Nonlinear Optimization: Optimality conditions

INSEAD, Spring 2006

Jean-Philippe Vert

Ecole des Mines de Paris

`Jean-Philippe.Vert@mines.org`

Outline

- General definitions
- Unconstrained problems
- Convex optimization
- Equality constraints
- Equality and inequality constraints

General definitions

Local and global optima

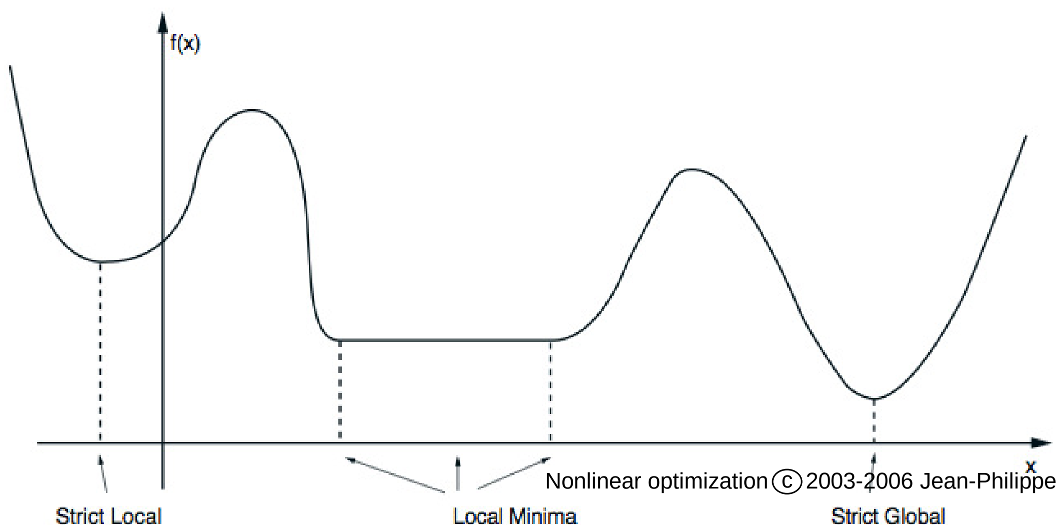
- (Strict) *global* minimum:

$$x^* \text{ s.t. } f(x^*) < (\leq) f(x), \quad \forall x \in \mathcal{X} \,.$$

- (Strict) *local* minimum:

$$x^* \text{ s.t. } f(x^*) < (\leq) f(x), \quad \forall x \in \mathcal{X} \bigcap \mathcal{N}(x^*) \,,$$

where $\mathcal{N}$ is a *neighborhood* of x^* (e.g., open ball).



Derivatives

A function

$$f : \mathbb{R}^n \to \mathbb{R}$$

is called *(Frechet) differentiable* at $x \in \mathbb{R}^n$ if there exists a vector $\nabla f(x)$, called the *gradient* of f at x, such that:

$$f(x+u) = f(x) + u^\top \nabla f(x) + o(\| u \|) \ .$$

In that case we have:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)^\top \ .$$

Second derivative

If each component of ∇f is itself differentiable, then f is called *twice differentiable* and the *Hessian* of f at x is the symmetric $n \times n$ matrix $\nabla^2 f$ with entries:

$$\left[\nabla^2 f\right]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \; .$$

In that case we have the following second-order expansion of f around x:

$$f(x+u) = f(x) + u^\top \nabla f(x) + \frac{1}{2} u^\top \nabla^2 f(x) u + o\left(\| u \|^2\right) \; .$$

Descent direction

For any differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $x \in \mathbb{R}^n$, the set of *descent directions* is the set of vectors:

$$\mathcal{D}_x = \left\{ d \in \mathbb{R}^n \;:\; d^\top \nabla f(x) < 0 \right\} .$$

If d is a descent direction of f at x, then there exists a scalar ϵ_0 such that

$$f(x + \epsilon d) < f(x), \quad \forall \epsilon \in (0, \epsilon) .$$

Feasible direction

At a feasible point x, a *feasible* direction $d \in \mathbb{R}^n$ is a direction such that $x + \epsilon d$ *is feasible* for sufficiently small $\epsilon > 0$. The set of feasible directions is formally defined as:

$$\mathcal{F}_x = \{d \in \mathbb{R}^n \ : \ d \neq 0 \text{ and } \exists \epsilon_0 > 0, \forall \epsilon \in (0, \epsilon_0), x + \epsilon d \in \mathcal{X}\} \ .$$

Examples

- $\mathcal{X} = \mathbb{R}^n \implies \mathcal{F}_x = \mathbb{R}^n$.
- $\mathcal{X} = \{x \ : \ Ax + b = 0\} \implies \mathcal{F}_x = \{d \ : \ Ad = 0\}$.

Optimality conditions

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{X} \end{array}$$

- a point $x \in \mathcal{X}$ is called *feasible*
- *How do we recognize a solution to a nonlinear optimization problem?*
- An *optimality condition* is a condition x must fulfill to be the solution (usually *necessary* but *not sufficient*).

Why optimality conditions?

- When solved, the conditions provide a set of minima candidates (although not easy in practice)
- Useful to design (e.g., stopping criterion) and analyse (e.g., convergence) optimization algorithms
- Useful for further analysis (e.g., sensitivity analysis in microeconomics)

A general optimality condition

A general necessary condition for a feasible point x to be a *local minimum* is that no little move from x in the feasible set decreases the objective function, i.e., that no feasible direction be a descent direction:

$$\mathcal{D}_x \cap \mathcal{F}_x = \emptyset \ .$$

We will now see how this principle translates in different contexts:

- unconstrained problems : $\mathcal{D} = \emptyset$,
- equality constraints : Lagrange theorem,
- equality/inequality constraints : KKT conditions.

Unconstrained optimization

First-order condition

Consider the unconstrained optimization problem:

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{R}^n \, . \end{array}$$

Théorème 1 *If x^* is a local minimum of f, and if f is differentiable in x^*, then:*

$$\nabla f\left(x^*\right) = 0 \, .$$

Proof

For a direction $d \in \mathbb{R}^n$, we have:

$$d^\top \nabla f(x^*) = \lim_{\epsilon \to 0} \frac{f(x^* + \epsilon d) - f(x^*)}{\epsilon} \geq 0 \, .$$

