Numerical optimization

Mines Nancy – Fall 2024

session 2 – existence and uniqueness of solutions
optimality conditions for unconstrained problems

Lecturer: Christophe Zhang

Course material:

- arche.univ-lorraine.fr/course/view.php?id=74098
- github.com/jflamant/mines-nancy-fall24-optimization

Outline

Context

Two categories of optimization problems

$$\min_{\mathbf{x}\in\mathbb{R}^N} f(\mathbf{x})$$

unconstrained optimization

$$\min_{\mathbf{x}\in\Omega} f(\mathbf{x})$$

constrained optimization

In this session

- Review of elementary results from multivariable calculus and algebra
- Notions of gradient, Hessian, and convexity
- Existence and uniqueness results for minimizers
- For unconstrained problems, necessary and sufficient optimality conditions for a point to be a minimizer.

Differentiability (I)

Let $f : U \to \mathbb{R}$ be a function defined on a open subset U of $\mathbb{R}^N$.

Definition

The function f is said to be differentiable at $a \in U$ if there exists a linear map $df_a : \mathbb{R}^N \to \mathbb{R}$ such that

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = df_{\mathbf{a}}(\mathbf{h}) + o(\|\mathbf{h}\|)$$

If f is differentiable at $\mathbf{a} \in U$, the *differential* $df_{\mathbf{a}}$ can be expressed in terms of *partial derivatives* of f at $\mathbf{a}$ such that

$$df_{\mathbf{a}}(\mathbf{h}) = \sum_{i=1}^{N} h_i \frac{\partial f}{\partial x_i}(\mathbf{a})$$

and as the limit, for any $\mathbf{h} \in \mathbb{R}^N$,

$$df_{\mathbf{a}}(\mathbf{h}) = \lim_{\varepsilon \to 0} \frac{f(\mathbf{a} + \varepsilon \mathbf{h}) - f(\mathbf{a})}{\varepsilon}$$

Differentiability (II)

Let $f : U \to \mathbb{R}$ be a function defined on a open subset U of $\mathbb{R}^N$.

Definition

The function f is said to be of class C^k on U if all its partial derivatives up to order k exist and are continuous on U.

If $f \in C^1$, we say it is *continuously differentiable*.

Theorem

If f admits partial derivatives at every point in a neighborhood of $\mathbf{a}$, and if the functions $\frac{\partial f}{\partial x_i}$ are continuous at $\mathbf{a}$, then f is differentiable at $\mathbf{a}$.

The converse is not true.

Example: the function $f(x_1, x_2) = |x_1 x_2|$ is differentiable at $(0,0)$ but its partial derivatives do not exist everywhere around the origin.

Gradient of a function

From now on, we assume that $f : \mathbb{R}^N \to \mathbb{R}$ is at least of class C^1.

For $a \in \mathbb{R}^N$, the vector

$$\nabla f(a) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{a}) \\ \dfrac{\partial f}{\partial x_2}(\mathbf{a}) \\ \vdots \\ \dfrac{\partial f}{\partial x_N}(\mathbf{a}) \end{bmatrix} \in \mathbb{R}^N$$

is called the gradient of f at point $\mathbf{a}$.

Remark The differential of f at a reads $df_{\mathbf{a}}(\mathbf{h}) = \nabla f(\mathbf{a})^\top \mathbf{h}$.

Hessian of a function

Let $f : \mathbb{R}^N \to \mathbb{R}$ be of class C^2.

For $\mathbf{a} \in \mathbb{R}^N$, the matrix

$$\nabla^2 f(\mathbf{a}) = \begin{bmatrix} \frac{\partial^2 f}{\partial {x_1}^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial {x_2}^2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_N}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_N \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial {x_N}^2}(\mathbf{a}) \end{bmatrix} \in \mathbb{R}^{N \times N}$$

is called the Hessian matrix of f at point $\mathbf{a}$ – also noted as Hess $f(\mathbf{a})$.

Important remark: since f is of class C^2 on $\mathbb{R}^N$, Schwarz's theorem ensures $\nabla^2 f(\mathbf{a})$ is a symmetric matrix for every $\mathbf{a} \in \mathbb{R}^N$ since

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) \text{ for all } (i, j)$$

Positive definiteness

Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a symmetric matrix ($\mathbf{A}^\top = \mathbf{A}$).

