Session 11: Optimality conditions

Optimization and Computational Linear Algebra for Data Science

Léo Miolane

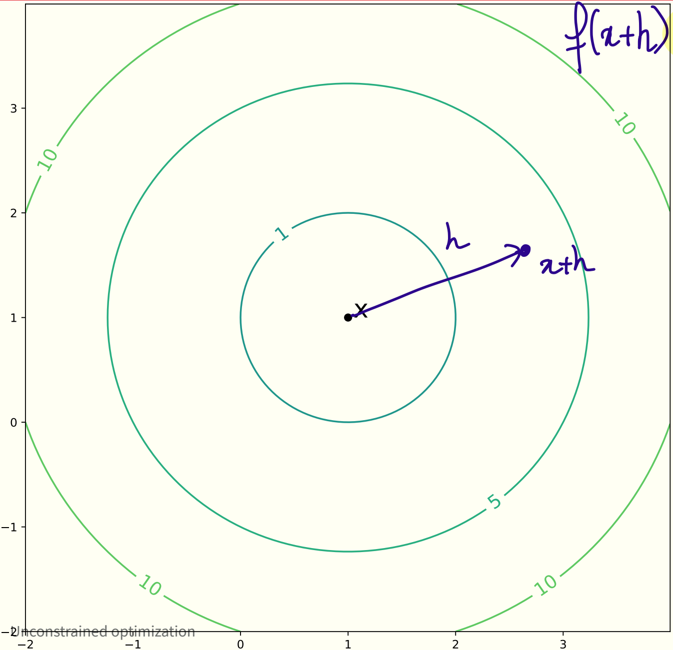
Contents

Unconstrained optimization

Questions about the video?

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

Hessian at a critical point

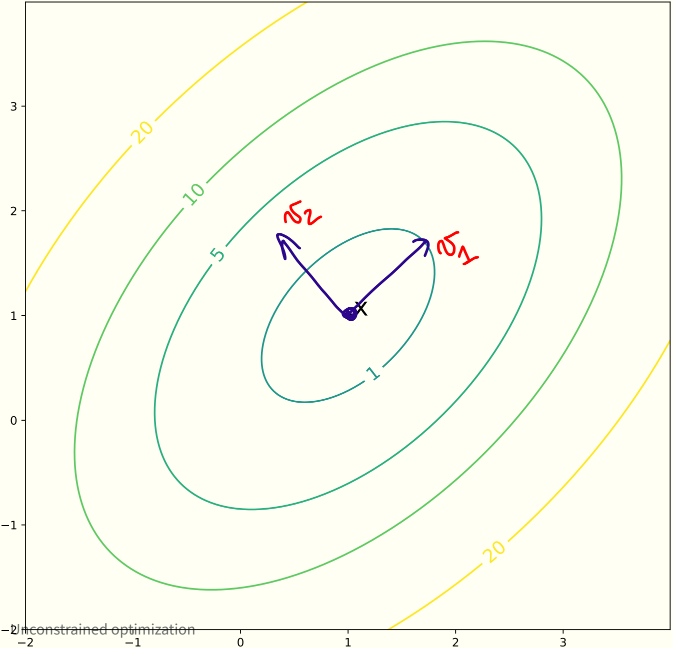


$$f(x+h) \approx f(x) + \cancel{\nabla f(x)\cdot h}$$
$$\frac{1}{2} h^T H_f(x)\, h$$

