



**HAL**  
open science

## Selected Topics on Continuous Optimization

Jean Charles Gilbert

► **To cite this version:**

Jean Charles Gilbert. Selected Topics on Continuous Optimization. Master. Palaiseau, France. 2019, pp.148. cel-01249369v1

**HAL Id: cel-01249369**

**<https://inria.hal.science/cel-01249369v1>**

Submitted on 4 Jan 2016 (v1), last revised 21 Mar 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Advanced Continuous Optimization

J. Ch. Gilbert (INRIA Paris-Rocquencourt)

— November 16, 2015 —

Lectures: September 21, 2015 → November 9, 2015

Examination: November 16, 2015

Keep a close eye to

<http://who.rocq.inria.fr/Jean-Charles.Gilbert/paris-saclay/optim.html>

## Outline

- 1 Background
  - Convex analysis
  - Nonsmooth analysis
  - Optimization
  - Algorithmics
- 2 Optimality conditions
  - First order optimality conditions
  - Second order optimality conditions
- 3 Linearization methods
  - The Josephy-Newton algorithm for inclusions
  - The SQP algorithm in constrained optimization
  - The semismooth Newton method for equations
- 4 Semidefinite optimization
  - Problem definition
  - Existence of solution and optimality conditions
  - An interior point algorithm

# Preliminaries

## Some features of the course I

- Smooth nonconvex optimization

- ▶ Dealing with problems like

$$(P_{EI}) \quad \begin{cases} \inf_x f(x) \\ c_E(x) = 0 \\ c_I(x) \leq 0, \end{cases}$$

where  $f : \mathbb{E} \rightarrow \mathbb{R}$ ,  $E$  and  $I$  form a partition of  $[1 : m]$ ,  $c_E : \mathbb{E} \rightarrow \mathbb{R}^{m_E}$ , and  $c_I : \mathbb{E} \rightarrow \mathbb{R}^{m_I}$  are smooth (possibly non convex) functions

- ▶ Or more generally

$$(P_A) \quad \begin{cases} \inf_x f(x) \\ c(x) \in K, \end{cases}$$

where  $f : \mathbb{E} \rightarrow \mathbb{R}$ ,  $c : \mathbb{E} \rightarrow \mathbb{F}$ , and  $K$  is a closed convex set in  $\mathbb{F}$

- ▶ The latter includes the linear (convex) SDO problem

$$(P_{\text{SDO}}) \quad \begin{cases} \inf_X \langle C, X \rangle \\ A(X) = b \\ X \succeq 0. \end{cases}$$



# Preliminaries

## Some features of the course II

- Finite dimension:  $\mathbb{E}$  and  $\mathbb{F}$  are Euclidean vector spaces

- ▶ The unit closed ball is compact
- ▶ A bounded sequence has a converging subsequence
- ▶ Nothing is said about optimal control problems ( $\rightarrow$  see other courses)

- Emphasis on algorithms

- ▶ Their analysis (Do they generate converging sequences? What is the speed of convergence? ...)
- ▶ This requires to know more on theoretical matters (optimality conditions, duality, nonsmooth analysis, ...)
- ▶ Implementation of algorithms is not forgotten (part II of the course)



# Preliminaries

## Outline of the course

### Three parts (60h)

- ① Advanced Continuous Optimization I (30h)
  - ▶ Theory and algorithms in smooth nonconvex optimization
  - ▶ **Goal:** improve the basic knowledge in optimization with some advanced concepts and algorithms
  
- ② Advanced Continuous Optimization II (20 h)
  - ▶ Implementation of algorithms (in Matlab on an easy to implement concrete problem)
  - ▶ **Goal:** improve algorithm understanding
  
- ③ Specialized course by [Yurii Nesterov](#) (12h)



# Preliminaries

## Outline of the course “Advanced Continuous Optimization I”

### Anticipated program (we’ll see ...)

- ① Background in convex analysis and optimization
  - ▶ **Convex analysis:** dual cone, Farkas lemma, asymptotic cone, ...
  - ▶ **Nonsmooth analysis:** multifunction, subdifferential, ...
  - ▶ **Optimization:** optimality conditions, ...
  
