Session 11: Optimality conditions

Optimization and Computational Linear Algebra for Data Science

Marylou Gabrié (based on material by Léo Miolane)

Contents

1. Critical points (Unconstrained optimization)
2. Constrained optimization and Lagrange multipliers
3. Convex constrained optimization

1. Critical points (Unconstrained optimization)

Definitions

Definition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. We say that $x \in \mathbb{R}^n$ is

- a critical point of f if $\nabla f(x) = 0$,
- a *global* minimizer of f if for all $x' \in \mathbb{R}^n$, $f(x) \leq f(x')$,
- a *local* minimizer of f if there exists $\delta > 0$ such that we have $f(x) \leq f(x')$ for all x' verifying $\|x - x'\| \leq \delta$.

1.1 Local extrema

Proposition

$$x \text{ is a local minimizer of } f \implies \nabla f(x) = 0.$$

Proposition

Assume that f is convex. Then

$$\nabla f(x) = 0 \iff x \text{ is a global minimizer of } f.$$

Looking at the Hessian

Proposition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Let $x \in \mathbb{R}^n$ be a critical point of f, i.e. $\nabla f(x) = 0$.
Then, if $H_f(x)$ is positive definite (that is, if all the eigenvalues of $H_f(x)$ are strictly positive), then x is a local minimizer of f.

Looking at the Hessian

Proposition

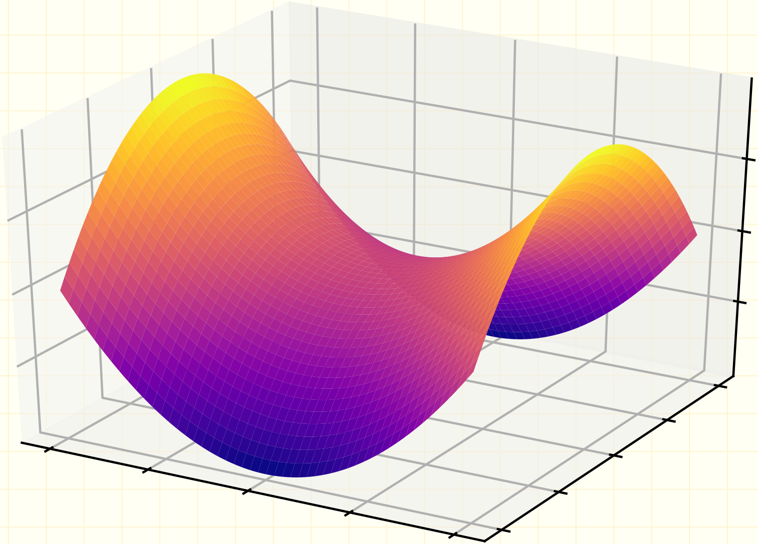
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Let $x \in \mathbb{R}^n$ be a critical point of f, i.e. $\nabla f(x) = 0$.
Then, if $H_f(x)$ is negative definite (that is, if all the eigenvalues of $H_f(x)$ are strictly negative), then x is a local maximizer of f.

1.2 Saddle points

Proposition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. Let $x \in \mathbb{R}^n$ be a critical point of f, i.e. $\nabla f(x) = 0$.
Then, if $H_f(x)$ admits strictly positive eigenvalues and strictly negative eigenvalues, then x is neither a local maximum nor a local minimum. We call x a saddle point.

1.2 Saddle points



1.3 Examples

Study the critical points of $f(x, y) = x^2 + xy^2 - x + 1$.

1.3 Examples

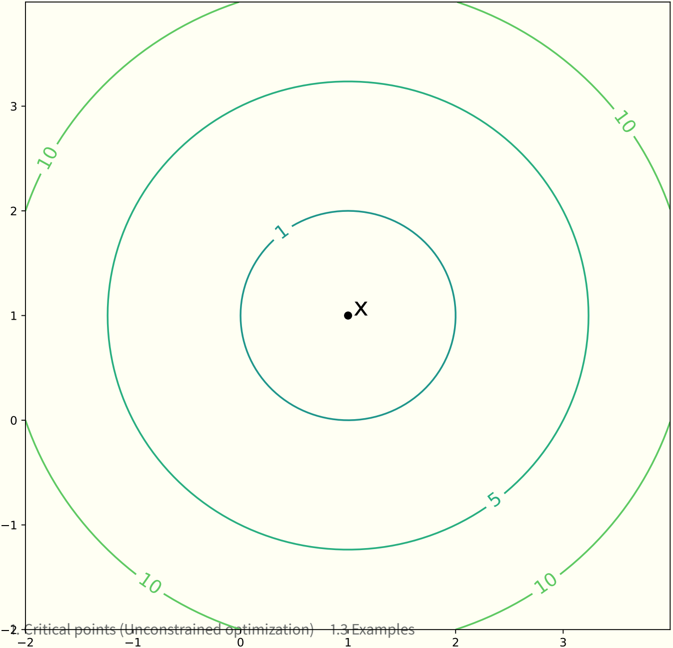
Study the critical points of $f(x, y) = x^2 + xy^2 - x + 1$.

