1 21st of September 2018 — A. Frangioni

1.1 Mathematical background for optimization problems

Definition 1.1 (Minimum problem)**.** *Let X be a set, called **feasible region** and let $f : X \to \mathbb{R}$ be* any *function, called **objective function** we call **problem** the following*

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

Definition 1.2 (Feasible solution)**.** *Let $x \in F$ be a solution of the minimum problem in which the domain is a superset of $X \subset F$. We say that x is a **feasible solution** if $x \in X$. On the other hand, x is **unfeasible** if $x \in F \setminus X$.*

Definition 1.3 (Optimal solution)**.** *Under the same hypothesis of the above definition, we define x_* such that $f(x_*) = f_*$ an **optimal solution**, where $f_* \leq f(x) \, \forall x \in X$, $\forall v > f_* \, \exists \, x \in X$ s.t. $f(x) < v$.*

It is possible to find problems where there is no optimal solution at all.

Example 1.1. *There are two cases in which it is not possible to find an optimal solution:*

1. *The domain is empty, which may be not trivial to prove, since it is an NP-hard problem sometimes;*

2. *We want to find the minimum of the objective function but it is unbounded below ($\forall M \, \exists x_M \in X$ s.t. $f(x_M) \leq M$). On the other hand, we need to maximize the function, but it is unbounded above.*

We can now rewrite the problem of solving an optimization problem as:

1. Finding x_* and proving it is optimal

2. Or proving $X = \emptyset$

3. Or constructively prove $\forall M \, \exists x_M \in X$ s.t. $f(x_M) \leq M$.

Typically $x \in \mathbb{R}$ actually mean $x \in \mathbb{Q}$ with up to k digits precision and most of the times we consider optimal a solution which is close to the true optimal value, modulo some error ($\bar{x}$, the approximately optimal).

Definition 1.4 (Absolute error)**.** *We call **absolute error** the gap between the real value and the one we obtained. Formally,*

$$f(\bar{x}) - f_* \leq \varepsilon$$

Definition 1.5 (Relative error)**.** *We define as **relative error** the absolute error, normalized by the true value of the function*

$$(\, f(\bar{x}) - f_* \,)/| \, f_* \, | \leq \varepsilon$$

Multi-objective Optimization:
What happens if we have more than one objective function?
Often you need more than one, say:

$$(P) \qquad \min\{[f_1(x), f_2(x)] \ : x \in X\}$$

with f_1, f_2 contrasting and/or with incomparable units (apples vs. oranges)

In multi-objective optimization, there does not typically exist a feasible solution that minimizes all objective functions simultaneously. Therefore, attention is paid to *Pareto optimal solutions*; that is, solutions that cannot be improved in any of the objectives without degrading at least one of the other objectives.

Definition 1.6 (Pareto frontier)**.** *A feasible solution $x^1 \in X$ is said to (Pareto) dominate another solution $x^2 \in X$, if*

$$f_i(x^1) \leq f_i(x^2) \quad for \quad all \quad indices \quad i \in \{1, 2, \ldots, k\} \quad and$$

$$f_j(x^1) \leq f_j(x^2) \quad for \quad all \quad indices \quad j \in \{1, 2, \ldots, k\}$$

A solution $x^* \in X$ (and the corresponding outcome $f(x^*)$) is called Pareto optimal, if there does not exist another solution that dominates it. The set of Pareto optimal outcomes is often called the ***Pareto frontier****.*

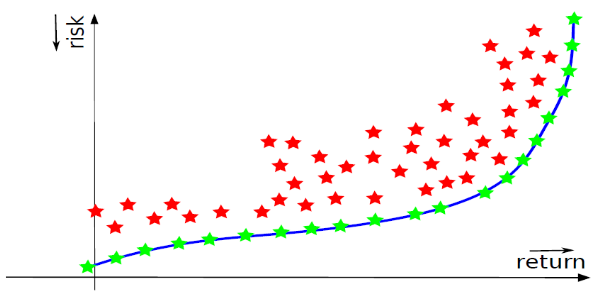

Figure 1.1: An example of Pareto frontier

We are provided with two practical solutions:

Scalarization: using a linear combination of the two functions: $f(x) = \alpha f_1(x) + \beta f_2(x)$;

