Nonlinear Programming

-

Optimality conditions for unconstrained and linearly-constrained problems

Marcello Sanguineti

marcello.sanguineti@unige.it

DIBRIS – Università di Genova

Unconstrained and constrained NLP

f: real function of n real variables

$$\begin{cases} \min f(x) \\ x \in \Omega \subseteq \mathbb{R}^n \end{cases}$$

f nonlinear

and/or

Ω defined by constraints, at least one of which is nonlinear

- If $\Omega = \mathbb{R}^n$: **unconstrained** (or **free**) **NLP**
- If $\Omega \subset \mathbb{R}^n$: **constrained NLP**
  - Only equality constraints
  - Equality constraints **and** inequality constraints

Minimum and maximum points

Given a set $\Omega \subseteq \mathbb{R}^n$ and a function f: $\Omega \rightarrow \mathbb{R}$, a point $x^* \in \Omega$ is a **global minimum** if

$$f(x^\star) \leq f(x) \qquad \forall\, x \in \Omega.$$

A point $\bar{x} \in \Omega$ is a **local minimum** if $\exists\, \epsilon > 0$:

$$f(\bar{x}) \leq f(x) \qquad \forall\, x \in \Omega \cap D(\bar{x}, \epsilon)$$

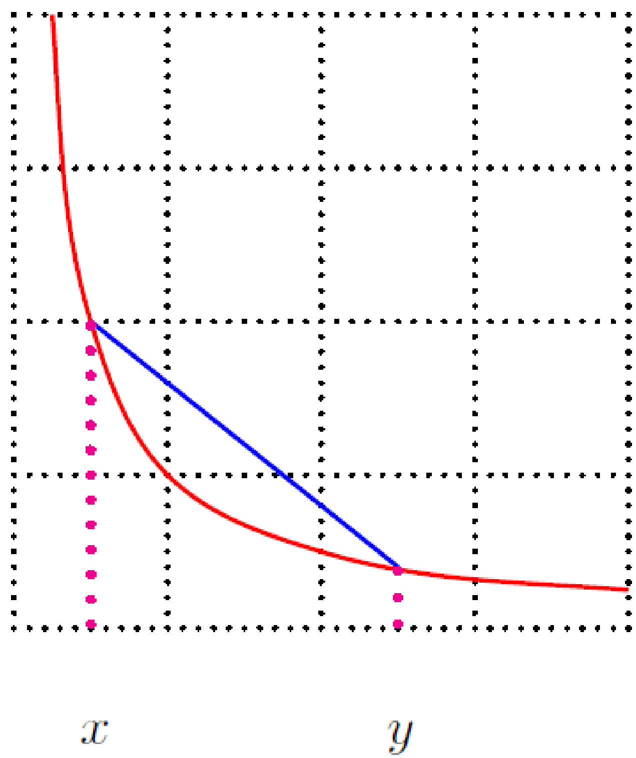
where

$$D(\bar{x}, \epsilon) = \{y \in \mathbb{R}^n : \|\bar{x} - y\| \leq \epsilon\}$$

The role of convexity

Let $f{:}\Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ with Ω nonempty convex set; f is **convex** if for all $x,y \in \Omega$, $\lambda \in [0, 1]$ one has

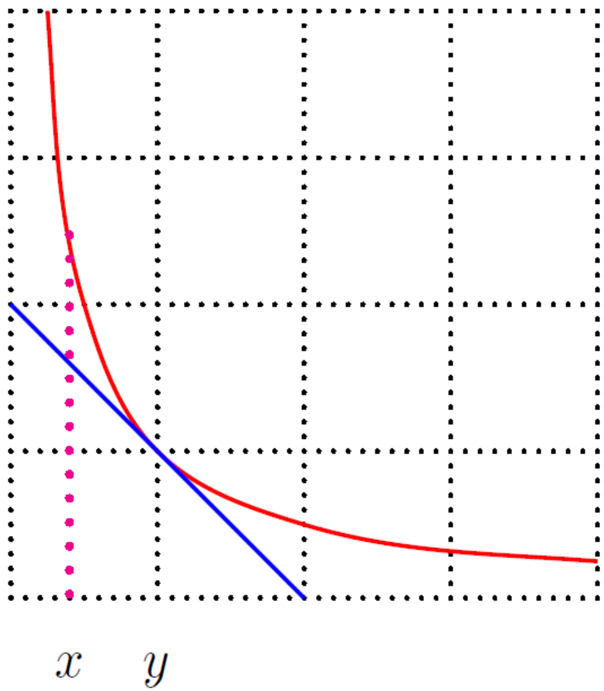
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$



Property. Every convex function is continuous in the interior of Ω. It is not necessarily differentiable.

