Optimization without constraints

We work in $(\mathbb{R}^d, \|\cdot\|)$, where $\|\cdot\|$ will almost always be the canonical Euclidean norm.

Let $f : D \subset \mathbb{R}^d \to \mathbb{R}$ be a function, where D denotes the domain of definition of f. We aim to answer these kinds of questions:

- is f bounded? does f reach its bounds? (*existence* of an extremum)
- where? (at a unique point?) (*uniqueness*)
- what about local extrema?
- effective search of extrema?

Vocabulary: "In mathematical analysis, the **maxima** and **minima** (the respective plurals of **maximum** and **minimum**) of a function, known collectively as **extrema** (the plural of **extremum**), are the largest and smallest value of the function, either within a given range (the **local** or **relative** extrema) or on the entire domain of a function (the **global** or **absolute** extrema)." (source: Wikipedia, *Maxima and minima*)

1 Overview on necessary or sufficient conditions in optimization

Sufficient condition for global existence: topological questions

Theorem. *If f is $\mathcal{C}^0$ and D is **compact** (*i.e.* closed and bounded), then f is bounded and attains its bounds on D.*

Sketch of proof.

Exercise. If $f : \mathbb{R}^d \to \mathbb{R}$ is $\mathcal{C}^0$ and **coercive** (*i.e.* $f(x) \to +\infty$ when $\|x\| \to \infty$), then f is bounded below and admits a global minimum.

Conditions for local optimization: differential calculus

Proposition (Necessary conditions of first order)**.** *If $x \in \mathring{D}$ is a local extremum of f and f is differentiable, then x is a* ***critical point*** *of f (*i.e. $\nabla f(x) = 0$*).*

Proposition (Necessary condition of second order)**.** *If $x \in \mathring{D}$ is a local minimum of f and f is $\mathcal{C}^2$, then x is a critical point of f and the Hessian matrix of f at x is a nonnegative matrix (*i.e. $Hf(x) \in S_d^+(\mathbb{R})$*).*

Proposition (Sufficient condition of second order)**.** *If f is $\mathcal{C}^2$, and $x \in D$ is a critical point of f such that the Hessian matrix of f at x is a positive matrix (*i.e. $Hf(x) \in S_d^{++}(\mathbb{R})$*), then x is a (strict) local minimum of f.*

Proof by Taylor formulae.

Convexity: local is global, necessary becomes sufficient

Assume that D is a convex set.

Proposition. *If f is convex, then a local minimum for f is a global minimum.*

Proof.

Proposition (Necessary and sufficient condition of first order)**.** *If f is convex and differentiable on D and $x \in D$, then if $\nabla f(x) = 0$, x is a (local and global) minimum of f.*

Proof.

Convexity supplies also more information on the points where extrema may be attained.

Proposition. *If $f : D \to \mathbb{R}$ is convex, then*

1. *the set of points at which f attains its minimum is convex,*
2. *if f attains a global maximum in $\mathring{D}$, f is constant.*

Furthermore, if f is strictly convex and has a minimum, it is attained at a unique point.

Example. Extrema of $(x, y) \mapsto x^2 + y^2$ on a rectangle containing 0.

2 Optimization trivia

2.1 Functions of one variable ($d = 1$)

In dimension 1, the search for extrema of f often comes down to the study of the sign of f' (study of *variations*).

Exercise. Determine local and global extrema (if they exist) for the following functions.

1. $f(x) = x(1-x)$ on $I = [0, 1]$.
2. $f(x) = 1 - e^{-x}$ on $I = \mathbb{R}^+$.
3. $f(x) = 3x^4 - 4x^3 + 6x^2 - 12x + 1$ on $I = \mathbb{R}$.
4. $f(x) = \frac{1}{\sqrt{x^2-x+1}}$ on $I = [0, 1]$.

Exercise. Let $f : [0, +\infty[\to \mathbb{R}$ be a $\mathcal{C}^0$ function with finite limit at $+\infty$.

1. Show that f is bounded, and admits a global maximum or minimum.
2. Give an example of function with a global maximum but no minimum.