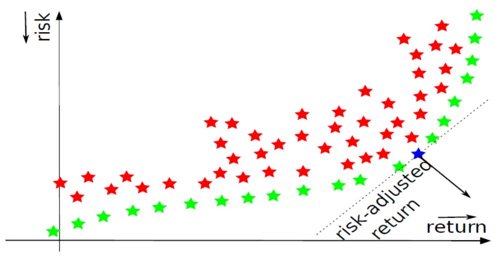

Figure 1.2: Maximize risk-adjusted return, $\min\{f_1(x) + \alpha f_2(x) \ : x \in X\}$

Budgeting: $f(x) = f_1(x)$, $X := X \cup \{\, f_2(x) \leq b \,\}$, which intuitively corresponds to taking into account only one objective function and add the others as constraints, provided that the values of the other functions are not too high.

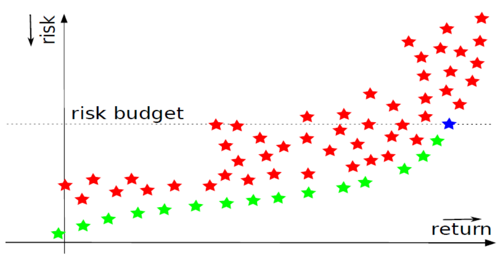

Figure 1.3: Maximize return with budget on maximum risk, $\min\{f_1(x) \ : f_2(x) \leq \beta \ : x \in X\}$

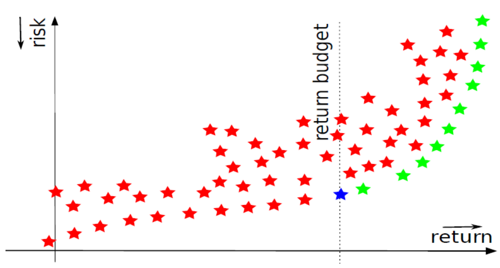

Figure 1.4: Minimize risk with budget on minimum return,, $\min\{f_2(x) \ : f_1(x) \leq \beta \ : x \in X\}$

1.1.1 Examples of Bad optimization problems

Here some problems that has no optimal solution:

- empty case ($X = \emptyset$): $\min\{x \ : \ x \in \mathbb{R} \wedge x \leq -1 \wedge x \geq 1\}$
- unbounded [below]: $\min\{x \ : \ x \in \mathbb{R} \wedge x \leq 0\}$
- bad X: $\min\{x \ : \ x \in \mathbb{R} \wedge x > 0\}$
- bad f and X: $\min\{x \ : \ x \in \mathbb{R} \wedge x > 0\}$
- bad f: let us consider an iterative algorithm that moves towards the optimum. It may happen that the function decreases and increases along a certain direction but its non-continuity leads to the impossibility of reaching the optimum. As an example, let us take the following

$$\min \left\{ f(x) = \begin{cases} x & \text{if } x > 0 \\ 1 & \text{if } x = 0 \end{cases} \quad : \quad x \in [0,1] \right\}$$

1.2 Infima, suprema and extended reals

Since we minimize/maximize stuff, infima/suprema are important:

Definition 1.7 (Totally ordered set)**.** *We say that set X is **totally ordered** if $\forall\, x, y \in X$, either $f(x) \leq f(y)$ or $f(y) \leq f(x)$.*

Definition 1.8 (Infima and suprema)**.** *Given a totally ordered set R and one of its subsets (say $S \subseteq R$)*

$$s \text{ is the } \textbf{infimum} \text{ of } S \Leftrightarrow \underline{s} = \inf S \quad \Leftrightarrow \quad \underline{s} \leq s \;\; \forall s \in S \;\; \wedge \;\; \forall t > \underline{s}\, \exists\, s \in S \text{ s.t. } s \leq t$$

$$s \text{ is the } \textbf{supremum} \text{ of } S \Leftrightarrow \bar{s} = \sup S \quad \Leftrightarrow \quad \bar{s} \geq s \;\; \forall s \in S \;\; \wedge \;\; \forall t < \bar{s}\, \exists\, s \in S \text{ s.t. } s \geq t$$