If a function f is differentiable $\Rightarrow f$ is convex iff $\forall x \in \Omega$ one has

$$f(x) \geq f(y) + (x-y)^T \nabla f(y)$$



Property. If f is twice differentiable with continuity ⇒ f is convex iff its Hessian matrix is positive semi-definite, i.e., letting

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

one has

$$v^T \nabla^2 f(x) v \geq 0 \qquad \forall\, v \in \mathbb{R}^n$$

or, equivalently, all the eigenvalues of $\nabla^2 f(x)$ are ≥ 0.

Example. For a **quadratic** function (A: symmetric):

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

we get

$$\nabla f(x) = Ax + b \quad \nabla^2 f(x) = A$$

⇒ f is convex iff A is positive semi-definite.

Fundamental property of convex functions.

If Ω is convex and nonempty and f is convex, then every local minimum is also a global minimum.

Proof. By contradiction, let $\bar{x}$ be a local minimum and x^* a global minimum with $f(x^*) < f(\bar{x})$. By definition, $\exists\epsilon$: $f(\bar{x}) \leq f(y)$ if $y \in D(\bar{x}, \epsilon) \cap \Omega$. Convexity implies that for all $\lambda \in [0, 1]$ one has

$$\begin{aligned} f(\lambda x^\star + (1-\lambda)\bar{x}) &\leq \lambda f(x^\star) + (1-\lambda) f(\bar{x}) \\ &< \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) \\ &= f(\bar{x}) \end{aligned}$$

Picking λ small enough, we have

$$\lambda x^\star + (1-\lambda)\bar{x} \in D(\bar{x}, \epsilon) \cap \Omega$$

from which the contradiction follows. ▪

Admissible directions

Let $\Omega \neq 0$ and let $x \in \Omega$. A vector $d \in \mathbb{R}^n$ is called **admissible direction** in x if $\exists\, \bar{\alpha} > 0$ such that for all $\alpha \in [0, \bar{\alpha})$ one has

$$x + \alpha d \in \Omega$$

If x is an internal point in Ω, then every direction is admissible in x

If $\Omega \equiv \mathbb{R}^n$, then every direction is admissible $\forall x$.

A vector d is called **admissible descent direction** in $x \in \Omega$ for the function f if $\exists\, \bar{\lambda} > 0$ such that for all $\lambda \in (0, \bar{\lambda})$ one has

$$f(x + \lambda d) < f(x), x + \lambda d \in \Omega$$

First-order necessary optimality conditions

Theorem. Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^* \in \mathbb{R}^n$, S be an open set of $\mathbb{R}^n$ containing x^* , and $f \in C^1(S)$. If x^* is a local minimum, then for every admissible direction d one has

$$d^T \nabla f(x^\star) \geq 0$$

Proof. Let d be an admissible direction; the Taylor series stopped at the first order is:

$$\begin{aligned} f(x^\star + \alpha d) &= f(x^\star) + (x^\star + \alpha d - x^\star)^T \nabla f(x^\star) + o \\ &= f(x^\star) + \alpha d^T \nabla f(x^\star) + o(\alpha) \end{aligned}$$

If α is "small enough" $\Rightarrow$

$$\begin{aligned} f(x^\star + \alpha d) - f(x^\star) &\geq 0 && \Rightarrow \\ \alpha d^T \nabla f(x^\star) + o(\alpha) &\geq 0 && \Rightarrow \\ d^T \nabla f(x^\star) + \frac{o(\alpha)}{\alpha} &\geq 0 && \Rightarrow \quad \text{for } \alpha \downarrow o \\ d^T \nabla f(x^\star) &\geq 0 \end{aligned}$$

▪

The theorem state that **in a local minimum no descent direction exist.**

Particular case: unconstrained problem

In this case, $\Omega = \mathbb{R}^n$. So

$$\begin{array}{rcll} d^T \nabla f(x^\star) & \geq & 0 & \forall\, d \in \mathbb{R}^n \Rightarrow \\ e_i^T \nabla f(x^\star) = [\nabla f(x^\star)]_i & \geq & 0 & \text{and} \\ -e_i^T \nabla f(x^\star) = -[\nabla f(x^\star)]_i & \geq & 0 & \Rightarrow \\ [\nabla f(x^\star)]_i & = & 0 & i = 1, \ldots, n \end{array}$$