The eigenvalues of the Hessian at x are

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

Hessian at a critical point

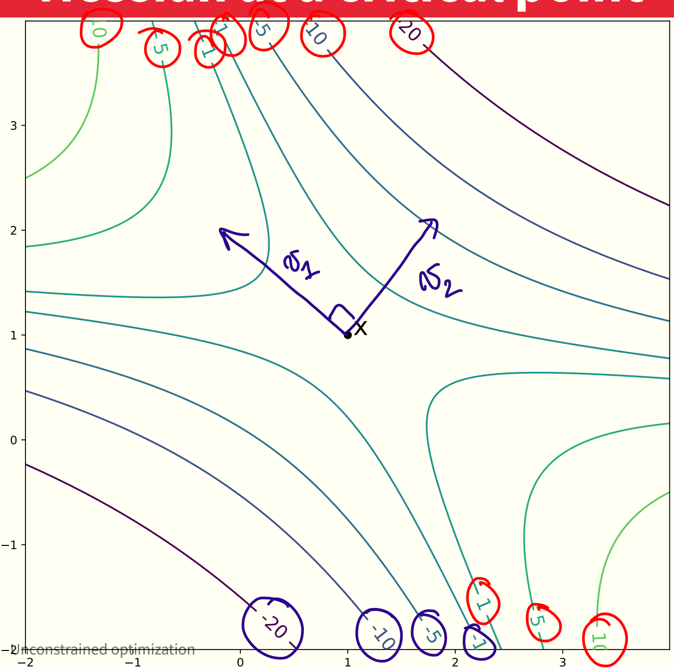


$h^T H_f(x) h$

The eigenvalues of the Hessian at x are

1. $\cancel{\begin{cases} \lambda_1 = 1 \\ \lambda_2 = -3 \end{cases}}$
2. $\cancel{\begin{cases} \lambda_1 = 2 \\ \lambda_2 = 2 \end{cases}}$
3. $\begin{cases} \lambda_1 = 1 \\ \lambda_2 = 3 \end{cases}$ $\quad v_1, v_2$
4. $\cancel{\begin{cases} \lambda_1 = -1 \\ \lambda_2 = -1 \end{cases}}$

Hessian at a critical point



The eigenvalues of the Hessian at x are

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

Constrained optimization

General formulation

Constrained optimization problems take the form:

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

constraints.

with variable $x \in \mathbb{R}^n$.

Example: minimize $x^T A x$ $f(x)$

subject to $\|x\|^2 = 1$

$h(x) = \|x\|^2 - 1$.

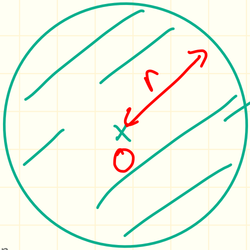
$\|x\|^2 - 1 = 0$

Feasible points

Definition

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

Example: minimize $\|Ax - y\|^2$
subject to $\|x\|^2 - r^2 \le 0$



set of feasible points.

"feasible set"

Question

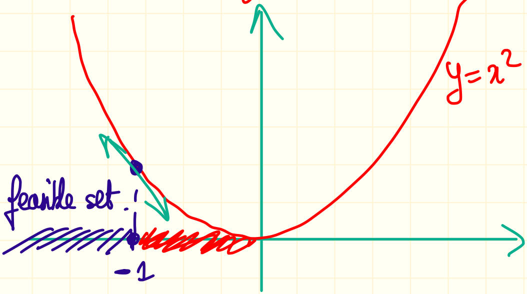
If x is a solution to

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

$f'(x) = 2x$

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

Unfortunatly not in general: ex: minimizing $f(x) = x^2$ subject to $x+1 \leq 0$



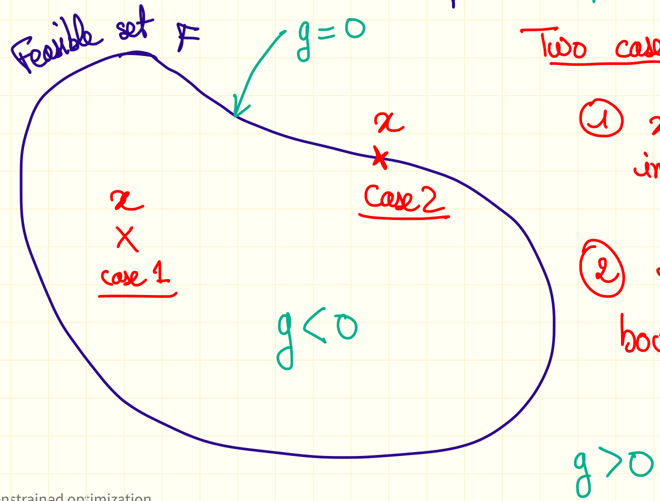
minimizer is $x^* = -1$

however $f'(x^*) = -2 \neq 0$

First order optimality condition

Consider minimize $f(x)$ subject to $g(x) \leq 0$ $(*)$

Let x be a solution of $(*)$ (provided it exists)



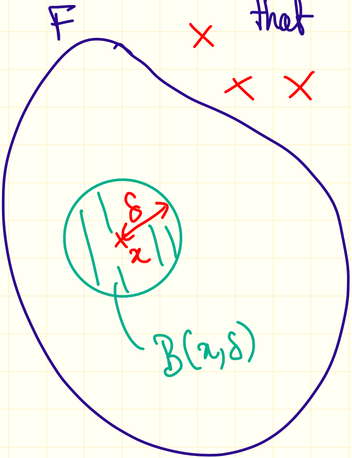
Two cases:

① x is strictly inside of F: $g(x)<0$

② x lies on the boundary of F
$g(x)=0$.