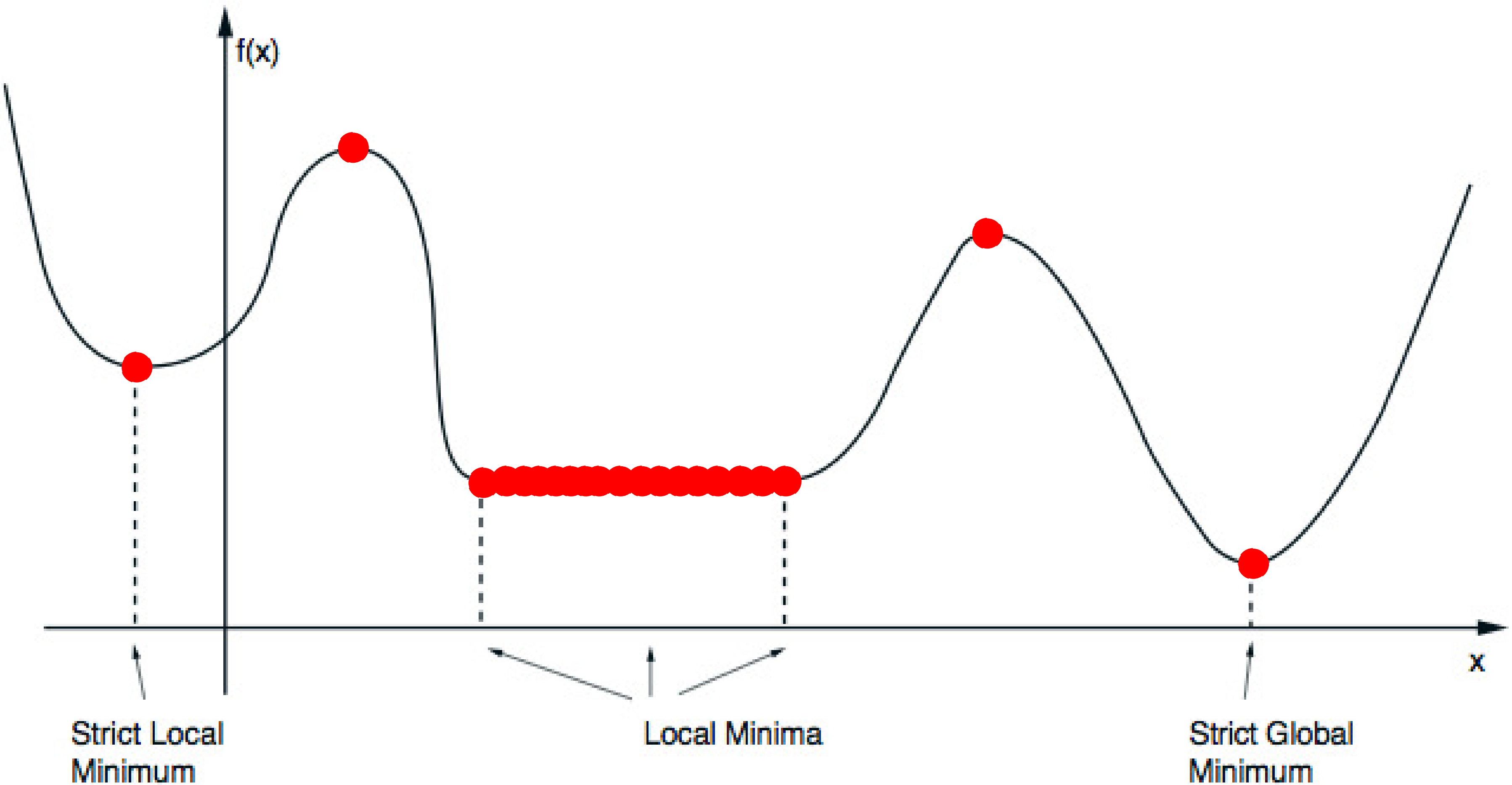
Similarly, for the direction $-d$, we obtain $-d^\top \nabla f(x) \geq 0$, therefore:

$$\forall d \in \mathbb{R}^n, \quad d^\top \nabla f(x^*) = 0 \, .$$

This shows that $\nabla f(x^*) = 0$. $\square$

Limits of first-order conditions

First-order conditions only detect *stationary points*



Positive (semi-)definite matrices

Let A be a *symmetric* $n \times n$ matrix.

- The eigenvalues of A are real.
- A is called *positive definite* (denoted $A \succ 0$) if all eigenvalues are *positive*, or equivalently:

$$x^\top A x > 0 \,, \quad \forall x \in \mathbb{R}^n, x \neq 0 \,.$$

- A is called *positive semidefinite* (denoted $A \succeq 0$) if all eigenvalues are *non-negative*, or equivalently:

$$x^\top A x \geq 0 \,, \quad \forall x \in \mathbb{R}^n \,.$$

Second order conditions

Théorème 2 *If x^* is a local minimum of f, and if f is twice differentiable in x^*, then:*

$$\nabla f\left(x^*\right) = 0 \quad \text{and} \quad \nabla^2 f\left(x^*\right) \succeq 0 \;.$$

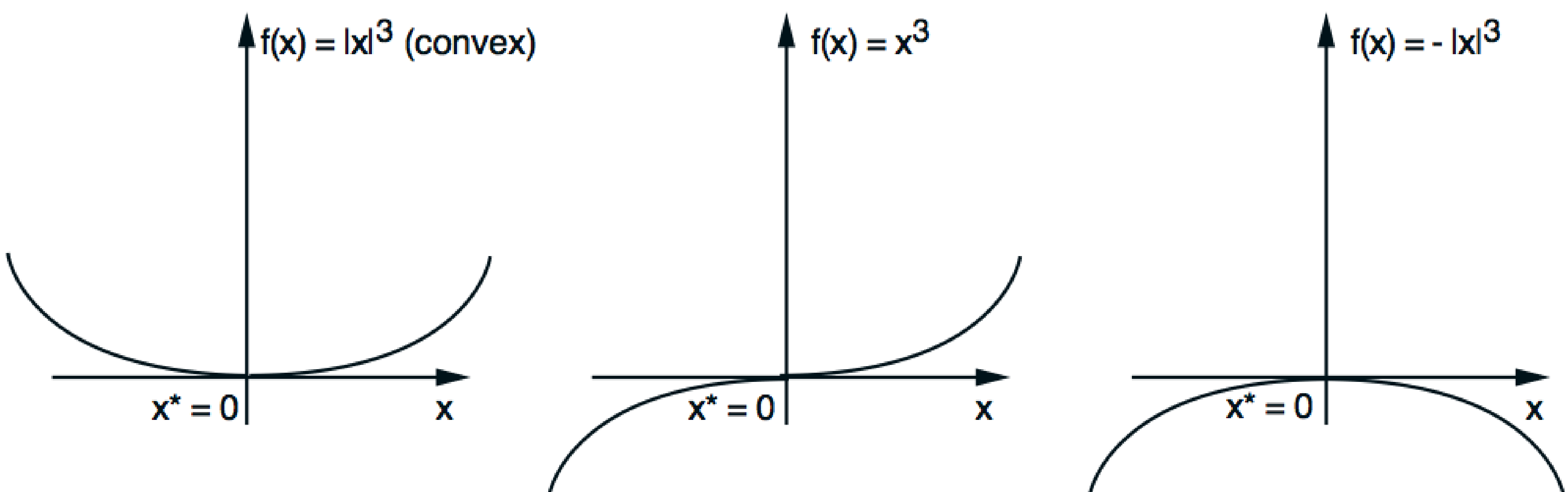
Conversely, if x^* satisfies:

$$\nabla f\left(x^*\right) = 0 \quad \text{and} \quad \nabla^2 f\left(x^*\right) \succ 0 \;,$$

then x^* is a strict *local minimum of f.*

Remark

- There may be points that satistfy the necessary first- and second-order conditions, but which are not local minima.
- There may be points that are local minima, but which do not satisfy the first- and second-order sufficient conditions.



Proof

Remember the Taylor expansion around x:

$$f(x+u) = f(x) + u^\top \nabla f(x) + \frac{1}{2} u^\top \nabla^2 f(x) u + o\left(\| u \|^2\right) .$$

At a local minimum x^* the first-order condition $\nabla f(x) = 0$ holds, and therefore for any direction $d \in \mathbb{R}^n$:

$$0 \leq \frac{f\left(x^* + \epsilon d\right) - f\left(x^*\right)}{\epsilon^2} = \frac{1}{2} d^\top \nabla^2 f(x^*) d + \frac{o\left(\epsilon^2\right)}{\epsilon^2} .$$

Taking the limit for $\epsilon \to 0$ gives $d^\top \nabla^2 f(x^*) d$ for any $d \in \mathbb{R}^n$, and therefore $\nabla^2 f(x^*) \succeq 0$.