Hence:

in unconstrained problems the local minimum points are **stationary points**, i.e.,

$$\nabla f(x^\star) = 0$$

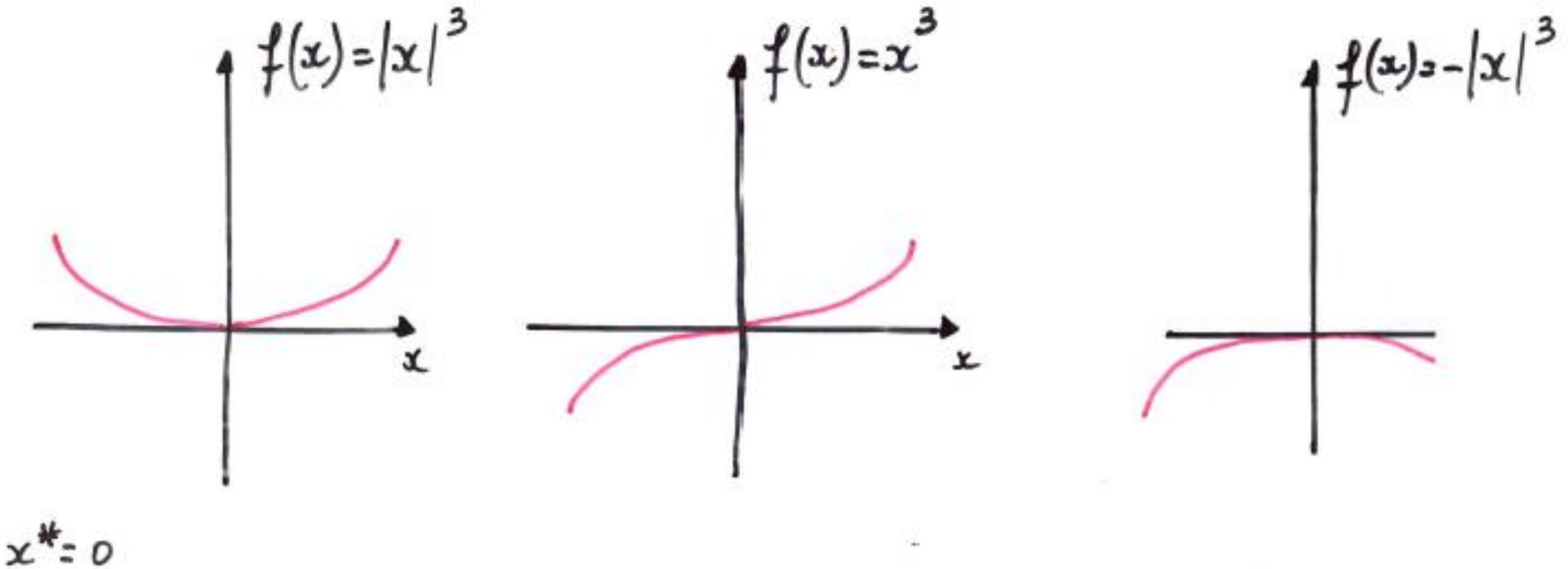
So, we proved the following result.

Theorem (I order necessary optimality condition for unconstrained NLP). Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^* \in \mathbb{R}^n$, S be an open set of $\mathbb{R}^n$ containing x^* , and $f \in C^1(S)$.

Then in order for x^* to be local minimum of f it has to be $\nabla f(x^*) = 0$.

Warning: the condition $\nabla f(x^*) = 0$ is **merely necessary**.

It must be satisfied by every local minimum point, but it can be satisfied also by maximum or inflection points.



Particular case: linear equality constraints

In this case, we have $\Omega = \{x \in \mathbb{R}^n\colon Ax = b\}$ with A matrix (m,n)

We compute the admissible directions

Let x be admissible and $d \in \mathbb{R}^n$, $\alpha>0$. Then

$$
\begin{array}{rrcll}
\alpha, d: & A(x+\alpha d) & = & b & \Leftrightarrow \\
& Ax + \alpha A d & = & b & \Leftrightarrow \\
& \alpha A d & = & 0 & \Leftrightarrow \\
& A d & = & 0 & \equiv \; d \perp \text{rows of } A
\end{array}
$$

If A has maximum rank, equal to the number of its rows, let A_B be a basis matrix (invertible square submatrix (m,m)):

$$
\begin{aligned}
A &= [A_B \quad A_N] \qquad \det(A_B) \neq 0 \\
Ad &= 0 \qquad \Leftrightarrow \\
A_B d_B + A_N d_N &= 0 \\
d_B &= -A_B^{-1} A_N d_N \\
d &= \begin{bmatrix} -A_B^{-1} A_N \\ I \end{bmatrix} d_N
\end{aligned}
$$

non serve tutta la slide

Let

$$Z = \begin{bmatrix} -A_B^{-1} A_N \\ I \end{bmatrix}$$

Z is a matrix $(n, n\text{-}m)$. Then

$$Ad = 0 \quad \Leftrightarrow \qquad d = Z d_N \qquad \text{for some } d_N \in \mathbb{R}^{n-m}.$$