Issue: inf S/sup S may not exist in $\mathbb{R}$

Definition 1.9 (Extended real)**.** *In the case of unbounded functions the value of infima or suprema are ∞, and we call **extended reals** $\overline{\mathbb{R}} = -\infty \cup \mathbb{R} \cup +\infty$.*

- For all $S \subseteq \mathbb{R}, \sup/\inf S \in \overline{\mathbb{R}}$
- $\inf S = -\infty$ just a convenient notation for there is no (finite) inf
- $\inf \emptyset = \infty$, $\sup \emptyset = -\infty$

1.3 (Monotone) Sequences in $\mathbb{R}$ and optimization

We are interested in studying sequences, because iterative methods start from a certain point and move towards the optimal, hopefully.

Sequence of iterates $\{x_i\} \subset X$ and $v_i = f(x_i)$. Typically we cant get f in finite time ($\exists i v_i = f$), but we can get as close as we want: there in the limit.

Definition 1.10 (Limit)**.** *Given a sequence $\{x_i\}$ the **limit** for $i \to \infty$ is defined as*

$$\lim_{i \to \infty} v_i = v \quad \Longleftrightarrow \quad \forall \varepsilon > 0 \;\exists\; h \text{ s.t. } |v_i - v| \leq \varepsilon \;\forall i \geq h$$

It may happen that a sequence has or does not have a limit. For example $\{\frac{1}{n}\}$ has limit 0 for $n \to +\infty$, while $\{(-1)^n\}$ does not.

Fact 1.1. *Let us be given a monotone sequence, then the sequence **does** have a limit.*

Notice that given a sequence either it is monotone or it can be "split" into two monotone sequences (for example $\{(-1)^n\}$ can be transformed into $\{(-1)^{2n}\}$ and $\{(-1)^{2n+1}\}$ and these two sequences are both monotone).

The obvious way to make $\{vi\}$ monotone: keep aside the best

$$v_i^* = \min\{v_h : h \leq i\} \qquad \text{(best value at interaction i)}$$

- $v_1^* \geq v_2^* \geq v_3^* \geq \;\ldots \Rightarrow v_\infty^* = \lim_{i\to\infty} v_i^* \geq f_*$ (asymptotic estimate)
- $\lim_{i\to\infty} v_i^* = v_\infty^* = f_* \Rightarrow \{v_i\}$ minimizing sequence (of values)

Forcing monotonicity on sequences in $\mathbb{R}$: the hard way

- Extract monotone sequences from $\{ v_i \}$ "the hard way":
  $\underline{v}_i = \inf\{ v_h \,:\, h \geq i \}$, $\bar{v}_i = \sup\{ v_h \,:\, h \geq i \}$
- $\underline{v}_1 \leq \underline{v}_2 \leq \underline{v}_3 \leq \ldots,\ \bar{v}_1 \geq \bar{v}_2 \geq \bar{v}_3 \geq \ldots \implies$ they still have a limit
- $\liminf_{i\to\infty} v_i := \lim_{i\to\infty} \underline{v}_i = \sup_i \underline{v}_i$
- $\limsup_{i\to\infty} v_i := \lim_{i\to\infty} \bar{v}_i = \inf_i \underline{v}_i$
- $\bar{v}_i \geq \underline{v}_i \implies \limsup_{i\to\infty} v_i \geq \liminf_{i\to\infty} v_i$
- $\lim_{i\to\infty} v_i = v \iff \limsup_{i\to\infty} v_i = v = \liminf_{i\to\infty} v_i$
- $\liminf_{i\to\infty} v_i = f_* \implies \{ v_i \}$ minimizing sequence (of values)
- A stronger definition: $\liminf_{i\to\infty} v_i = f_* \implies \lim_{i\to\infty} v_i^* = f_*$

FIGURE 1.5: Forcing monotonicity on sequences in $\mathbb{R}$: the hard way

1.4 Vector spaces and topology

1.4.1 Euclidean space $\mathbb{R}^n$

Single numbers are not enough, except for objective function values.