Proof (cont.)

Conversely suppose that x^* is such that $\nabla f\left(x^*\right) = 0$ and $\nabla^2 f\left(x^*\right) \succ 0$. Let $\lambda > 0$ be the smallest eigenvalue of $\nabla^2 f\left(x^*\right)$, then we have:

$$d^\top \nabla^2 f(x^*) d \geq \lambda \| \, d \, \|^2 \,, \quad \forall d \in \mathbb{R}^d \,.$$

The Taylor expansion therefore gives for all d:

$$\begin{aligned} f\left(x^* + d\right) - f\left(x^*\right) &= \frac{1}{2} d^\top \nabla^2 f\left(x^*\right) d + o\left(\| \, d \, \|^2\right) \\ &\geq \frac{\lambda}{2} \| \, d \, \|^2 + o\left(\| \, d \, \|^2\right) \\ &= \left(\frac{\lambda}{2} + \frac{o\left(\| \, d \, \|^2\right)}{\| \, d \, \|^2} \right) \| \, d \, \|^2 \quad \square \end{aligned}$$

Summary

- $\nabla f(x) = 0$ defines a *stationary point* (including but not limited to local and global minima and maxima).
- If x^* is a stationary point and $\nabla^2 f\left(x^*\right) \succ 0$ (resp. $\prec 0$) and x^* is a *local minimum* (resp. maximum).
- If $\nabla^2 f\left(x^*\right)$ has *strictly positive and negative eigenvalues* then x^* is neither a local minimum nor a local maximum.

Example

$$f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2 - x_2 + 1 \ .$$

f is infinitely differentiable. Its gradient and Hessian are:

$$\nabla f(x_1, x_2) = \begin{pmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{pmatrix} \ ,$$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{pmatrix} \ .$$

Example (cont.)

There are two stationary points: $x_a = (1, -1)^\top$ and $x_b = (2, -3)^\top$. The corresponding Hessian are:

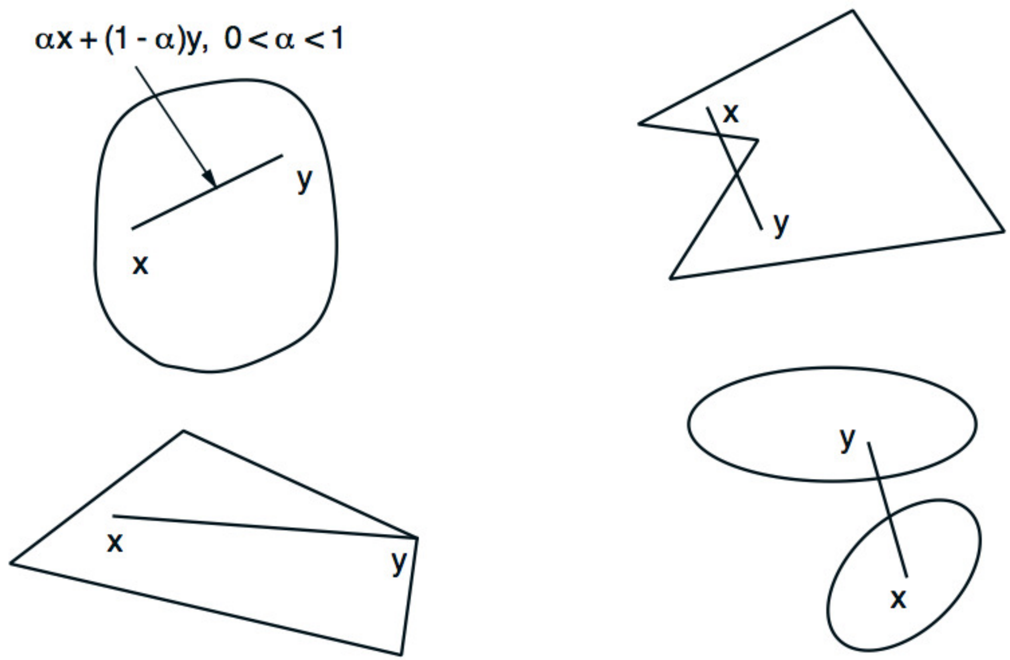
$$\nabla^2 f(x_a) = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(x_b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

- $\det\left(\nabla^2 f(x_a)\right) = -1$ so the Hessian has a negative and a positive eigenvalue: x_a is neither a local maximum nor a local minimum
- $\nabla^2 f(x_b) \succ 0$ so x_b is a local minimum.

Convex optimization

Convex set

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \Longrightarrow \theta x_1 + (1-\theta) x_2 \in C$$



Convex Sets

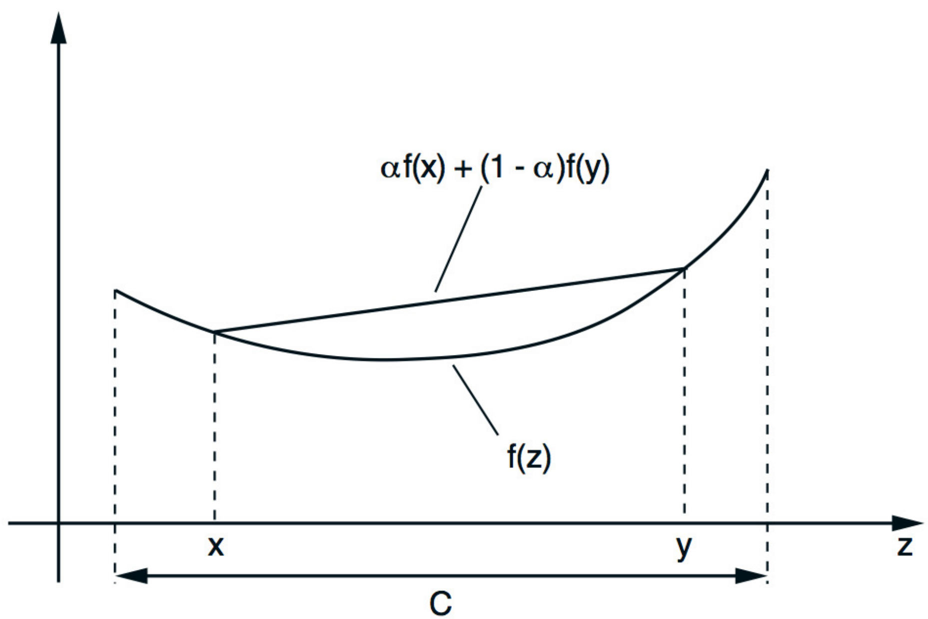
Nonconvex Sets

Convex function

If C is a convex set, then $f : C \to \mathbb{R}$ is called *convex* if

$$\begin{cases} x_1, x_2 \in C \\ 0 \leq \theta \leq 1 \end{cases} \implies f\left(\theta x_1 + (1-\theta)x_2\right) \leq \theta f\left(x_1\right) + (1-\theta) f\left(x_2\right) \ .$$

A function is called *concave* is $-f$ is convex. It is *strictly convex* is the inequality is strict for $x_1 \neq x_2$ and $\theta \in (0,1)$.