Exercise. Let I be an interval of $\mathbb{R}$ and $f : I \to \mathbb{R}$ be derivable. If x is a local minimum of f and f' cancels only at x, show that x is actually a global minimum of f.

Exercise. Build an example of a non bounded below strictly convex function on $\mathbb{R}$.

Exercise (Uniformly strictly convex function). Let $\alpha > 0$ and f be a $\mathcal{C}^2$ function on $\mathbb{R}$ such that $f'' \geq \alpha$ on $\mathbb{R}$. Show that f attains a global minimum at a unique point.

Exercise (Newton's method). The Newton's method is an iterative algorithm for finding zeroes of a function. This algorithm relies on a classical fixed point determination argument. It can be applied to the derivative of a function in order to localize effectively its points of annulation, hence a potential extremum for the initial function.

Let $f : [a, b] \to \mathbb{R}$ be a $\mathcal{C}^2$ function with $f' > 0$, and such that $f(a) < 0 < f(b)$. Newton's method aim to determine a fixed point of

$$F(x) = x - \frac{f(x)}{f'(x)},$$

considering the sequence (x_n) iteratively defined by $x_{n+1} = F(x_n)$, $x_0 \in [a, b]$.

1. Check that f cancels at a unique point z which is a fixed point of F.
2. In general, (x_n) does only converge for x_0 close enough to z:
   (a) Using a second order Taylor formula, show that for all x in $[a, b]$, there exists t in $[a, x]$ such that
   $$F(x) - z = \frac{1}{2}\frac{f''(t)}{f'(x)}(x - z)^2.$$
   (b) Deduce that there exists $C > 0$ such that
   $$|F(x) - z| \leq C|x - z|^2$$
   for all x in $[a, b]$, and $\alpha > 0$ such that $[z - \alpha, z + \alpha]$ is included in $[a, b]$ and stable by F. Deduce that if x_0 is in $[z - \alpha, z + \alpha]$, (x_n) converges to z.
3. Show that if f is convex on $[a, b]$, $[z, b]$ is stable by F. Check that if x_0 is in $[z, b]$, (x_n) is nonincreasing and converges to z.
4. Give a geometrical interpretation of the construction of (x_n).

2.2 Schematic strategy of optimization for $d > 1$

First step (*à ne pas négliger*). Global considerations:

- domain of the function (is it compact?),
- regularity of the function ($\mathcal{C}^0$? $\mathcal{C}^1$? $\mathcal{C}^2$? $\mathcal{C}^\infty$???),
- is f convex?
- quick check for obvious bounds (ex: explicitly nonnegative function?),
- obvious behaviour at ∞ for some fixed variables? (this can easily exclude that a function is globally bounded/bounded below/bounded above)

Second step. If f is differentiable, look for the critical points, *i.e.* solve $\nabla f(x) = 0$.
Third step. If x is a critical point, and f is $\mathcal{C}^2$, determine $Hf(x)$ (note that it is unnecessary to calculate Hf at the other points! Sometimes simplifications can be made). Then:

- if $Hf(x)$ has positive eigenvalues, f attains a strict local minimum at f,
- if $Hf(x)$ has negative eigenvalues, f attains a strict local maximum at f,
- if $Hf(x)$ has (nonzeroes) eigenvalues of opposite signs, x is called a **saddle point** (or **minimax point** or **mountain pass**) of f: there exists a direction along which f attains a minimum at x, and other along which f attains a maximum at x,
- **else, EVERYTHING HAPPENS!** (it does!) $\to$ Taylor expansion of f at x.

Fourth step (à ne pas négliger non plus). Conclusion: are the critical points local extrema? global extrema? are there other potential local extrema on the border that are not critical points? Use cautiously the previously stated necessary and sufficient conditions for optimization.

