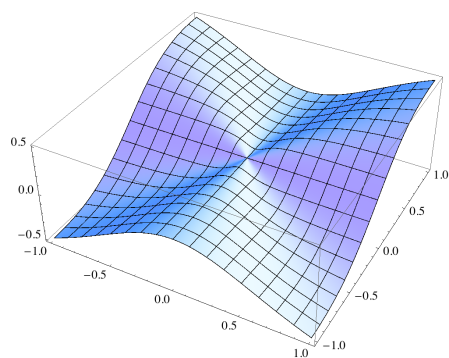FIGURE 1.4: The function $f(x_1, x_2) = \|[x_1, x_2]\| = |x_1| + |x_2|$ has some kinks.

FIGURE 1.5: The function $f(x_1, x_2) = \frac{x_1^2 x_2}{x_1^2 + x_2^2}$ may be put to 0 in $(0,0)$ for continuity, but it is still not differentiable in $(0,0)$ although the function admits directional derivatives in $(0,0)$.
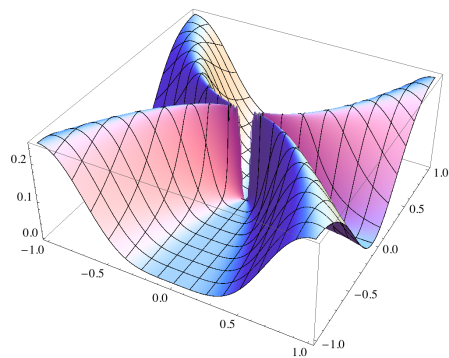

FIGURE 1.6: The function $f(x_1, x_2) = \left(\frac{x_1^2 x_2}{x_1^4 + x_2^2}\right)^2$ is not continuous in 0. There are some directions that lead to the limit 0, while there are some other directions (the parabolas above) that do not lead to value 0.

Definition 1.20 (First order model). *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $x \in dom(f)$. We term **first***

6

order model *of f at point x*

$$L_x(y) = \nabla f(x)(y - x) + f(x)$$

Fact 1.12. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $x \in dom(f)$ such that f is differentiable at x. Then $(L_x, f(x)) \perp S(f, f(x)) \perp \nabla f(x)$.*

Geometrically speaking, we can observe that a function is smooth whenever its levelsets are smooth.