Necessary condition of the first order:

$$\begin{aligned} d^T \nabla f(x^\star) &\geq 0 & \forall \text{ admissible } d \\ (Z d_N)^T \nabla f(x^\star) &\geq 0 & \forall d_N \in \mathbb{R}^{n-m} \\ d_N^T Z^T \nabla f(x^\star) &\geq 0 & \forall d_N \quad \Leftrightarrow \\ Z^T \nabla f(x^\star) &= 0 & \end{aligned}$$

$Z^T \nabla f(x^*) \in \mathbb{R}^{n-m}$: **projected gradient** of f in x^*.

Hence:

when there are only linear equality constraints, in the local minimum points the projected gradient is equal to zero.

Example

$$\min_{x+y=1} x^2 + y^2$$

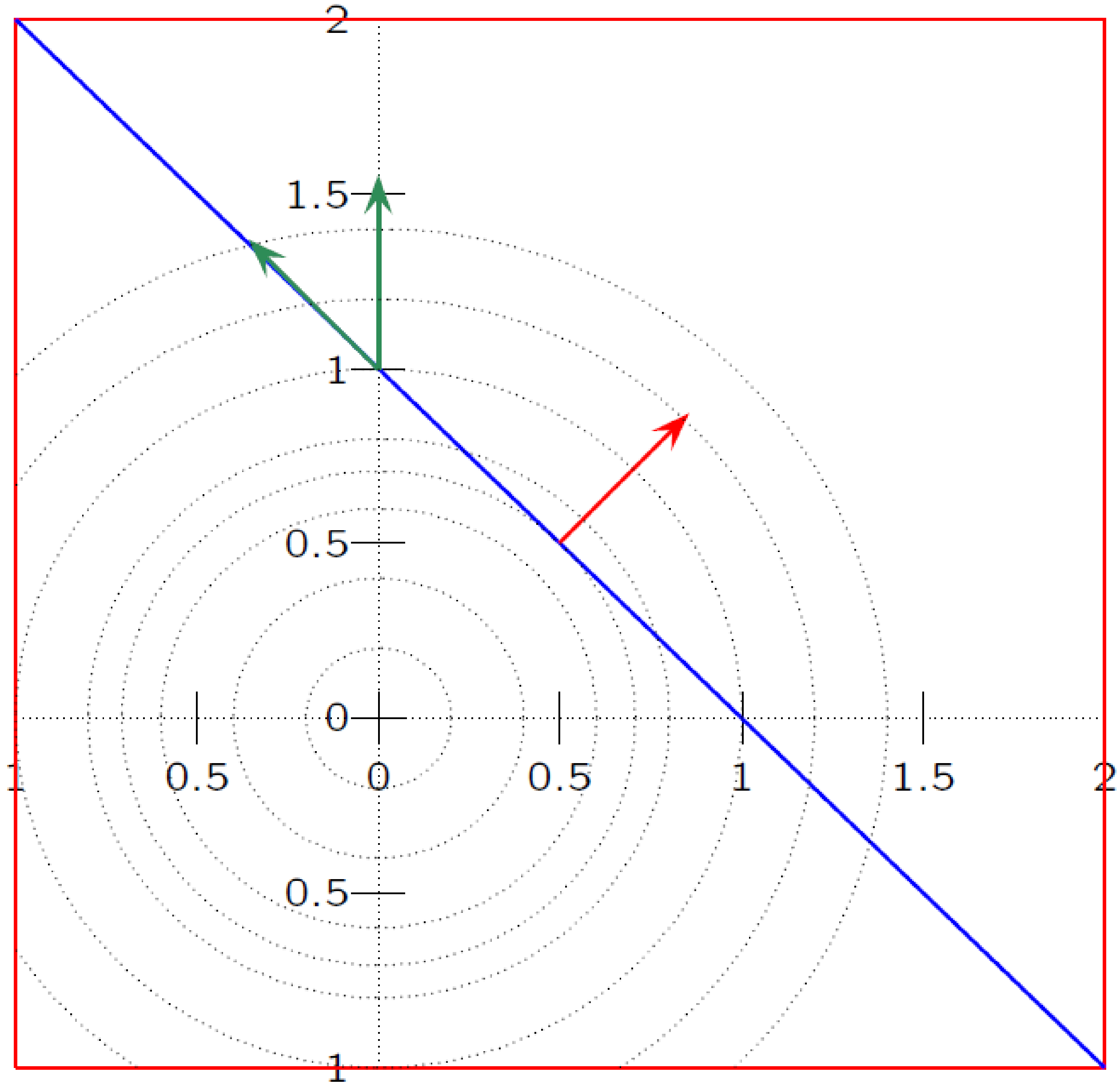
$$\begin{aligned} \Omega &= \{(x,y) : x + y = 1\} \\ &= \left\{(x,y) : \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1\right\} \end{aligned}$$