First order optimality condition

Case 1: In that case there exists $\delta > 0$ such that $\underbrace{B(x,\delta)}_{\{x' \mid \|x'-x\| \leq \delta\}} \subset F$



- Then, for all $x' \in B(x,\delta)$ we have $x' \in F$, hence

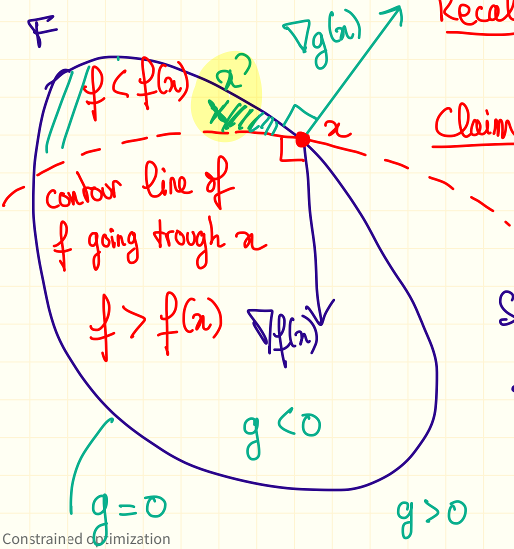
$$f(x) \leq f(x')$$ because x solution to (*)

$\longrightarrow$ x is a local minimizer of f

$\longrightarrow$ $\boxed{\nabla f(x) = 0}$.

First order optimality condition

Case 2 : $g(x) = 0$, "the constraint is active at x"



Recall that the gradient is orthogonal to the contour lines.

Claim: There exists some $\lambda \geq 0$ such that $\boxed{\nabla f(x) = -\lambda \nabla g(x)}$

Suppose not, then there exists $x' \in F$ such that $f(x') < f(x)$

contradiction

First order optimality condition

Conclusion of the 2 cases: there exists $\lambda \geq 0$ such that

$$\begin{cases} \nabla f(x) + \lambda \nabla g(x) = 0 \\ \lambda = 0 \quad \text{if} \quad g(x) < 0. \end{cases}$$

λ is called a Lagrange multiplier.

First order optimality condition

Theorem

If x is a solution and if $\nabla h_1(x), \ldots, \nabla h_p(x), \overbrace{\{\nabla g_i(x) \mid g_i(x) = 0\}}^{\text{gradients of active ineq. constraints}}$ are linearly independent, then there exists $\lambda_1, \ldots, \lambda_m \geq 0$ and $\nu_1, \ldots, \nu_p \in \mathbb{R}$ such that:

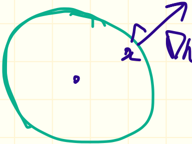
$$\boxed{\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0.}$$

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

$$h_i(x) = 0 \iff \begin{cases} h_i(x) \leq 0 \\ -h_i(x) \leq 0 \end{cases}$$

$\|x\|^2 = 1$

$h(x) = \|x\|^2 - 1$



$\nabla h(x) = 2x$

Example

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

Minimize $\langle x, u \rangle = f(x)$ (*)
subject to $\|x\|^2 = 1.$

$h(x) = \|x\|^2 - 1 = 0$

Let x be a solution (assuming it exists)

By the theorem, there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(x) + \lambda \nabla h(x) = 0$$

$$u + \lambda 2x = 0$$

- $\lambda \neq 0$ because $u \neq 0$, hence $x = -\frac{1}{2\lambda} u$
- Since $\|x\| = 1$, $1 = \|x\| = \frac{1}{2|\lambda|} \|u\|$

$\longrightarrow |\lambda| = \frac{1}{2}\|u\| \longrightarrow \lambda = \pm \frac{1}{2}\|u\|$

Example

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

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

$$\left\langle \frac{u}{\|u\|}, u \right\rangle = \frac{\langle u, u \rangle}{\|u\|}$$

if x is a solution of $(*)$

then $\longrightarrow$ $x = +\dfrac{u}{\|u\|}$ or $x = -\dfrac{u}{\|u\|}$

$$f\left(\frac{u}{\|u\|}\right) = \|u\| > -\|u\| = f\left(\frac{-u}{\|u\|}\right)$$

$\longrightarrow$ $\boxed{x = \dfrac{-u}{\|u\|} \text{ is the only solution}}$

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.

$h_i(x) = \langle a_i, x \rangle + b_i$

for some $a_i \in \mathbb{R}^n$

$b_i \in \mathbb{R}$.