2.3 Functions of two variables ($d = 2$)

Dealing with two-dimensional matrices makes easier the discussion on the sign of eigenvalues. With Monge's notation:

Proposition. *If f is $\mathcal{C}^2$ around a critical point a, and*

$$r = \frac{\partial^2 f}{\partial x^2}(a), \qquad s = \frac{\partial^2 f}{\partial x \partial y}(a) = \frac{\partial^2 f}{\partial y \partial x}(a), \qquad t = \frac{\partial^2 f}{\partial y^2}(a),$$

then:

- *if $rt - s^2 > 0$ and $r > 0$, a is a strict local minimum of f,*
- *if $rt - s^2 > 0$ and $r < 0$, a is a strict local maximum of f,*
- *if $rt - s^2 < 0$, a is not a local extremum but a saddle point,*
- *if $rt - s^2 = 0$, everything happens.*

Proof.

Exercise. Find the local extrema of the following functions and describe their other critical points. Are the local extrema global?

1. $f(x,y) = x^2(1+y)^3 + y^2$.
2. $f(x,y) = 3x^3 + xy^2 - xy$.
3. $f(x,y) = x^4 + y^8$.
4. $f(x,y) = 4x^2 - xy - x^3 + y^2$.
5. $f(x,y) = x^2 - xy + y^2 + 3x - 2y - 1$.
6. $f(x,y) = \frac{1}{1-x} + \frac{1}{1-y} + \frac{1}{x+y}$.

Exercise. Let $D = \{(x,y) \in \mathbb{R}^+, x+y \leq 4\}$, and $f(x,y) = x^2 + 2y^2 - 4(x+y)$. Justify that f admits global extrema and determine their values.

A bit painful, right? This is a motivation for the introduction of Lagrange multipliers!

2.4 Higher dimension

Exercise. Let us optimize the following functions on their domain of definition.

1. $f(x,y,z) = (x-2)^2 + y^2 + z^2$.
2. $f(x,y,z) = x^4 - y^2 + z^2$.
3. $f(x,y,z) = x^3 + y^3 + z^3 + 3xyz$.
4. $f = (x^2+y^2+z^2)\ln(x^2+y^2+z^2-5)$.

Exercise (Quadratic optimization). For $A \in S_d^{++}(\mathbb{R})$ and $B \in \mathbb{R}^d$, let $J(x) = \frac{1}{2}\langle Ax, x\rangle + \langle B, x\rangle$. Show that J is strictly convex and coercive. Express its critical point in terms of A and B.

Exercise (Rayleigh quotient). For $A \in S_d^{+}(\mathbb{R})$, define $f(x) = \frac{\langle Ax, x\rangle}{\|x\|^2}$. It is $\mathcal{C}^\infty$ on $\mathbb{R}^d \setminus \{0\}$.

1. Show that f is bounded on its domain of definition. *Hint:* reduce to the unit ball.
2. Determine the critical points of f.
3. Show that the Hessian matrix of f at a critical point $x^\star$ is $Hf(x^\star) = \frac{2}{\|x^\star\|^2}(A - f(x^\star)I_d)$. Deduce that the critical points who do not give extrema are saddle points.

3 Levelsets, gradient field and gradient descent

As before, $f : D \subset \mathbb{R}^d \to \mathbb{R}$ is a $\mathcal{C}^1$ function. But here d will almost always be 2.

Definition. The **levelset** of f associated with a value c is the (possibly empty) subset of $\mathbb{R}^d$ defined by $\{x \in D | f(x) = c\}$. We will denote this set by f^c.

The **gradient field** of f is the vector field $x \mapsto \nabla f(x)$.

Lemma. *If $\gamma : I \subset \mathbb{R} \to D$ is a $\mathcal{C}^1$ path, and $v(t) = f(\gamma(t))$ is the value along this path, then*

$$v'(t) = \nabla f(\gamma(t)) \cdot \dot{\gamma}(t) \qquad \forall t \in I.$$

Consequence. The value of the function increases when *following the gradient field* (this is the intuitive statement for : along a path $\gamma(t)$ solving the ODE $\dot{\gamma}(t) = \nabla f(\gamma(t))$).