- ② Abstract optimality conditions
  - ▶ First order optimality conditions for  $(P_A)$
  - ▶ Second order optimality conditions for  $(P_{EI})$
  
- ③ Linearization methods (Newton)
  - ▶ Josephy-Newton algorithm for nonlinear inclusions
  - ▶ SQP algorithm for  $(P_{EI})$
  - ▶ Semismooth Newton algorithm for nonlinear equations



# Preliminaries

Outline of the course “Advanced Continuous Optimization I”

- 4 Proximal and augmented Lagrangian methods
  - ▶ Proximal algorithm for solving inclusions and optimization problems
  - ▶ Augmented Lagrangian algorithm in optimization
  - ▶ Application to convex quadratic optimization (error bound and global linear convergence)
  
- 5 Interior point methods
  - ▶ Interior point methods in convex optimization
  - ▶ Examples of use in linear, second order cone and semidefinite optimization
  - ▶ Interior point methods in nonlinear optimization

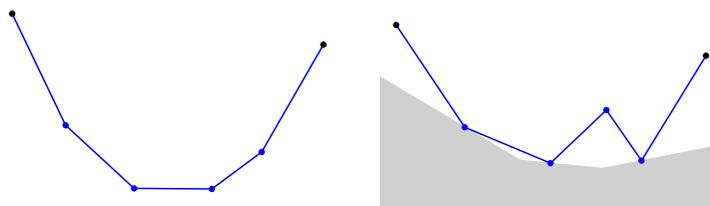
# Preliminaries

Outline of the course “Advanced Continuous Optimization II”

Goal: mastering the implementation of an algorithm

Choose one of the following topics (20 h)

- 1 Implementation of an SQP algorithm, application to the hanging chain problem



- 2 Implementation of an SDO algorithm, application to the OPF problem (Shor relaxation of its QCQP modelling)

# Background

## Convex analysis (notation)

- $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ .
- $\mathbb{R}_+ := \{t \in \mathbb{R} : t \geq 0\}$  and  $\mathbb{R}_{++} := \{t \in \mathbb{R} : t > 0\}$ .
- $B, \bar{B}$ : open and closed unit balls centered at the origin; for  $r \geq 0$ :  
 $B(x, r) = x + rB$  and  $\bar{B}(x, r) = x + r\bar{B}$ .
- $\mathbb{E}, \mathbb{F}, \mathbb{G}$  usually denote Euclidean vector spaces.

# Background

## Convex analysis (relative interior I)

- The **affine hull** of  $P \subset \mathbb{E}$  is the smallest affine space containing  $P$ :

$$\text{aff } P := \bigcap \{A : A \text{ is an affine space containing } P\}.$$

- The **relative interior** of  $P \subset \mathbb{E}$  is its interior in  $\text{aff } P$ :

$$\text{ri } P := \{x \in P : \exists r > 0 \text{ such that } [B(x, r) \cap \text{aff } P] \subset P\}.$$

In **finite dimension**, there holds

$C$ convex and nonempty	$\implies$	$\begin{cases} \text{ri } C \neq \emptyset, \\ \text{aff } C = \text{aff}(\text{ri } C). \end{cases}$
-------------------------	------------	-------------------------------------------------------------------------------------------------------

# Background

Convex analysis (relative interior II)

## Proposition (relative interior criterion)

Let  $C$  be a nonempty convex set and  $x \in \mathbb{E}$ . Then

$$\begin{aligned} x \in \text{ri } C \text{ and } y \in \overline{C} &\implies [x, y[ \subset \text{ri } C. \\ x \in \text{ri } C &\iff \forall x_0 \in C \text{ (or aff } C), \exists t > 1: (1-t)x_0 + tx \in C. \end{aligned}$$

Let  $C$  be a nonempty convex set. Then

- $\text{ri } C$  is convex,
- $\overline{C}$  is convex and  $\text{aff } C = \text{aff } \overline{C}$ ,
- $\text{ri } \overline{C} = \overline{C}$  and  $\text{ri } \overline{C} = \text{ri } C$  (i.e., the last operation prevails).

A point  $x \in C$  is said **absorbing** if  $\forall d \in \mathbb{E}, \exists t > 0$  such that  $x + td \in C$ .

- $x \in \text{int } C \iff x$  is absorbing.