Ex:

$$\begin{array}{ll} \text{minimize} & \|Ax - y\|^2 \\ \text{s.t.} & \|x\|^2 - r^2 \leq 0. \end{array}$$

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 $\lambda_1, \ldots, \lambda_m \geq 0, \nu_1, \ldots, \nu_p \in \mathbb{R}$ such that:

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

$\nabla L_{\nu,\lambda}(x) = 0$

x minimizes $L_{\nu,\lambda}$

$\lambda_i = 0$ if $g_i(x) < 0$

For $\lambda_i \geq 0$, $\nu_i \in \mathbb{R}$, define the "Lagrangian"

$$\boxed{L_{\nu,\lambda}(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)}$$

$L_{\nu,\lambda}$ is convex as a sum of convex functions.

Example: Ridge regression

(∗) minimize $\|Ax - y\|^2$
subject to $\|x\|^2 \le r^2$.) $g(x) = \|x\|^2 - r^2 \le 0$

$$L_\lambda(x) = \|Ax - y\|^2 + \lambda \|x\|^2 \underbrace{- \lambda r^2}_{\text{constant}}$$

From the theorem

x solution of (∗) $\iff$ there exists some λ such that x minimizes L_λ

+ ...

$r \nearrow$ $\lambda \searrow$

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:

$$(*) \quad \begin{array}{ll} \text{minimize} & \|x-u\|^2 = f(x) \quad \text{convex} \\ \text{subject to} & x \perp v. \end{array}$$

$h(x) = \langle x, v \rangle = 0$

By the KKT Theorem

x is solution of $(*)$ $\iff$ There exists $\lambda \in \mathbb{R}$ such that

$$\begin{cases} \nabla f(x) + \lambda \nabla h(x) = 0 \\ h(x) = 0 \end{cases}$$

$\iff$ x, λ are solutions to

$$\begin{cases} 2x - 2u + \lambda v = 0 \\ \langle x, v \rangle = 0 \end{cases}$$

Example

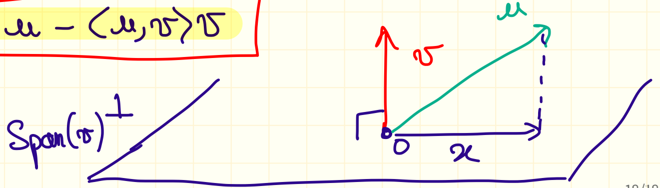
Let's solve

$$\begin{cases} 2x - 2u + \lambda v = 0 \\ \langle x, v \rangle = 0 \end{cases} \iff \begin{cases} 2\overset{=0}{\cancel{\langle x, v \rangle}} - 2\langle u, v \rangle + \lambda \overset{=1}{\|v\|^2} = 0 \\ 2x - 2u + \lambda v = 0 \\ \langle x, v \rangle = 0 \end{cases}$$

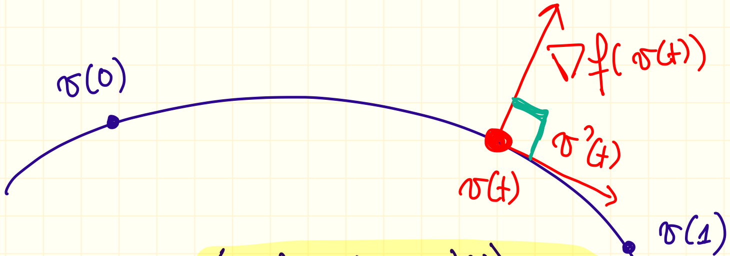
$$\iff \begin{cases} \lambda = 2\langle u, v \rangle \\ x = u - \langle u, v \rangle v \end{cases}$$

Conclusion: the problem (*) admits a unique solution

$$\boxed{x = u - \langle u, v \rangle v}$$



Questions?



Show $\langle \nabla f(\gamma(t)), \gamma'(t) \rangle = 0$

Since $\gamma(t)$ is on the contour line for all t, $f(\gamma(t)) = \text{cste}$ for all t.

$$\hookrightarrow \frac{d}{dt}\Big(f(\underbrace{\gamma(t)}_{\in \mathbb{R}^n}) \Big) = 0 \quad , 0 = \frac{d}{dt}\Big(f(\gamma(t)) \Big) = \overbrace{\langle \nabla f(\gamma(t)), \gamma'(t) \rangle}^{=0}$$

Questions?

$f, g : \mathbb{R} \longrightarrow \mathbb{R}$

$$\boxed{\frac{d}{dt}\left(f(g(t))\right) = f'(g(t)) \cdot g'(t)}$$