Exercise. Draw in $\mathbb{R}^2$ some levelsets of the following functions, and represent the gradient field:

1. $f(x,y) = 3x + 4$.
2. $f(x,y) = xy$.
3. $f(x,y) = \frac{1}{x^2+y^2}$.
4. $f(x,y) = x^2 - y^2$.
5. $f(x,y) = x^2 - y$.
6. $f(x,y) = y^2 - x^2 + x^3$.

Definition. Let $x \in D$ be such that $\nabla f(x) \neq 0$, and take $c = f(x)$. The **tangent space** of the levelset f^c at x is defined by

$$T_x f^c = \left\{ v \in \mathbb{R}^d, \exists \mathcal{C}^1 \text{ path } \gamma : [-r, r] \to D \text{ with } \left| \begin{array}{l} \gamma(0) = x \\ \dot{\gamma}(0) = v \\ f(\gamma(t)) = c \;\; \forall t \in [-r, r] \end{array} \right. \right\}.$$

The tangent space is a local linear approximation of the levelset.

Proposition. *With the previous assumptions,* $T_x f^c = \left\{ v \in \mathbb{R}^d, \nabla f(x) \cdot v = 0 \right\}$. *In particular, it is a* $(d-1)$*-dimensional vector space.*

The gradient field is perpendicular to the (tangent spaces of the) levelsets.

Gradient descent

"**Gradient descent** is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point. If, instead, one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient **ascent**. Gradient descent is also known as **steepest** descent." (source: Wikipedia, *Gradient descent*)

Elementary implementation of an ascent gradient algorithm

Take x_0 in D (*initial condition*), fix $\alpha > 0$ (*step size multiplier*) and $\varepsilon > 0$ (*precision*).
As long as $\|\nabla f(x_k)\| \geq \varepsilon$, define $x_{k+1} = x_k + \alpha \nabla f(x_k)$.
End when $\|\nabla f(x_k)\| \leq \varepsilon$.

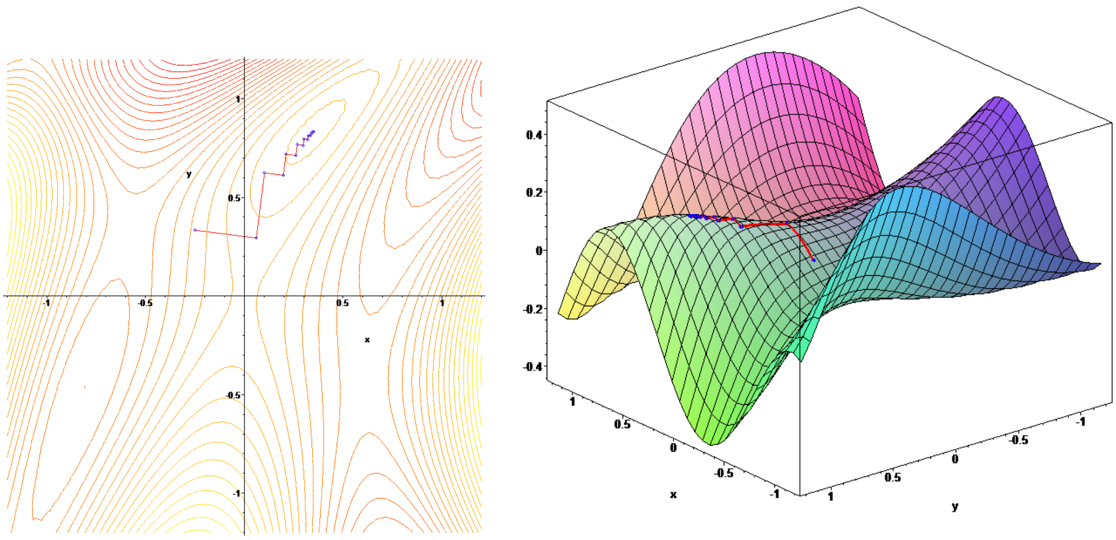

Figure 1: Implementation of a gradient ascent for $f(x, y) = \sin\left(\frac{1}{2}x^2 - \frac{1}{4}y^2 + 3\right)\cos(2x + 1 - e^y)$ (source: Wikipedia, *Gradient descent*)