Definition

The symmetric matrix $\mathbf{A}$ is said to be positive semi-definite (PSD) if for every $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0$. It is said to be positive definite (PD) if for every $\mathbf{x} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$, $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$.

We note $\mathbf{A} \succeq 0$ when it is PSD, and $\mathbf{A} \succ 0$ when it is PD.

Remarks

- $\mathbf{A} \succeq 0$ if and only if all its eigenvalues are positive (non-negative)
- $\mathbf{A} \succ 0$ if and only if all its eigenvalues are stricty positive
- Positive (semi)-definiteness of Hessian matrix is crucial in the study of optimization problems

Taylor's theorem

Amongst different versions of Taylor's theorem, this one will be useful for proofs later on.

Theorem (see e.g., Nocedal)

Suppose that $f : \mathbb{R}^N \to \mathbb{R}$ is continuously differentiable and that $\mathbf{p} \in \mathbb{R}^N$. Then for some $t \in (0,1)$,

$$f(\mathbf{x} + \mathbf{p}) = f(\mathbf{x}) + \nabla f(\mathbf{x} + t\mathbf{p})^\top \mathbf{p}$$

Moreover if f is twice continuously differentiable,

$$\nabla f(\mathbf{x} + \mathbf{p}) = \nabla f(\mathbf{x}) + \int_0^1 \nabla^2 f(\mathbf{x} + t\mathbf{p})\mathbf{p} \mathrm{d}t$$

and

$$f(\mathbf{x} + \mathbf{p}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{p} + \frac{1}{2}\mathbf{p}^\top \nabla^2 f(\mathbf{x} + t\mathbf{p})\mathbf{p}$$

for some $t \in (0,1)$.

Outline

Existence of solutions

Let $f : \Omega \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ and consider the following optimization problem

$$(\mathcal{P}) \qquad \min_{\mathbf{x} \in \Omega} f(\mathbf{x})$$

Theorem (existence in finite dimensions)

Suppose that f is continuous and that $\Omega \subset \mathbb{R}^n$. If one of the following condition is satisfied:

- *Ω is compact (i.e., closed and bounded – because of the finite dimension);*
- *Ω is closed and f is coercive (i.e., such that $f(\mathbf{x}) \xrightarrow[\|\mathbf{x}\| \to +\infty]{} +\infty$),*

then the problem $(\mathcal{P})$ admits (at least) one solution.

Remark the set $\mathbb{R}$ is both open and closed; therefore, optimization problems of the form $\min_{x \in \mathbb{R}^N} f(\mathbf{x})$ with f coercive admit at least one solution.

Outline

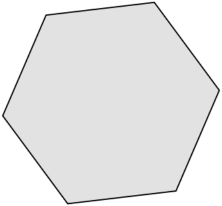
Convex sets and convex functions

convexity is a key tool to study uniqueness of solutions in optimization

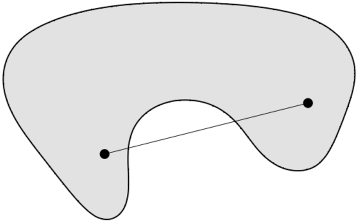
Definition (Convex set)

A set $\mathcal{C}$ is convex if for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$ and for every $\theta \in [0, 1]$, one has

$$\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2 \in \mathcal{C}$$



$\mathcal{C}_1$ is convex



$\mathcal{C}_2$ is non-convex

Convex sets and convex functions

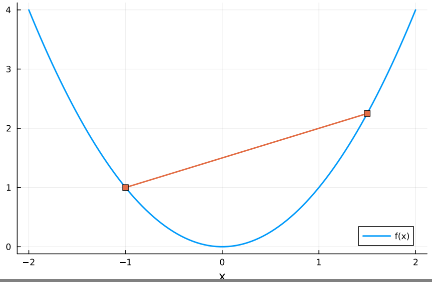
convexity is a key tool to study uniqueness of solutions in optimization

Definition (Convex function)