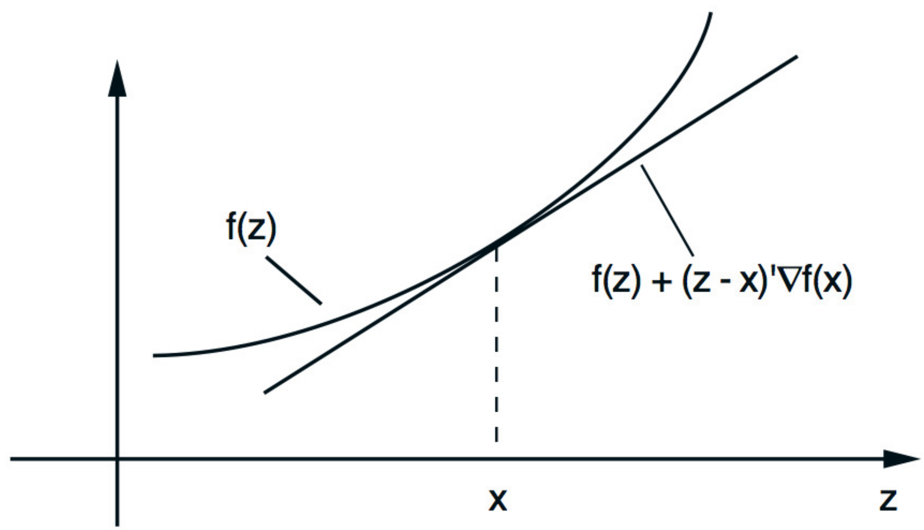
Examples on $\mathbb{R}$

- Convex:
  - affine: $f(x) = ax + b$ for any $a, b \in \mathbb{R}$.
  - exponential: $f(x) = \exp(ax)$ for any $a \in \mathbb{R}$.
  - powers: x^α for $x > 0$ and $\alpha \geq 1$ or $\alpha \leq 0$.
- Concave:
  - affine: $f(x) = ax + b$ for any $a, b \in \mathbb{R}$.
  - logarithm: $f(x) = \log(x)$ for $x > 0$.
  - powers: x^α for $x > 0$ and $0 \leq \alpha \leq 1$.

First-order convexity condition

Let f be defined over a convex open set C. If f is differentiable, then f is convex if and only if:

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x), \quad \forall x, y \in C.$$



Implication: $\nabla f(x^*) = 0 \implies x^*$ *global minimum.*

Second-order convexity condition

Let f be defined over a convex open set C. If f is twice differentiable, then f is convex if and only if:

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in C.$$

If $\nabla^2 f(x) \succ 0$ for all $x \in C$, then f is *strictly* convex.

Example

- *Quadratic function*:

$$\begin{aligned} f(x) &= (1/2)x^\top Px + q^\top x + b\,, \\ \nabla f(x) &= Px + q\,, \\ \nabla^2 f(x) &= P\,, \end{aligned}$$

  is convex if and only if $P \succeq 0$.

- *Least-squares objective*:

$$\begin{aligned} f(x) &= \| Ax - b \|_2^2\,, \\ \nabla f(x) &= 2A^\top (Ax - b)\,, \\ \nabla^2 f(x) &= 2A^\top A\,, \end{aligned}$$

  is *always convex*.

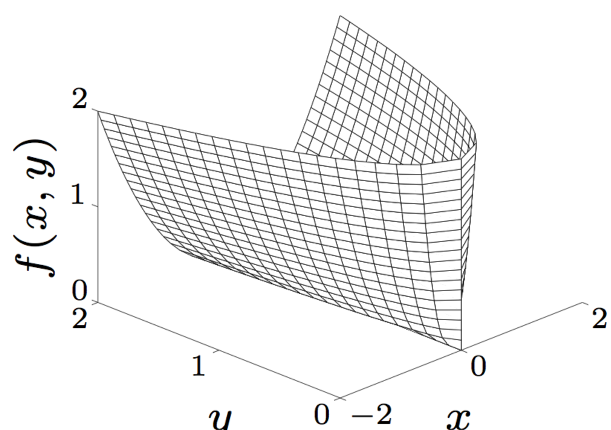
Example

The *quadratic-over-linear* function:

$$f(x,y) = \frac{x^2}{y}\,, \quad x \in \mathbb{R}, y > 0\,,$$

is convex. Indeed it is twice differentiable on its domain and:

$$\nabla^2 f(x,y) = \frac{2}{y^3}\begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix} = \frac{2}{y^3}\begin{pmatrix} y \\ -x \end{pmatrix}\begin{pmatrix} y \\ -x \end{pmatrix}^\top \succeq 0\,.$$



More examples

- The *sum-log-exp* function is convex:

$$f(x) = \log \sum_{i=1}^{n} e^{x_i} \; .$$

- The *geometric mean* is concave:

$$f(x) = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} \; .$$

Left as exercice (hint: compute Hessians and show that $v^\top \nabla f(x) v \geq 0$ for all $v \in \mathbb{R}^n$).

Minima of convex function

Théorème 3 *Let C be a convex set and $f : C \to \mathbb{R}$ be a convex function.*

- *Any* local *minimum of f is also a* global *minimum.*
- *If f is* strictly *convex, then there exists* at most one global minimum *of f.*

Proof

If x_1 is a local minimum of f but not a global minimum, there exists x_2 s.t. $f(x_2) < f(x_1)$. By convexity it holds for any $\theta \in [0,1]$:

$$f\left(\theta x_1 + (1-\theta)x_2\right) \leq \theta f\left(x_1\right) + (1-\theta) f\left(x_2\right) < f\left(x_1\right) \ ,$$

which contradicts the fact that x_1 is a local minimum.

If f is strictly convex and x_1 and x_2 are two global minima, then their average $u = \left(x_1 + x_2\right)/2$ satisfies $f(u) \leq \left(f(x_1) + f(x_2)\right)/2$, with strict inequality if $x_1 \neq x_2$: this is not possible, therefore $x_1 = x_2$. $\square$