Definition 1.11 (Euclidean vector space)**.** *We call* ***Euclidean space***

$$\mathbb{R}^n := \{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \ :\ x_i \in \mathbb{R},\ i = 1, \ldots, n\}$$

Equivalently, we can characterize the Euclidean space as Cartesian product of $\mathbb{R}$ n *times:* $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \ldots \mathbb{R}$. *Closed under sum and scalar multiplication*

The main operations on elements of the Euclidean space (vectors) are:

SUM: $x + y := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$

SCALAR MULTIPLICATION: $\alpha x = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix}$

Usually $x \in \mathbb{R}^n$ usually considered column vector $\in \mathbb{R}^{n\times 1}$, otherwise a row vector is x^T.

Definition 1.12 (Finite vector space)**.** *each $x \in \mathbb{R}^n$ can be obtained from a finite basis. (canonical base is u_i having 1 in position i and 0 elsewhere)*

Notes for vector space:

- Not all vector spaces are finite
- Not a totally ordered set

Concept of limit requires topology. So, in order to be able to compute limits in a vector space we need some topology definitions: norm, scalar product, distance.

1.4.2 (Euclidean) norm

Definition 1.13. *Let $x \in \mathbb{R}^n$ we define the **euclidean norm** of a vector:*

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} {x_i}^2} = \sqrt{< x, x >}$$

Fact 1.2. *The norms on a vector space have the following properties:*

1. $\|x\| \geq 0$ *and* $\forall x \in \mathbb{R}^n$, $\|x\| = 0 \iff x = 0$;
2. $\|\alpha x\| = |\alpha| \, \|x\|, \ \forall x \in \mathbb{R}^n, \ \alpha \in \mathbb{R}$;
3. $\|x + y\| \leq \|x\| + \|y\|, \ \forall x, y \in \mathbb{R}^n$ *(triangle inequality).*

Fact 1.3 (Cauchy-Schwartz inequality)**.** *Let $x, y \in \mathbb{R}^n$. The following holds:*

$$< x, y >^2 \leq < x, x >< y, y > \equiv |< x, y >| \leq \|x\| \, \|y\| , \ \forall x, y \in \mathbb{R}^n$$

Fact 1.4 (Parallelogram Law)**.**

$$2\|x\|^2 + 2\|y\|^2 = \|x + y\|^2 + \|x - y\|^2$$

Fact 1.5.

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2 < x, y >$$

1.4.3 A useful norm generalization: p-norm

Many (but not all) derive from p-norm:

Definition 1.14 (p-norm). *Let $p \geq 1$ be a real number, the p-norm of a vector $x = (x_1, \ldots, x_n)$ is defined as follow:*

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

We require $p \geq 1$ for the general definition of the p-norm because the triangle inequality fails to hold if $p < 1$. The p-norm is convex for $p \geq 1$, nonconvex for $p < 1$

Here some norm derived from p-norm:

- $\|x\|_1 := \sum_{i=1}^{n} |x_i|$
- $\|x\|_\infty := \max\{|x_i| \ : \ i = 1, \ldots, n\}$
- $\|x\|_i := |\{i \ : \ |x_i| > 0\}|$
- Other ones (e.g. for matrices . . .)

Fact 1.6. *For any given finite-dimensional vector space V (e.g. $\mathbb{R}^n$ is a finite vector space), all norms on V are equivalent in the sense that given two norms $\|\cdot\|_A$, $\|\cdot\|_B$:*

$$\exists\, 0 < \alpha < \beta \ \ s.t \ \ \alpha \|x\|_A \leq \|x\|_B \leq \beta \|x\|_A \qquad \forall x \in V$$

Therefore convergence in one norm implies convergence in any other norm. This rule may not apply in infinite-dimensional vector spaces such as function spaces, though

Fact 1.7 ((Holders inequality).

$$< x, y >^2 \leq \|x\|_p \|y\|_q \quad 1/p + 1/q = 1$$

Definition 1.15 (Ball). *We term **ball** centered in $\bar{x}$ and having ε as radius as the set of points that are close enough to $x \in \mathbb{R}^n$: $B(\bar{x}, \varepsilon) = \{x \in \mathbb{R}^n \ : \ \|x - \bar{x}\| \leq \varepsilon\}$.*

Let's take a unit ball, if the center of the unit-ball is in the origin (0,0), then each point on the unit-ball will have the same p-norm (i.e. 1). The unit ball therefore describes all points that have "distance" 1 from the origin, where "distance" is measured by the p-norm. In Figure 1.6 we may observe the different shapes of the same ball varying the value of p in the p-norm.