$$A_B = [1] \qquad A_N = [1]$$

$$\begin{aligned} Z &= \begin{bmatrix} -1^{-1} \cdot 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ d &= \begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} d_y \\ Z^T \nabla f(x,y) &= \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 2x \\ 2y \end{bmatrix} = 2(y - x) \end{aligned}$$

BY enforcing the first-order necessary condition and the admissibility, we get

$$\begin{aligned} 2(y-x) &= 0, x + y = 1 \Rightarrow \\ x &= y = 1/2 \end{aligned}$$


2
1.5
1
0.5
0
0.5
1
1
0.5
0
0.5
1
1.5
2

Lagrange multipliers for linear equality constraints

If A has maximum rank equal to $m \Rightarrow m$ linearly indep. rows (vectors in $\mathbb{R}^n$).

Z: n-m linearly indipendent **columns**. Each of them is a vector of $\mathbb{R}^n$.

Each **row** of A is orthogonal to each **column** of Z- Indeed:

$$AZ = [A_B \quad A_N] \begin{bmatrix} -A_B^{-1} A_N \\ I \end{bmatrix} = A_B \cdot (-A_B^{-1} A_N) + A_N \cdot I = 0$$

Orthogonality $\Rightarrow$ indipendence, hence the set of n vectors (m rows, n-m columns)

$$a_1, \ldots, a_m, Z_1, \ldots, Z_{n-m}$$

is a set of n **linearly indipendent** vectors, hence a **basis** for $\mathbb{R}^n$

$\Rightarrow$ any element of $\mathbb{R}^n$ can be written in a unique way as a linear combination of such vectors:

$$\forall v \in \mathbb{R}^n, \quad \exists! \quad \lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_{n-m} :$$
$$v \quad = \quad \sum_{i=1}^{m} \lambda_i a_i + \sum_{i=1}^{n-m} \mu_i Z_i$$

This means that $\exists \lambda \in \mathbb{R}^n$, $\mu \in \mathbb{R}^{n\text{-}m}$ such that

$$v = A^T \lambda + Z\mu$$

In particular, $\nabla f(x^*) \in \mathbb{R}^n$. So, $\exists \lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^{n-m}$ such that

$$\nabla f(x^\star) = A^T \lambda + Z\mu$$

By the first-order necessary condition, we have

$$\begin{aligned}
Z^T \nabla f(x^\star) &= 0 & \Rightarrow \\
Z^T (A^T \lambda + Z\mu) &= 0 & \Rightarrow \\
(AZ)^T \lambda + Z^T Z\mu &= 0 & \Rightarrow \\
(Z^T Z)\mu &= 0
\end{aligned}$$

Let us verify that $Z^T Z$ is invertible. We have

$$v^T Z^T Z v = (Zv)^T (Zv) = 0 \qquad \text{iff } Zv=0,$$

but $Zv=0$ iff $v=0$, because all the columns of Z are independent. So $Z^T Z$ is **positive definite**, i.e., all its eigenvalues are positive ⇒ strictly positive determinant ⇒ invertible. So

$$Z^T Z\, \mu=0 \text{ iff } \mu=0.$$

Summing up, we conclude that the first-order necessary condition is

$$\exists \lambda \in \mathbb{R}^m : \nabla f(x^\star) = A^T \lambda$$

We introduce the **Lagrangian function**:

$$L(x, \lambda) = f(x) - \lambda^T (Ax - b)$$

In terms of the Lagrangian function, we have the following:

The first-order condition is equivalent to the stationarity of the Lagrangian function:

$$\begin{aligned} \frac{\partial L}{\partial x} &= \nabla f(x) - A^T \lambda := 0 \\ \frac{\partial L}{\partial \lambda} &= (Ax - b) := 0 \end{aligned}$$

λ: **Lagrange multipliers**

Geometric interpretation: in a minimum point the gradient is a linear combination of the rows of A.

Given a linear constraint

$$\alpha^T x = b$$

the vector α is orthogonal to the hyperplane $\alpha^T x = b$ and points towards the half-space $\alpha^T x \geq b$.