Optimality conditions

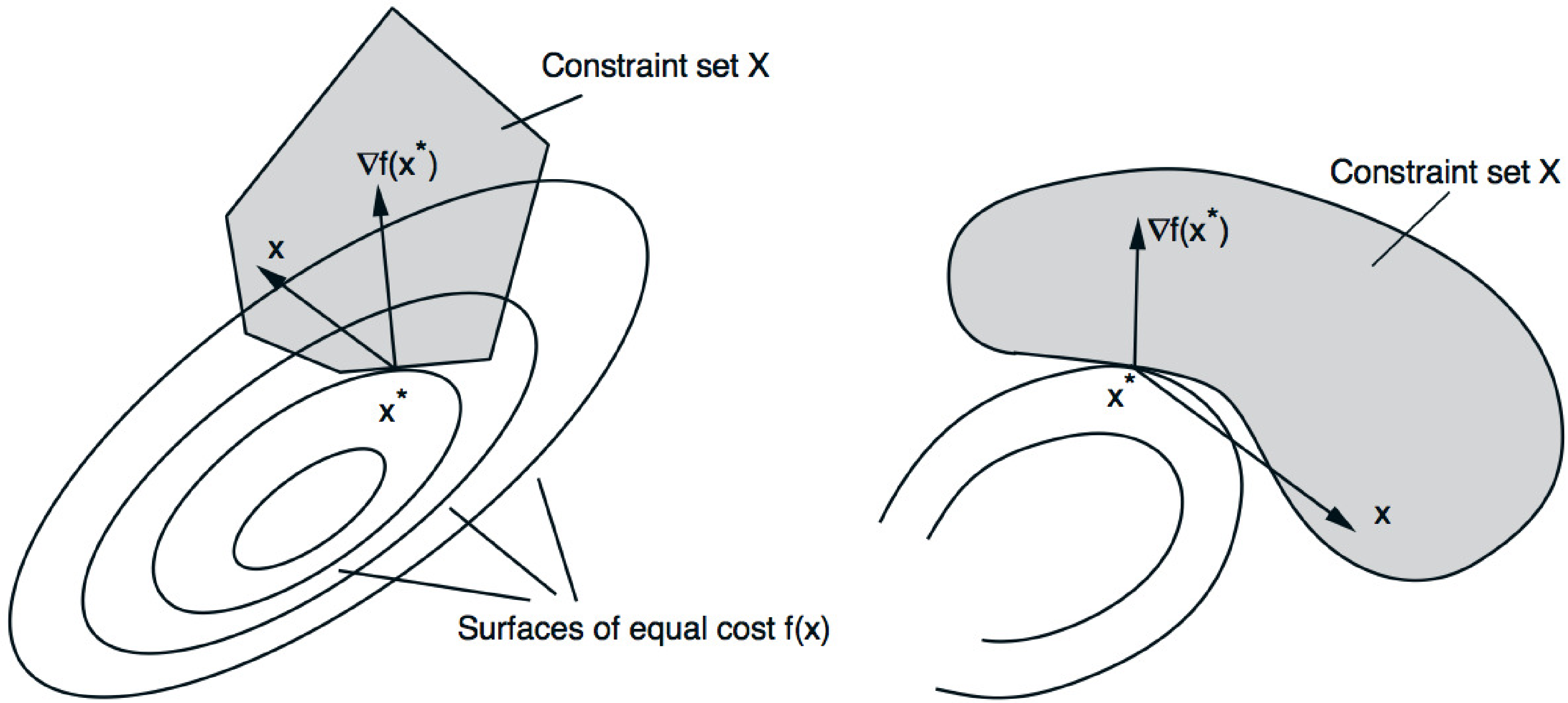
Théorème 4 *Let $\mathcal{X}$ be an convex set, and $f : \mathcal{X} \to \mathbb{R}$ continuously differentiable (not necessarily convex).*

- *If x^* is a local minimum of f over $\mathcal{X}$, then*

$$\nabla f\left(x^*\right)^\top \left(x - x^*\right) \geq 0\,, \quad \forall x \in \mathcal{X}\,.$$

- *If f is convex, then this condition is also* sufficient *for x^* to be a local and therefore global minimum of f over $\mathcal{X}$.*

Illustration



Left: at a local minimum, the gradient $\nabla f\left(x^*\right)$ makes an angle less than or equal to 90 degrees with all feasible variations $x - x^*$. Right: the optimality condition fails if $\mathcal{X}$ is not convex: x^* is a local minimum, but $\nabla f\left(x^*\right)^\top\left(x - x^*\right) < 0$.

Proof

- Let x^* be a local minimum, and suppose there exists $x \in \mathcal{X}$ with $\nabla f(x^*)^\top (x - x^*) < 0$. Then by Taylor expansion we get:

$$f(x^* + \epsilon(x - x^*)) = f(x^*) + \epsilon \nabla f(x^*)^\top (x - x^*) + o(\epsilon),$$

  and therefore for ϵ small enough we have $f(x^* + \epsilon(x - x^*)) < f(x^*)$ which is a contradiction since $x^* + \epsilon(x - x^*)$ is a feasible point by convexity of $\mathcal{X}$.

- If f is convex we have the general property:

$$f(x) \geq f(x^*) + \nabla f(x^*)^\top (x - x^*)$$

  for every $x \in \mathcal{X}$, and therefore $f(x) \geq f(x^*)$ under the hypothesis of the theorem. $\square$

Example

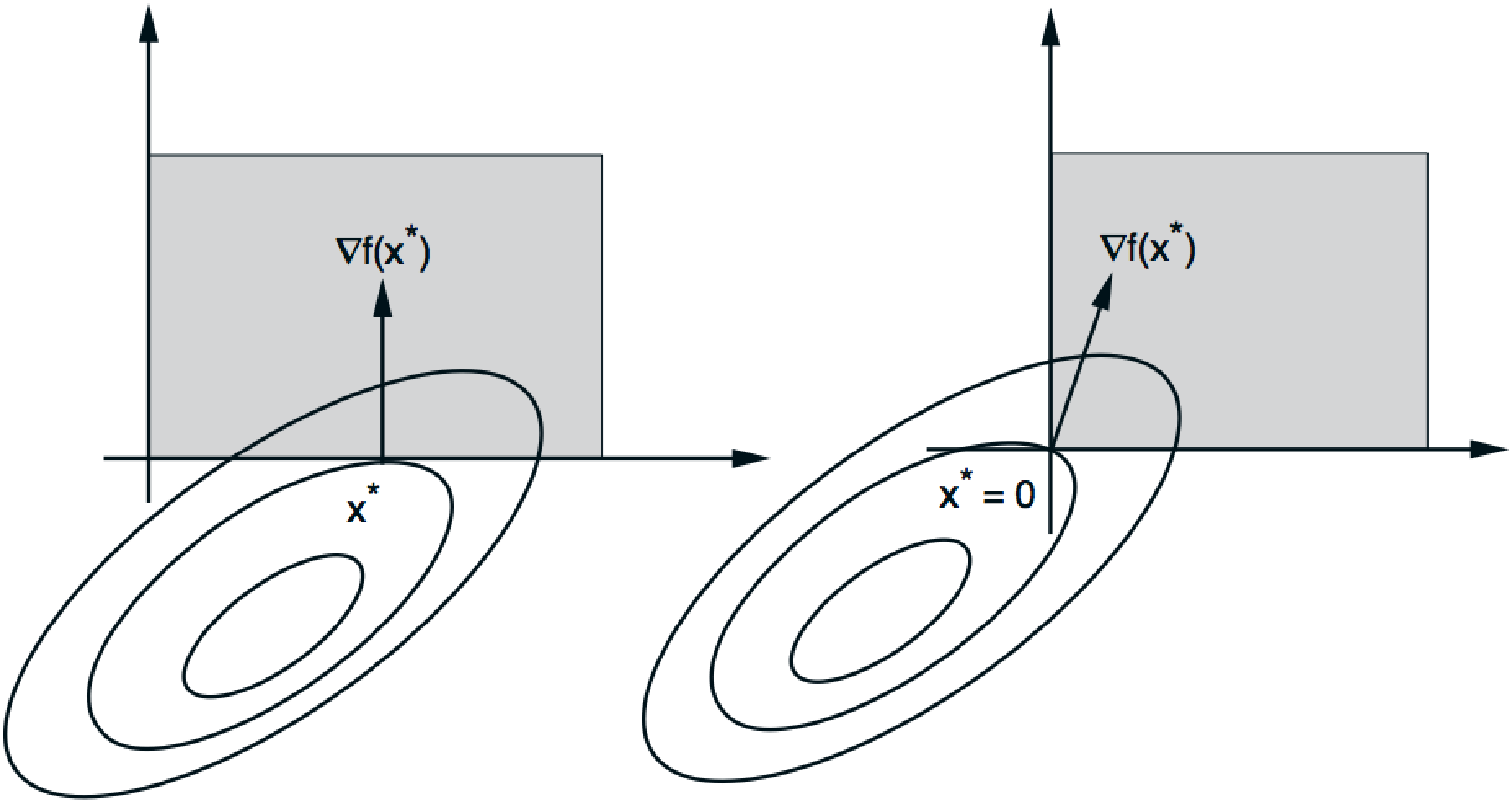
Let $\mathcal{X} = \{x : x \geq 0\}$. The necessary condition for x^* to be a local minimum of f is:

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^*)(x_i - x_i^*) \geq 0 \quad \forall x_i \geq 0 \, .$$

This implies:

$$\frac{\partial f}{\partial x_i}(x^*) \begin{cases} \geq 0 & \forall i \, , \\ = 0 & \textit{if } x_i > 0 \, . \end{cases}$$

Illustration



Optimization with equality constraints

Equality constraints

Here we consider optimization problems where the constraints are specified in terms of equality constraints:

$$\begin{aligned} \textit{minimize} \quad & f(x) \\ \textit{subject to} \quad & h_i(x) = 0 \,, \quad i = 1, \ldots, m \,, \end{aligned}$$

where f and $h_i : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable.

For notational convenience we introduce $h : \mathbb{R}^n \to \mathbb{R}^m$ where $h = (h_1, \ldots, h_m)$ and write the constraint compactly:

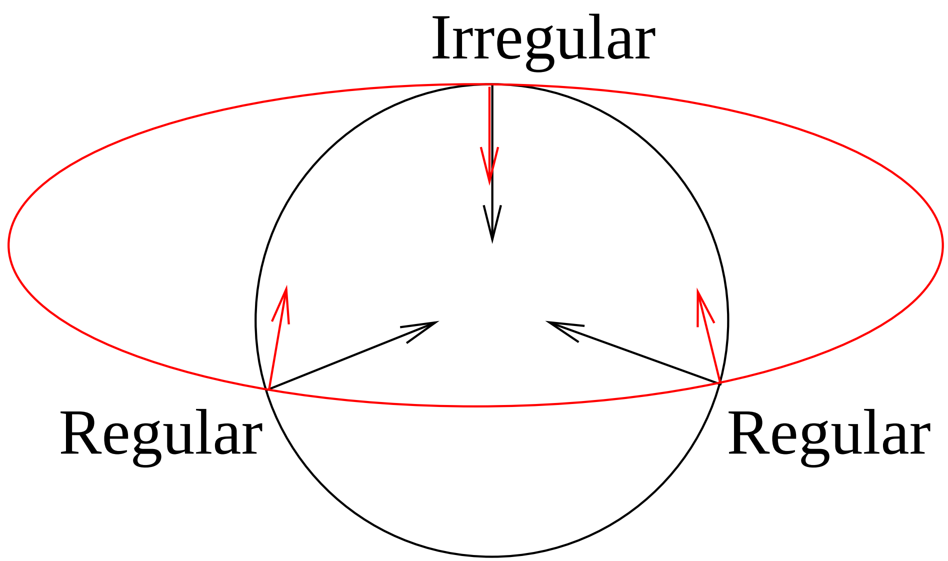
$$h(x) = 0 \,.$$

Regular points

A feasible vector x is called *regular* if the constraint gradients:

$$\nabla h_1(x), \ldots, \nabla h_m(x)$$

is *linearly independent*.



Lagrange Multiplier Theorem

Théorème 5 *Let x^* be a* local minimum *of f subject to $h(x) = 0$, and a* regular *point. Then there exist unique scalars $\lambda_1^*, \ldots, \lambda_m^* \in \mathbb{R}$ called* Lagrange multipliers *such that:*

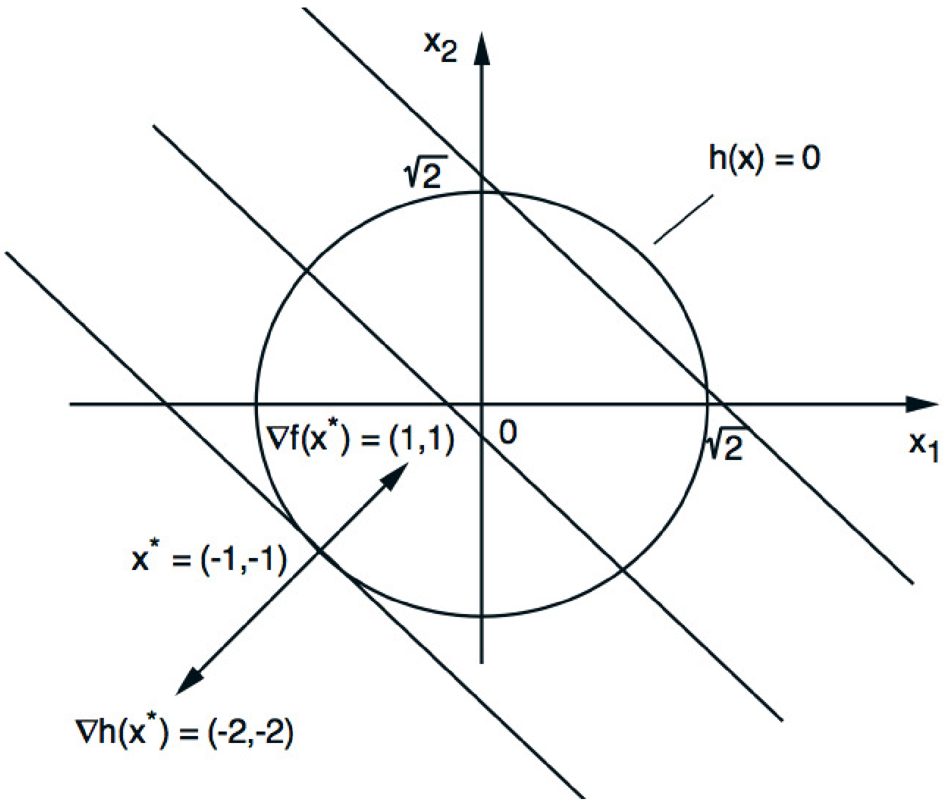
$$\nabla f\left(x^*\right) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i\left(x^*\right) = 0\ .$$

If in addition f and h are twice continuously differentiable *we have:*

$$y^\top \left(\nabla^2 f\left(x^*\right) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_{i\left(x^*\right)} \right) y \geq 0\ , \quad \forall y \text{ s.t. } y^\top \nabla h\left(x^*\right) = 0\ .$$