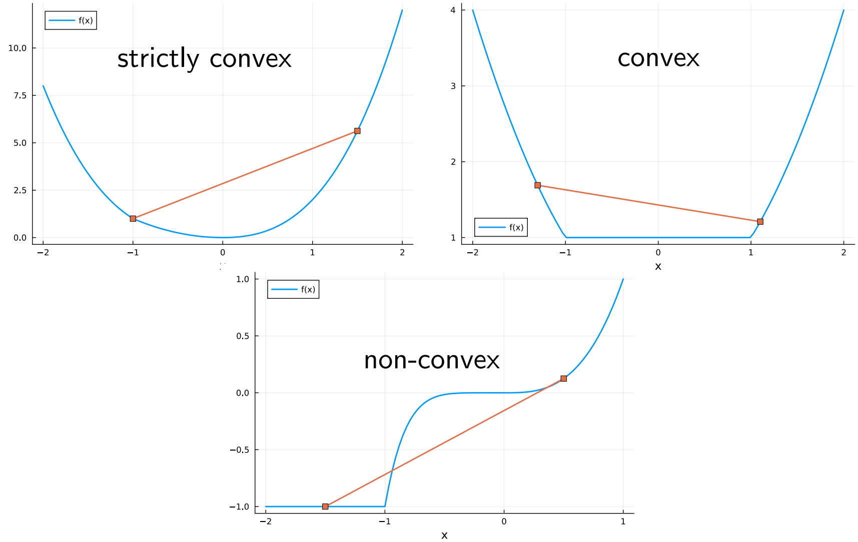
Let $\Omega \subset \mathbb{R}^N$ be convex. The function $f : \Omega \to \mathbb{R}$ is said to be convex if for every $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and for every $\theta \in [0, 1]$, one has

$$f(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2)$$

and strictly convex if the inequality is strict.



Examples of convex and non-convex functions (1-D)



Useful characterization results for convex functions

Theorem (first order)

Let $f : \mathbb{R}^N \to \mathbb{R}$ be differentiable. These statements are equivalent:

1. *f is convex on $\mathbb{R}^N$*
2. *for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$*
3. *for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, $(\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}))^\top (\mathbf{y} - \mathbf{x}) \geq 0$*

The equivalence is preserved for strict convexity, with $\mathbf{x} \neq \mathbf{y}$ and strict inequalities.

Theorem (second order)

Let $f : \mathbb{R}^N \to \mathbb{R}$ be twice differentiable. We have equivalence between

1. *f is convex*
2. *for all $\mathbf{x} \in \mathbb{R}^N$, $\nabla^2 f(\mathbf{x}) \succeq 0$, i.e. the Hessian is positive semidefinite.*

This time, the equivalence is not preserved in the strict case: if $\nabla^2 f \succ 0$ then f is strictly convex, the converse is not true.

Uniqueness in optimization

Consider the following optimization problem

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x})$$

Theorem

Suppose that f is a convex function and that $\Omega \subset \mathbb{R}^N$ is a convex set. Then:

1. *any local minimizer is a global minimizer*
2. *if f is strictly convex, there is at most one (global) minimizer.*

Proof: by contradiction (Exercise!)

More comments on convex functions

- By definition, f is (strictly) concave iff $-f$ is (strictly) convex
- Optimization problems of the form

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x})$$

where f is a convex function and $\Omega \subset \mathbb{R}^N$ is a convex set are called convex optimization problems. It is one of the most successful field in numerical optimization.

- the special (and important!) case of quadratic functions

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{p}^\top \mathbf{x} + \mathbf{r} \qquad (\text{with } \mathbf{Q} = \mathbf{Q}^\top)$$

Exercise show that f is (resp. strictly) convex iff $\mathbf{Q} \succeq 0$ (resp. $\mathbf{Q} \succ 0$). What can we say about the minimizers of the problem $\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x})$?

Outline

Optimality conditions for unconstrained optimization

We consider problems of the form

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x})$$

we assume that f has "good" properties on $\mathbb{R}^N$, i.e., f is twice continuously differentiable (or at least continuously differentiable).

Provided minimizers exist (not always guaranteed!) we will derive conditions on a point $\mathbf{x}^*$ to be a minimizer of the problem.