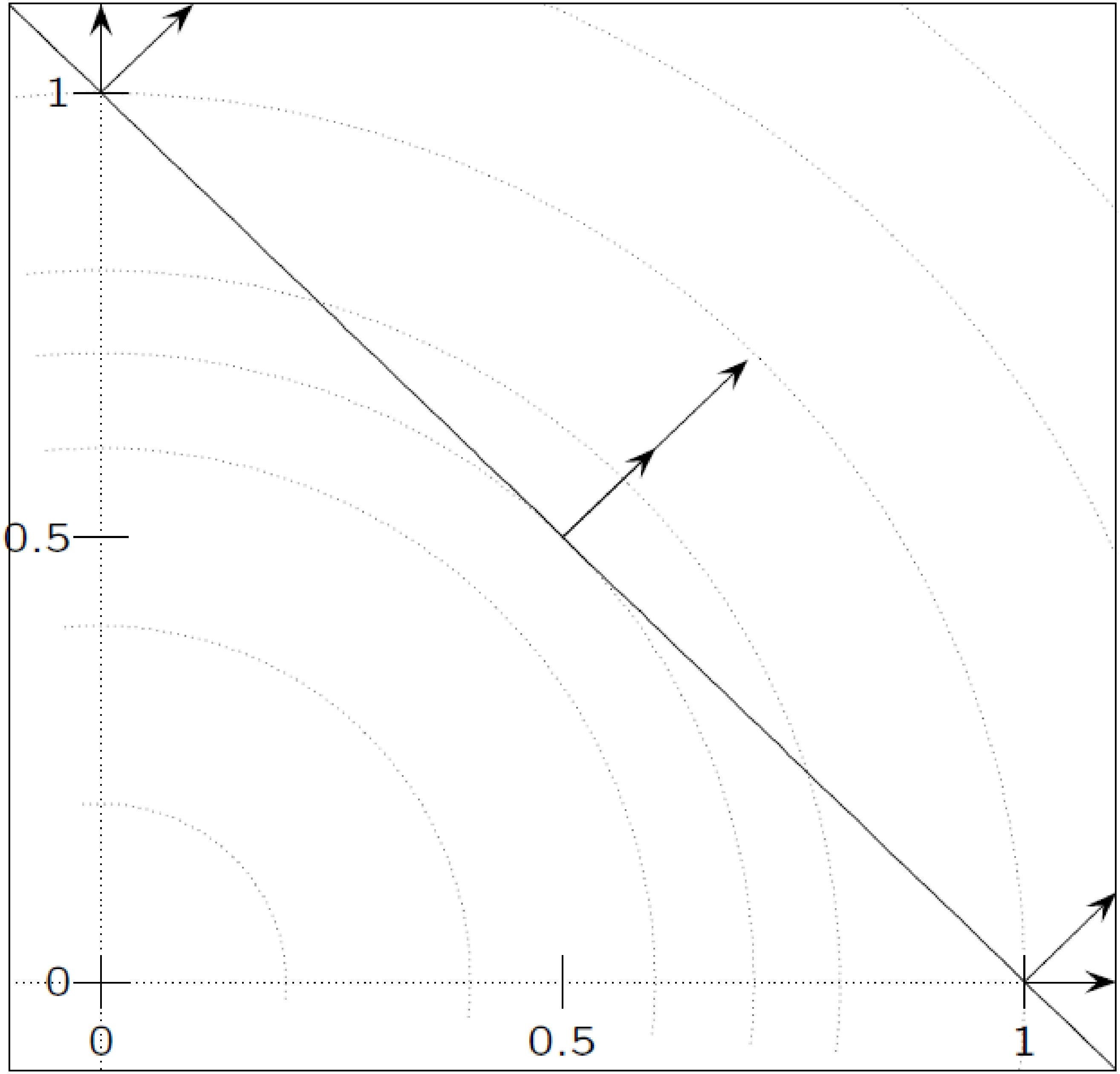
So, the first-order necessary condition is verified in x iff x is an admissible solution and in x the gradient can be obtained by a linear combination of vectors that are orthogonal to the constraints.

Back to the example

$$\min_{x+y=1} x^2 + y^2$$

$$\begin{aligned} \Omega &= \{(x,y) : x+y=1\} \\ &= \left\{(x,y) : [1 \quad 1] \begin{bmatrix} x \\ y \end{bmatrix} = 1\right\} \end{aligned}$$

$$\begin{aligned} \exists ? \lambda \in \mathbb{R} \quad &: \quad \nabla f(x,y) = A^T \lambda \\ \begin{bmatrix} 2x \\ 2y \end{bmatrix} &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \lambda \qquad \Rightarrow \\ \lambda &= 1 \\ x = y &= 1/2 \end{aligned}$$


1
0.5
0
0
0.5
1

Back to LP: dual variables as Lagrange multipliers

Linear Programming with only equality constraints (=standard form)

$$\min_x c^T x$$
$$Ax = b$$

Dual problem:

$$\max_\lambda \lambda^T b$$
$$\lambda^T A = c^T$$

The dual constraint can be written as $A^T \lambda = c = \nabla (c^T x)$

So, **the Lagrange multipliers are the variables of the dual problem**.