12 / 109

# Background

Convex analysis (dual cone and Farkas lemma)

- The **(positive) dual cone** of a set  $P \subset \mathbb{E}$  is defined by

$$P^+ := \{d \in \mathbb{E} : \langle d, x \rangle \geq 0, \forall x \in P\}.$$

The **(negative) dual cone** of a set  $P$  is  $P^- := -P^+$ .

## Lemma (Farkas, generalized)

Let  $\mathbb{E}$  and  $\mathbb{F}$  be two Euclidean spaces,  $A : \mathbb{E} \rightarrow \mathbb{F}$  a linear map, and  $K$  a nonempty convex cone of  $\mathbb{E}$ . Then

$$\overline{A(K)} = \{y \in \mathbb{F} : A^*y \in K^+\}^+.$$

- $A^* : \mathbb{F} \rightarrow \mathbb{E}$  is defined by:  $\forall (x, y) \in \mathbb{E} \times \mathbb{F}, \langle A^*y, x \rangle = \langle y, Ax \rangle$ .
- One cannot get rid of the closure on  $A(K)$  in general.
- If  $K$  is polyhedral, then  $A(K)$  is polyhedral, hence closed.
- For  $K = \mathbb{E}$ , one recovers  $\mathcal{R}(A) = \mathcal{N}(A^*)^\perp$ .



13 / 109

# Background

## Convex analysis (tangent and normal cones)

- Tangent cone

- ▶ Let  $C$  be a **closed** convex set of  $\mathbb{E}$  and  $x \in C$ .
- ▶ The **cone of feasible directions** for  $C$  at  $x$  is  $T_x^f C := \mathbb{R}_+(C - x)$ .
- ▶ The **tangent cone** to  $C$  at  $x$  is the closure of the previous one

$$T_x C \equiv T_C(x) = \overline{\mathbb{R}_+(C - x)}.$$

- Normal cone

- ▶ Let  $C$  be a convex set of  $\mathbb{E}$  and  $x \in C$ .
- ▶ The **normal cone** to  $C$  at  $x$  is

$$N_x C \equiv N_C(x) = \{d \in \mathbb{E} : \langle x' - x, d \rangle \leq 0, \forall x' \in C\}.$$

- ▶ When  $C$  is closed, there holds

$$N_x C = (T_x C)^-.$$



# Background

## Convex analysis (convex polyhedron)

A **convex polyhedron** in  $\mathbb{E}$  is a set of the form

$$P = \{x \in \mathbb{E} : Ax \leq b\},$$

where  $A : \mathbb{E} \rightarrow \mathbb{R}^m$  is a linear map and  $b \in \mathbb{R}^m$ . It is a closed set. For  $x \in P$ , define

$$I(x) := \{i \in [1 : m] : (Ax - b)_i = 0\}.$$

- If  $T : \mathbb{E} \rightarrow \mathbb{F}$  is linear, then  $T(P)$  is a convex polyhedron.
- If  $P_1$  and  $P_2$  are polyhedra, then  $P_1 + P_2$  is a polyhedron.
- $T_x P = T_x^f P = \{d \in \mathbb{E} : (Ad)_{I(x)} \leq 0\}$ .
- $N_x P = \text{cone}\{A^* e_i : i \in I(x)\}$ .



# Background

## Convex analysis (asymptotic function I)

The **domain** and the **epigraph** of a function  $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$  are the sets

$$\text{dom } f := \{x \in \mathbb{E} : f(x) < +\infty\} \quad \text{and} \quad \text{epi } f := \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} : f(x) \leq \alpha\}.$$

Let  $\overline{\text{Conv}}(\mathbb{E})$  be the set of **closed** (i.e.,  $\text{epi } f$  is closed) **proper** (i.e.,  $\text{epi } f \neq \emptyset$ ) **convex** (i.e.,  $\text{epi } f$  is convex) functions.

### Proposition (asymptotic function $f^\infty$ )

If  $f \in \overline{\text{Conv}}(\mathbb{E})$ , then

- ①  $(\text{epi } f)^\infty$  is the epigraph of a function  $f^\infty : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ ,
- ② for all  $x \in \text{dom } f$  and all  $d \in \mathbb{E}$

$$f^\infty(d) = \lim_{t \rightarrow \infty} \frac{f(x + td) - f(x)}{t} = \lim_{t \rightarrow \infty} \frac{f(x + td)}{t},$$

- ③  $\text{dom } f^\infty \subset (\text{dom } f)^\infty$ ,
- ④  $f^\infty \in \overline{\text{Conv}}(\mathbb{E})$ .