Illustration: regular case

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 = 2 \ . \end{aligned}$$



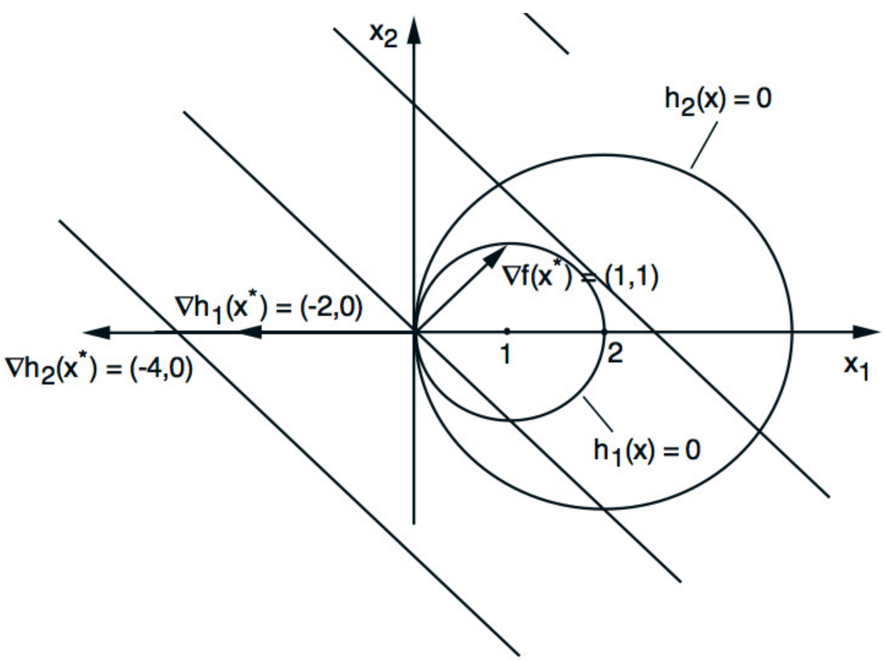
$$\nabla f(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\nabla h(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

$$x^* = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \lambda^* = 1/2 \ .$$

Illustration: irregular case

$$\begin{array}{ll} \text{minimize} & x_1 + x_2 \\ \text{subject to} & (x_1 - 1)^2 + x_2^2 = 1 \ , \\ & (x_1 - 2)^2 + x_2^2 = 4 \ . \end{array}$$



$$x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \nabla f(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\nabla h_1(x) = \begin{pmatrix} -2 \\ 0 \end{pmatrix} \quad \nabla h_2(x) = \begin{pmatrix} -4 \\ 0 \end{pmatrix}$$

Proof

- Introduce, for $k = 1, 2, \ldots$, the cost function:

$$F^k(x) = f(x) + \frac{k}{2} \| h(x) \|^2 + \frac{\alpha}{2} \| x - x^* \|^2 ,$$

where $\alpha > 0$ and x^* is a local minimum, and let

$$x^k = \arg\min_{x \in S} F_k(x) ,$$

where S is a small ball around x^* s.t. $f(x^*) < f(x)$ for all feasible points of S.

- Observe that:

$$F^k\left(x^k\right) = f(x^k) + \frac{k}{2} \| h(x^k) \|^2 + \frac{\alpha}{2} \| x^k - x^* \|^2 \leq F^k\left(x^*\right) = f\left(x^*\right) .$$

Proof (cont.)

- Taking the limit when $k \rightarrow \infty$, this shows that any limit point $\bar{x}$ of $\left(x^k\right)_{k=1,\ldots}$ satisfies $h\left(\bar{x}\right) = 0$, $f\left(\bar{x}\right) = f\left(x^*\right)$ and $\bar{x} = x^*$. Therefore x^* is the only limit point:

$$\lim_{k \rightarrow +\infty} x^k = x^*$$

.

- As a result, for k large enough, x^k is an interior point of S and is an *unconstrained* local minimum of $F^k(x)$.

- From the first-order optimality condition we therefore have, for sufficiently large k:

$$0 = \nabla F^k\left(x^k\right) = \nabla f\left(x^k\right) + k\nabla h\left(x^k\right) h\left(x^k\right) + \alpha\left(x^k - x^*\right) \quad . \tag{1}$$

Since $\nabla h\left(x^*\right)$ has rank m, the same is true for $\nabla h\left(x^k\right)$ if k is sufficiently large, and therefore $\nabla h\left(x^k\right)^\top \nabla h\left(x^k\right)$ *is invertible*.

Proof (cont.)

- We therefore obtain:

$$kh\left(x^k\right) = -\left(\nabla h\left(x^k\right)^\top \nabla h\left(x^k\right)\right)^{-1} \nabla h\left(x^k\right)^\top \left(\nabla f\left(x^k\right) + \alpha\left(x^k - x^*\right)\right) \;.$$

- By taking the limit when $k \to +\infty$:

$$\lim_{k\to+\infty} kh\left(x^k\right) = -\left(\nabla h\left(x^*\right)^\top \nabla h\left(x^*\right)\right)^{-1} \nabla h\left(x^*\right)^\top \nabla f\left(x^*\right) \triangleq \lambda^* \;.$$

- Take now the limit in (1) to obtain:

$$\nabla f\left(x^*\right) + \nabla h\left(x^*\right)\lambda^* = 0 \;.$$

- The second-order condition is also obtained by taking a limit from the second-order optimality condition of x^k [Bersteskas p.288]. $\square$

Lagrangian function

Define the *Lagrangian function* $L : \mathbb{R}^{m+n} \to \mathbb{R}$ by

$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) \; .$$

Then, if x^* is a local minimum which is regular, the Lagrange multiplier conditions are written as a system of $n+m$ equations with $n+m$ unknowns:

$$\nabla_x L\left(x^*,\lambda^*\right) = 0 \; , \quad \nabla_\lambda L\left(x^*,\lambda^*\right) = 0 \; ,$$

$$y^\top \nabla^2_{xx} L\left(x^*,\lambda^*\right) y \geq 0 \; , \quad \forall y \textbf{ s.t. } \nabla\left(x^*\right)^\top y = 0 \; .$$

Example

$$\text{minimize} \quad \frac{1}{2}\left(x_1^2 + x_2^2 + x_3^2\right)$$
$$\text{subject to} \quad x_1 + x_2 + x_3 = 3 \; .$$

Minimize a convex function over a convex set $\Longrightarrow$ a unique global minimum.
First-order necessary conditions:

$$x_1^* + \lambda^* = 0 \; , \quad x_2^* + \lambda^* = 0 \; ,$$
$$x_3^* + \lambda^* = 0 \; , \quad x_1 + x_2 + x_3 = 3 \; .$$

Solution:

$$\lambda^* = -1 \; , \quad x_1^* = x_2^* = x_3^* = 1 \; .$$

Example: Portfolio Selection

Investment of 1 unit of wealth among n *assets* with *random rates of return* e_i $(i = 1, \ldots, n)$ with mean and covariences:

$$\bar{e}_i = E\left[e_i\right] \ ,$$
$$Q_{ij} = E\left[\left(e_i - \bar{e}_i\right)\left(e_j - \bar{e}_j\right)\right] \ .$$

The *return* $r = \sum x_i e_i$ has mean $\sum x_i \bar{e}_i$ and variance $x^\top Q x$. A possible investment strategy is:

$$\text{minimize} \quad x^\top Q x$$
$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1 \, , \quad \sum_{i=1}^{n} \bar{e}_i x_i = m \ .$$

How does the solution vary with m?

Example: Portfolio Selection (cont.)

Let λ_1 and λ_2 be the Lagrange multipliers. The optimality condition is:

$$2Qx^* + \lambda_1 u + \lambda_2 \bar{e} \ ,$$

where $u = (1, \ldots, 1)^\top$ and $\bar{e} = (\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n)^\top$ (assuming u and $\bar{e}$ are linearly independent). This yields:

$$x^* = -\frac{1}{2} Q^{-1} u \lambda_1 - \frac{1}{2} Q^{-1} \bar{e} \lambda_2 \ .$$

But $u^\top x^* = 1$ and $\bar{e}^\top x^* = m$, therefore:

$$\begin{aligned} 1 = u^\top x^* &= -\frac{1}{2} u^\top Q^{-1} u \lambda_1 - \frac{1}{2} u^\top Q^{-1} \bar{e} \lambda_2 \ , \\ m = \bar{e}^\top x^* &= -\frac{1}{2} \bar{e}^\top Q^{-1} u \lambda_1 - \frac{1}{2} \bar{e}^\top Q^{-1} \bar{e} \lambda_2 \ . \end{aligned}$$

Example: Portfolio Selection (cont.)