Critical points' critical points

- Global minimizer $\implies$ local minimizer $\implies$ critical point.

- Critical point + positive definite Hessian $\implies$ local minimizer.

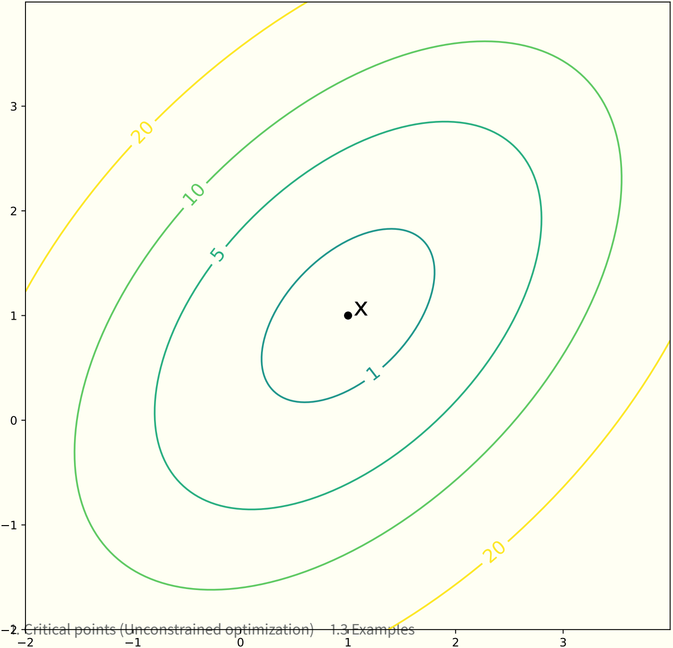
Example: Hessian at critical point



The eigenvalues of the Hessian at x are

1. $\begin{cases} \lambda_1 = 1 \\ \lambda_2 = -1 \end{cases}$
2. $\begin{cases} \lambda_1 = 1 \\ \lambda_2 = 1 \end{cases}$
3. $\begin{cases} \lambda_1 = 2 \\ \lambda_2 = 1 \end{cases}$
4. $\begin{cases} \lambda_1 = -1 \\ \lambda_2 = -1 \end{cases}$

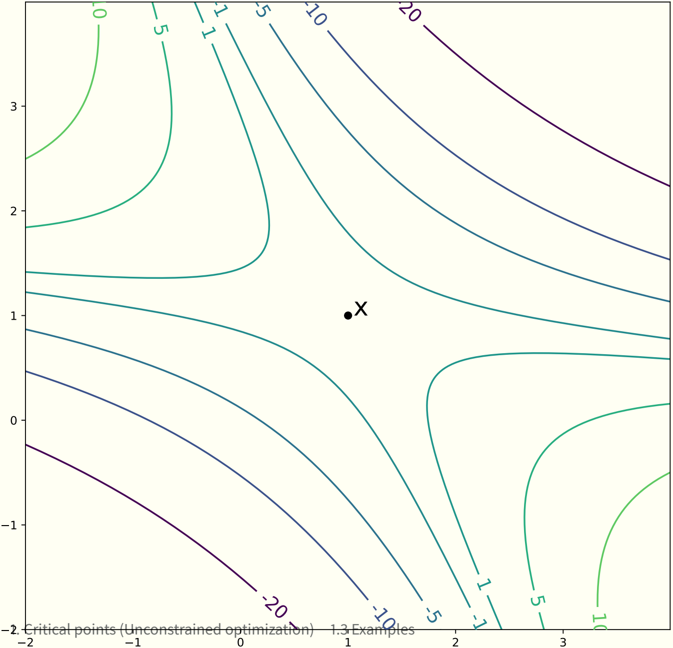
Example: Hessian at critical point



The eigenvalues of the Hessian at x are

1. $\begin{cases} \lambda_1 = 1 \\ \lambda_2 = -3 \end{cases}$
2. $\begin{cases} \lambda_1 = 2 \\ \lambda_2 = 2 \end{cases}$
3. $\begin{cases} \lambda_1 = 1 \\ \lambda_2 = 3 \end{cases}$
4. $\begin{cases} \lambda_1 = -1 \\ \lambda_2 = -1 \end{cases}$