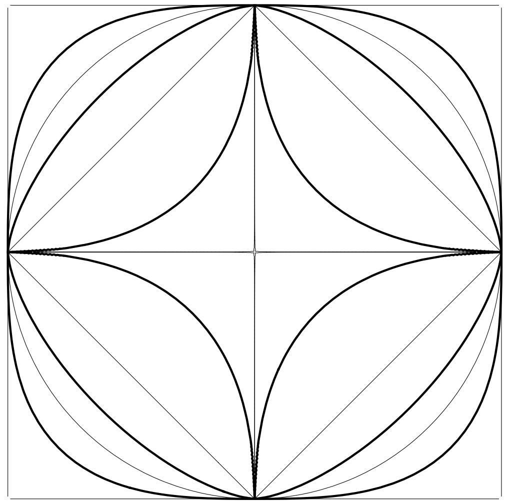

FIGURE 1.6: The shapes of balls centered in the origin of radius 1 varying the value of p-norm.

1.4.4 (Euclidean) Scalar Product

Definition 1.16 (Scalar product)**.** *Let $x, y \in \mathbb{R}^n$ we define the **scalar product** between these two vectors*

$$< x, y > := \; y^T x = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots + x_n y_n$$

Fact 1.8. *A scalar product has the following properties:*

1. $< x, y >=< y, x > \quad \forall x, y \in \mathbb{R}^n$ *(symmetry)*
2. $< x, x >\geq 0, \; \forall x \in \mathbb{R}^n, < x, x >= 0 \iff x = 0$*;*
3. $< \alpha x, y >= \alpha < x, y >, \; \forall x \in \mathbb{R}^n, \alpha \in \mathbb{R}$*;*
4. $< x + y, z >=< x, z > + < y, z >, \; \forall x, y, z \in \mathbb{R}^n$*.*

Geometric interpretation of the scalar product, an important characterization of the scalar product is the one that uses angles:

$$< x, y >= \|x\| \, \|y\| \cos \theta$$

- $x \perp y \iff < x, y >= 0$ (orthogonality condition)
- $< x, y > > \; 0 \iff$ "x and y point in the same direction"

More General: $< x, y >_M := y^T M x$ with $M \succ 0$ $(x \longrightarrow M^{-1/2} x)$

1.4.5 (Euclidean) Distance

Definition 1.17 ((Euclidean) distance)**.** *The **Euclidean distance** between points x and y is the length of the line segment connecting them. In Cartesian coordinates, if $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ are two points in Euclidean n-space, then the distance (d) from x to y, or from y to x is given by*

$$d(x, y) := \|x - y\| = \sqrt{(x_1 - y_1)^2 + ... + (x_n - y_n)^2}$$

Fact 1.9. *The distance has the following properties:*

1. $d(x, y) \geq 0 \; \forall x, y \in \mathbb{R}^n$, $d(x, y) = 0 \iff x = y$
2. $d(\alpha x, 0) = |\alpha| d(x, 0) \;\; \forall x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$
3. $d(x, y) \leq d(x, z) + d(z, y) \;\; \forall x, y, z \in \mathbb{R}^n$ *(triangle inequality)*

1.5 Limit of a sequence in $\mathbb{R}^n$

We have now all the tools to define the notion of limit of a sequence in $\mathbb{R}^n$. **Limit of a sequence in R^n**

Definition 1.18 (Limit of a sequence in the Euclidean space)**.** *Let $\{x_i\} \subset \mathbb{R}^n$ be a sequence in $\mathbb{R}^n$. The **limit** of $\{x_i\}$ for $i \to +\infty$ is the following:*

$$\lim_{i \to \infty} x_i = x \equiv \{x_i\} \to x$$

$$\Updownarrow$$

$$\forall \varepsilon > 0 \; \exists h \; s.t. \; d(x_i, x) \leq \varepsilon \; \forall i \geq h$$

$$\Updownarrow$$

$$\forall \varepsilon > 0 \; \exists h \; s.t. \; x_i \in \mathcal{B}(x, \varepsilon) \; \forall i \geq h$$

$$\Updownarrow$$

$$\lim_{i \to \infty} d(x_i, x) = 0$$