Conditions are of several types

- Necessary, sufficient or necessary and sufficient
- first-order (i.e. involving ∇f) or second-order (i.e. involving $\nabla^2 f$)

Necessary conditions (unconstrained case)

Theorem (First-order)

If $\mathbf{x}^\star$ is local minimizer and f is continuously differentiable in a open neighborhood of $\mathbf{x}^\star$, then $\nabla f(\mathbf{x}^\star) = \mathbf{0}$.

We call $\mathbf{x}^\star$ a stationary or critical point of f if $\nabla f(\mathbf{x}^\star) = \mathbf{0}$.

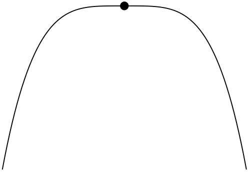
Any local minimizer must be a stationary point.

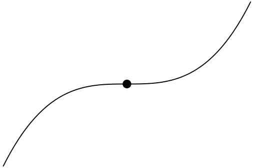
Theorem (Second order)

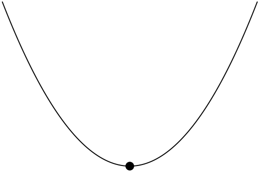
If $\mathbf{x}^\star$ is local minimizer and f is twice continuously differentiable in a open neighborhood of $\mathbf{x}^\star$, then $\nabla f(\mathbf{x}^\star) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x}^\star) \succeq 0$ (is positive semidefinite).

Exercise: prove these results by contradiction using Taylor's theorem.

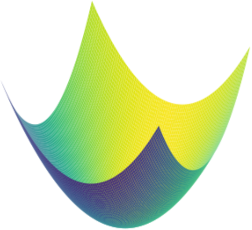
Examples: when $\nabla f(\mathbf{x}^\star) = 0$ is not enough ...

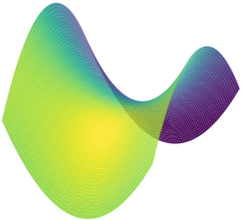






$f''(x^\star) = 0$ $\qquad$ $f''(x^\star) = 0$ $\qquad$ $f''(x^\star) > 0$





$\nabla^2 f(\mathbf{x}^\star) \succ 0$ $\qquad$ $\nabla^2 f(\mathbf{x}^\star) \succeq 0$ $\qquad$ $\nabla^2 f(\mathbf{x}^\star) \succeq 0$

Sufficient conditions (unconstrained case)

Theorem

Suppose that $\nabla^2 f$ is continuous in an open neighborhood of $\mathbf{x}^\star$ and that $\nabla f(\mathbf{x}^\star) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x}^\star) \succ 0$ (is positive definite). Then $\mathbf{x}^\star$ is a strict local minimizer of f.

Remarks

- Sufficient conditions guarantee that the minimizer is a *strict* local minimizer. (Compare with the necessary conditions)
- These sufficient conditions are not necessary: a point $\mathbf{x}^\star$ can fail to satisfy the conditions and yet be a strict minimizer.

  Example: $f(x) = x^4$; the point $x^\star = 0$ is a strict local minimizer (and global as well) but $f''(0) = 0$ shows that the Hessian is not positive definite at this point.

Exercises

- Quadratic function

  Let $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(\mathbf{x}) = x_1^2 - x_2^2$. Compute its gradient and Hessian. List critical points and their properties.

- Rosenbrock function

  Let $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(\mathbf{x}) = (1 - x_1)^2 + 5(x_2 - x_1^2)^2$.

  Does the point $[1, 1]^\top$ satisfy the necessary conditions? the sufficient conditions?

Necessary and sufficient conditions (unconstrained case)

When f is convex there is a simple characterization of optimal points.

Theorem

Suppose that $f : \mathbb{R}^N \to \mathbb{R}$ is convex. A point $\mathbf{x}^\star$ is a local minimizer (hence global) of f if and only if $\mathbf{x}^\star$ is a stationary point of f, i.e., such that $\nabla f(\mathbf{x}^\star) = \mathbf{0}$.

Remarks

- if f is strictly convex, then the theorem gives a characterization of the unique global minimizer of the problem (when it exists).
- finding points such that $\nabla f(\mathbf{x}^\star) = \mathbf{0}$ (stationary points) is the foundation for many unconstrained optimization algorithms, even in the non-convex case.
- in the next session, we'll see a very important of this result to solve a very important category of optimization problems, called least-squares problems