Example: Hessian at critical point



The eigenvalues of the Hessian at x are

1. $\begin{cases} \lambda_1 = 1 \\ \lambda_2 = -3 \end{cases}$
2. $\begin{cases} \lambda_1 = 2 \\ \lambda_2 = 2 \end{cases}$
3. $\begin{cases} \lambda_1 = 1 \\ \lambda_2 = 3 \end{cases}$
4. $\begin{cases} \lambda_1 = -1 \\ \lambda_2 = -1 \end{cases}$

2. Constrained optimization

General formulation

Constrained optimization problems take the form:

$$
\begin{array}{lll}
\text{minimize} & f(x) & \text{with variable } x \in \mathbb{R}^n. \\
\\
\text{subject to} & g_i(x) \leq 0, & i = 1, \ldots, m \\
 & h_i(x) = 0, & i = 1, \ldots, p,
\end{array}
$$

Example:

2.1 Feasible points

Definition

A point $x \in \mathbb{R}^n$ is *feasible* if it satisfies all the constraints:
$g_1(x) \leq 0, \ldots, g_m(x) \leq 0$ and $h_1(x) = 0, \ldots, h_p(x) = 0$.

Example:

Question

If x is a solution to

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array}$$

do we have $\nabla f(x) = 0$?

2.2 First order optimality condition

Consider the following problem:

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad g_(x) \leq 0 \qquad (\star)$$

Case 1: “x inside”

- There exists $\delta > 0$ such that $B(x, \delta) \subset F$.

Case 2: "x on boundary"

- The constraint is active at x: $g(x) = 0$.

Cases summary

Case 1: "x inside" $\Rightarrow \nabla f(x) = 0$.

Case 2: active constraint, $g(x) = 0 \Rightarrow \nabla f(x) = -\lambda \nabla g(x)$ for $\lambda \geq 0$

.

First order optimality condition

Theorem

If x is a solution and if $\nabla h_1(x), \dots, \nabla h_p(x), \{\nabla g_i(x) \mid g_i(x) = 0\}$ are linearly independent, then there exists $\lambda_1, \dots, \lambda_m \geq 0$ and $\nu_1, \dots, \nu_p \in \mathbb{R}$ such that:

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0.$$

Moreover, for all $i \in \{1, \dots, m\}$, if $g_i(x) < 0$ then $\lambda_i = 0$.

Example

Let $u \in \mathbb{R}^n$ be a non-zero vector.

$$\begin{array}{ll} \text{Minimize} & \langle x, u \rangle \\ \text{subject to} & \|x\|^2 = 1. \end{array}$$

Example

Let $u \in \mathbb{R}^n$ be a non-zero vector.

$$\begin{array}{ll} \text{Minimize} & \langle x, u \rangle \\ \text{subject to} & \|x\|^2 = 1. \end{array}$$

3. Convex constrained optimization

General formulation

We say that the constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array}$$

is convex when $f, g_1, \ldots, g_m$ are convex and $h_1, \ldots, h_p$ are affine.

Karush-Kuhn-Tucker Theorem

Theorem (KKT)

Assume that the problem is convex and that there exists a feasible point x_0 such that $g_i(x_0) < 0$ for all i.
Then x is a solution if and only if x is feasible and there exists $\lambda_1, \dots, \lambda_m \geq 0, \nu_1, \dots, \nu_p \in \mathbb{R}$ such that:

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0. \\ \lambda_i g_i(x) = 0, \ \text{ for all } \ i \in \{1, \dots, p\}. \end{cases}$$

Example: Ridge regression

$$
\begin{array}{ll}
\text{minimize} & \|Ax - y\|^2 \\
\text{subject to} & \|x\|^2 \le r^2.
\end{array}
$$

Example: Ridge regression

$$\begin{array}{ll} \text{minimize} & \|Ax - y\|^2 \\ \text{subject to} & \|x\|^2 \leq r^2. \end{array}$$

Example

Let $u, v \in \mathbb{R}^n$ such that $\|v\| = 1$. Solve:

$$\begin{array}{ll} \text{minimize} & \|x-u\|^2 \\ \text{subject to} & x \perp v. \end{array}$$

Example

Questions?

Questions?