The necessary condition for optimality is the existence of an admissible solution for the dual.

Alternative method to obtain the first-order necessary conditions for problems with linear equality constraints.

$$\begin{aligned} d^T \nabla f(x^\star) &\geq 0 \quad \forall\, d : Ad = 0 \\ 0 &= \min_d \nabla^T f(x^\star) d \\ Ad &= 0 \end{aligned}$$

Dual problem:

$$\max_\lambda \lambda^T 0$$

$$\lambda^T A = \nabla^T f(x^\star)$$

The necessary condition of the I order holds in x^* iff

$$\exists\, \lambda : A^T \lambda = \nabla f(x^\star).$$

Particular case: linear inequality constraints

$$\Omega = \{x \in \mathbb{R}^n : Ax \geq b\}$$

Given x^* admissible, let $A_=$ be the submatrix of A corresponding to the active constraints in x^* (subset of the rows of A). Let $b_=$ be the vector of the corresponding right-hand sides. Similarly, let us define $A_>$ and $b_>$:

$$\begin{aligned} A_= x^\star &= b_= \\ A_> x^\star &> b_> \end{aligned}$$

The inactive constraints do not bind the choice of the admissible direction. Therefore, a direction d is admissible iff

$$\begin{aligned} A_= x^\star + \alpha A_= d &\geq b_= \qquad \text{i.e.,} \\ \alpha A_= d &\geq b_= - A_= x^\star = 0 \\ A_= d &\geq 0 \end{aligned}$$

First-order necessary condition:

$$d^T \nabla f(x^\star =\geq 0 \quad \forall\, d : A_= d \geq 0$$

$$0 = \min_{d : A_= d \geq 0} \nabla^T f(x^\star) d$$

Dual problem:

$$\begin{array}{rcl} \max \lambda^T 0 & & \\ \lambda^T A_= & = & \nabla^T f(x^\star) \\ \lambda & \geq & 0 \end{array}$$

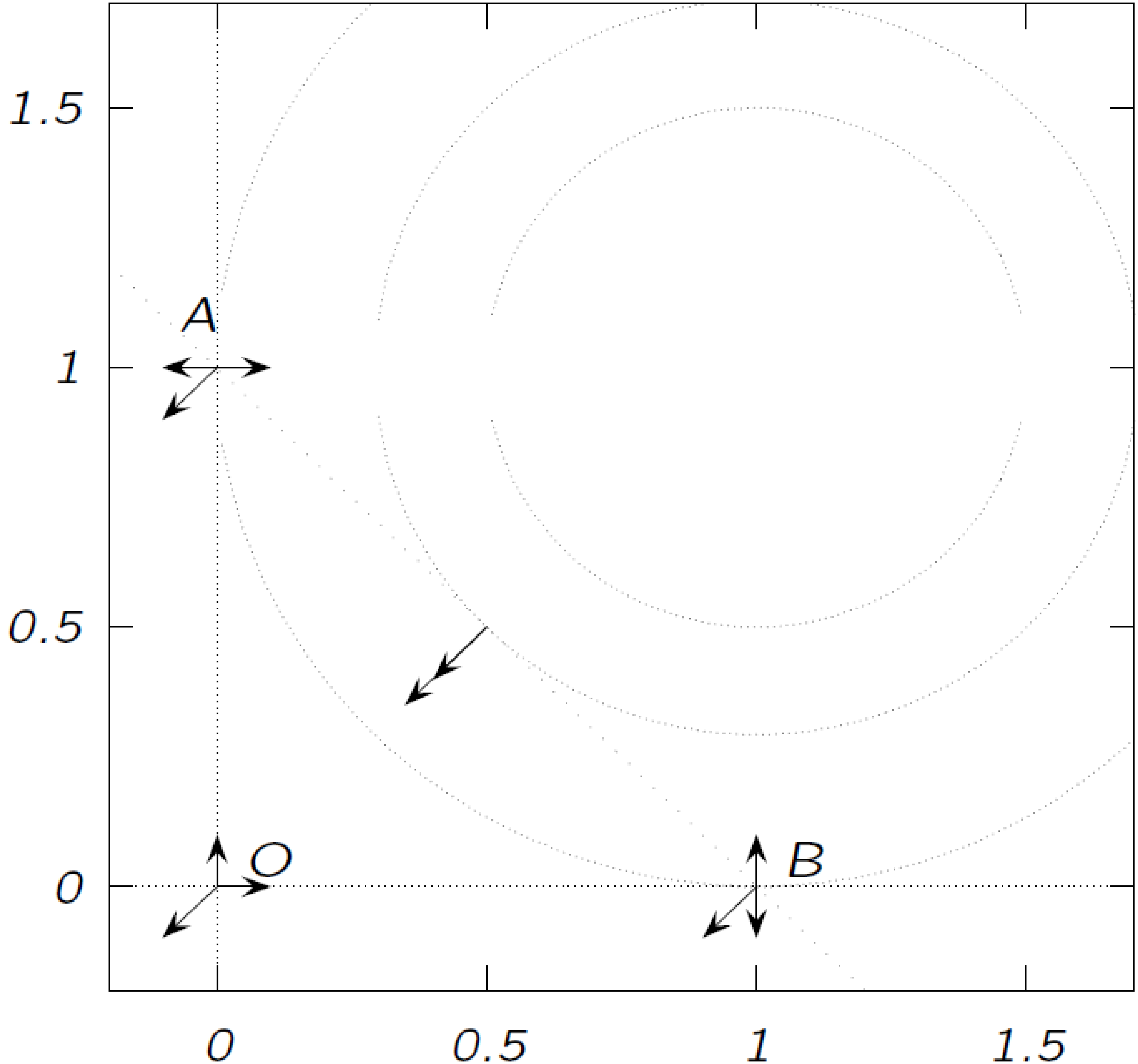
⇒ the first-order necessary condition is equivalent to