Solving in λ_1 and λ_2 yields:

$$\lambda_1 = \xi_1 + \xi_2 m \; ,$$
$$\lambda_2 = \xi_3 + \xi_4 m \; ,$$

for some scalar ξ_i. Back to x^* we obtain:

$$x^* = mv + w$$

for some vectors v and w that depend on Q and $\bar{e}$. The corresponding variance of return is:

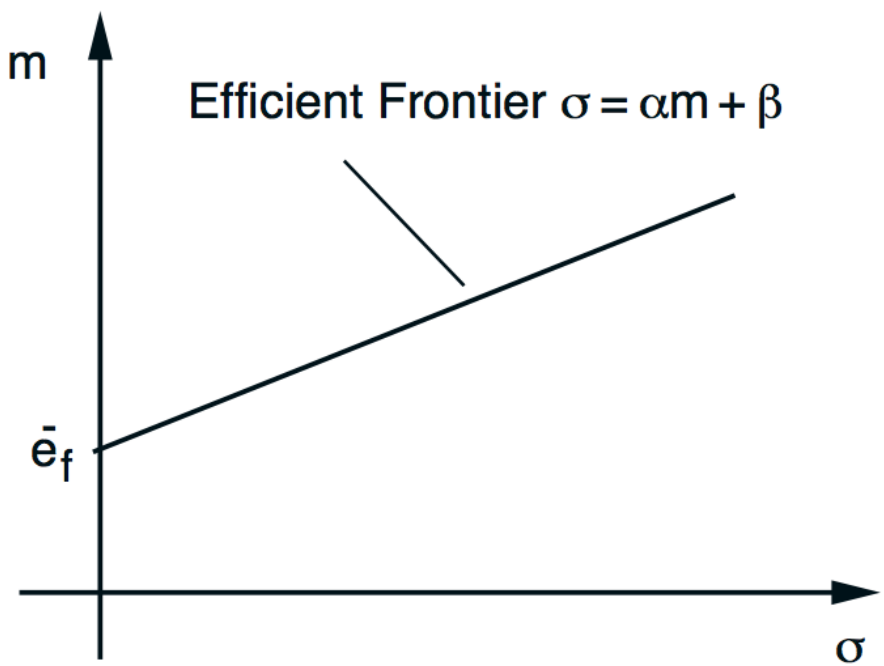
$$\sigma^2 = (mv + w)^\top Q \, (mv + w) = (\alpha m + \beta)^2 + \gamma \; ,$$

where α, β and γ are some scalars that depend on Q and $\bar{e}$.

Example: Portfolio Selection (cont.)

If one asset is *riskless*, then $\sigma^2 = 0$ must be a possible solution (setting m equal to the return of the riskless assset). This implies $\gamma = 0$ and therefore:

$$\sigma = |\,\alpha m + \beta\,|$$



This defines the *efficient frontier*. Each point of the efficient frontier can be achieved by a mixture of two portfolios.

Optimization with inequality constraints

Inequality constraints

Here we consider optimization problems where the constraints are specified in terms of equality and inequality constraints:

$$
\begin{aligned}
\textit{minimize} \quad & f(x) \\
\textit{subject to} \quad & h_i(x) = 0 \,, \quad i = 1, \ldots, m \,, \\
& g_j(x) \leq 0 \,, \quad j = 1, \ldots, r \,,
\end{aligned}
$$

where f and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^r$ are continuously differentiable. For convenience we rewrite the problem as :

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & h(x) = 0 \,, \quad g(x) \leq 0 \,.
\end{aligned}
$$

Active constraints

For any feasible point x, the set of *active inequality constraints* is denoted by:

$$A(x) = \{j \mid g_j(x) = 0\} \ .$$

If $j \notin A(x)$, we say that the j-th constraint is *inactive*. If x^* is a local minimum to the inequality constrained problem (ICP), it is also a local minimum to the same ICP without the inactive constraints at x^*. If a contraint is active, it can be treated "as an equality constraint".

A feasible vector x is said to be *regular* if the equality constraint gradients $\nabla h_i(x), i = 1, \dots, m$ and the active inequality constraint gradients $\nabla g_j(x)$, $j \in A(x)$ are linearly independent.

KKT optimality conditions

Théorème 6 **[Karush(1939),Kuhn and Tucker (1951)]** *Let x^* be a* local minimum *of f subject to $h(x) = 0$, $g(x) \leq 0$ and a* regular *point. Then there exist unique Lagrange multipliers $\lambda = (\lambda_1^*, \ldots, \lambda_m^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ such that the following* KKT conditions *are satisfied:*

$$\nabla_x L\left(x^*, \lambda^*, \mu^*\right) = 0 \,,$$

$$\mu_j^* \geq 0 \,, \quad j = 1, \ldots, r,$$

$$\mu_j^* = 0 \,, \quad \forall j \notin A\left(x^*\right)$$

where the Lagrangian function is:

$$L\left(x, \lambda, \mu\right) = f\left(x\right) + \sum_{i=1}^{m} \lambda_i h_i\left(x\right) + \sum_{j=1}^{r} \mu_j g_j(x) \,.$$

Proof (sketch)

The proof is similar to the proof of the Lagrange theorem of equality constrained problems, with the penalized function:

$$F^k(x) = f(x) + \frac{k}{2} \| h(x) \|^2 + \frac{k}{2} \sum_{j=1}^{r} \left(g_j^+(x) \right)^2 + \frac{\alpha}{2} \| x - x^* \|^2 ,$$

where:

$$g_j^+(x) = \max\left(0, g_j(x)\right) , \quad j = 1, \ldots, r . \quad \square$$

Example

$$\text{minimize} \quad \frac{1}{2}\left(x_1^2 + x_2^2 + x_3^2\right)$$
$$\text{subject to} \quad x_1 + x_2 + x_3 \leq -3 \ .$$

Minimization of a convex function over a convex set has a single local (global) optimum x^*. Every point is regular so x^* must satisfy the KKT conditions:

$$x_1^* + \mu^* = 0 \ , \quad x_2^* + \mu^* = 0 \ , \quad x_3^* + \mu^* = 0 \ .$$

Example (cont.)

There are two possibilities

- The constraint is *inactive*:

$$x_1^* + x_2^* + x_3^* < -3 \, ,$$

  in which case $\mu^* = 0$. Then we obtain $x_1^* = x_2^* = x_3^* = 0$ which leads to a contradiction.

- The constraint is *inactive*:

$$x_1^* + x_2^* + x_3^* = -3 \, .$$

  Then we obtain $x_1^* = x_2^* = x_3^* = -1$ and $\mu^* = 1$, which satisfies all KKT conditions. This is the unique candidate for a local minimum, it is therefore the unique global solution.

Summary

- The KKT conditions *generalize* the unconstrained and equality-constrained cases.
- These conditions are only *necessary*: they provide conditions a *regular local optimum* must fulfill.
- *Irregular* local optima are not covered by these conditions.
- The conditions can be used to *find candidate* regular local optima.
- Sometimes the conditions are sufficient: see next lessons about *duality*.
- Lagrange multipliers are useful for *sensitivity analysis* : see next lessons.