$$\exists\, \lambda \geq 0 : A_=^T \lambda = \nabla f(x^\star).$$

Geometric interpretation: the gradient is a linear combination with non-negative coefficients of the vectors normal to the **active** constraints.

Example.

$$
\begin{aligned}
\min (x-1)^2 + (y-1)^2 & & \\
-x - y & \geq & -1 \\
x, y & \geq & 0
\end{aligned}
$$


1.5
1
0.5
0
A
O
B
0
0.5
1
1.5
1.5

First-order necessary condition:

$$\nabla f(x,y) = \begin{bmatrix} 2(x-1) \\ 2(y-1) \end{bmatrix}$$

We have to analyze the following cases:

1. (x,y) internal point
2. $(x,y) \in (A,B)$
3. $(x,y) \equiv A$
4. $(x,y) \in (O,A)$
5. $(x,y) \equiv O$
6. $(x,y) \in (O,B)$
7. $(x,y) \equiv B$

1. Internal point

$$\begin{bmatrix} 2(x-1) \\ 2(y-1) \end{bmatrix} = 0$$

from which $x = y = 1$, not admissible.

2. $(x,y) \in (A,B)$. Active constraint: $-x-y \geq -1 \Rightarrow$ I order necessary condition:

$$\begin{aligned} \exists \lambda \in \mathbb{R}^+ : \begin{bmatrix} 2(x-1) \\ 2(y-1) \end{bmatrix} &= \begin{bmatrix} -1 \\ -1 \end{bmatrix} \lambda \qquad \Rightarrow \\ -x-y &= -1 \qquad \Rightarrow \\ x &= y = 1/2 \\ \lambda &= -2(1/2-1) > 0 \end{aligned}$$

The point $(1/2, 1/2)$ satisfies the necessary condition.

3. $(x,y) = (0,1)$. Active constraints: $-x-y \geq 0$, $x \geq 0$. So, $A_= = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}$

$$\begin{aligned} \begin{bmatrix} -2 \\ 0 \end{bmatrix} &= \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} \lambda \qquad \Rightarrow \\ -\lambda_1 + \lambda_2 &= -2 \\ -\lambda_1 &= 0 \\ \lambda^T &= \begin{bmatrix} 0 & -2 \end{bmatrix} \end{aligned}$$

from which we obtain

Lagrange multipliers $\ngeq 0$.

4. $(x,y) \in (O,A)$. Active constraint: $x \geq 0 \Rightarrow$

$$\begin{aligned} \begin{bmatrix} -2 \\ 2(y-1) \end{bmatrix} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \lambda \\ x &= 0 \\ y &\in (0,1) \end{aligned}$$

which is impossible.

5. $(x,y) = (0,0)$. Active constraints: $x \geq 0$, $y \geq 0$.

$$\begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \lambda$$

$$\lambda = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \ngeq 0$$

6. etc.