# Background

## Convex analysis (asymptotic function II)

The **sublevel set** of  $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$  of level  $\nu \in \mathbb{R}$  is the set

$$L_\nu(f) := \{x \in \mathbb{E} : f(x) \leq \nu\}.$$

### Proposition (existence of a bounded set of minimizers)

If  $f \in \overline{\text{Conv}}(\mathbb{E})$ , then

- ①  $\forall \nu \in \mathbb{R}$  such that  $L_\nu(f) \neq \emptyset$ , there holds

$$[L_\nu(f)]^\infty = \{d \in \mathbb{E} : f^\infty(d) \leq 0\},$$

- ② the following properties are equivalent:

- ⓐ  $\exists \nu \in \mathbb{R} : L_\nu(f)$  is nonvoid and bounded,
- ⓑ  $\arg \min f$  is nonvoid and bounded,
- ⓒ  $\forall \nu \in \mathbb{R} : L_\nu(f)$  is bounded,
- ⓓ  $\forall d \in \mathbb{E} \setminus \{0\} : f^\infty(d) > 0$ .

# Background

## Nonsmooth analysis (multifunction I)

- A **multifunction**  $T$  (or **set-valued mapping**) between two sets  $\mathbb{E}$  and  $\mathbb{F}$  is a function from  $\mathbb{E}$  to  $\mathcal{P}(\mathbb{F})$ , the set of the subsets of  $\mathbb{F}$ . Notation:

$$T : \mathbb{E} \multimap \mathbb{F} : x \mapsto T(x) \subset \mathbb{F}.$$

Same concept as a **binary relation** (i.e., the data of a part of  $\mathbb{E} \times \mathbb{F}$ ).

- A **graph**, the **domain**, the **range** of  $T : \mathbb{E} \multimap \mathbb{F}$  are defined by

$$\mathcal{G}(T) := \{(x, y) \in \mathbb{E} \times \mathbb{F} : y \in T(x)\},$$

$$\mathcal{D}(T) := \{x \in \mathbb{E} : (x, y) \in \mathcal{G}(T) \text{ for some } y \in \mathbb{F}\} = \pi_{\mathbb{E}}\mathcal{G}(T),$$

$$\mathcal{R}(T) := \{y \in \mathbb{F} : (x, y) \in \mathcal{G}(T) \text{ for some } x \in \mathbb{E}\} = \pi_{\mathbb{F}}\mathcal{G}(T).$$

- The **image** of a part  $P \subset \mathbb{E}$  by  $T$  is

$$T(P) := \bigcup_{x \in P} T(x).$$



# Background

## Nonsmooth analysis (multifunction II)

- The **inverse** of a multifunction  $T : \mathbb{E} \multimap \mathbb{F}$  (it always exists!) is the multifunction  $T^{-1} : \mathbb{F} \multimap \mathbb{E}$  defined by

$$T^{-1}(y) := \{x \in \mathbb{E} : y \in T(x)\}.$$

Hence

$$y \in T(x) \iff x \in T^{-1}(y).$$

- When  $\mathbb{E}, \mathbb{F}$  are topological/metric spaces, a multifunction  $T : \mathbb{E} \multimap \mathbb{F}$  is said to be
  - ▶ **closed at**  $x \in \mathbb{E}$  if  $y \in T(x)$  when  $(x_k, y_k) \in \mathcal{G}(T)$  converges to  $(x, y)$ ,
  - ▶ **closed** if  $\mathcal{G}(T)$  is closed in  $\mathbb{E} \times \mathbb{F}$  (i.e.,  $T$  is closed at any  $x \in \mathbb{E}$ ).
  - ▶ **upper semicontinuous at**  $x \in \mathbb{E}$  if  $\forall \varepsilon > 0, \exists \delta > 0, \forall x' \in x + \delta B : T(x') \subset T(x) + \varepsilon B$ .



# Background

## Nonsmooth analysis (multifunction III)

- When  $\mathbb{E}, \mathbb{F}$  are vector spaces, a multifunction  $T : \mathbb{E} \multimap \mathbb{F}$  is said to be **convex** if  $\mathcal{G}(T)$  is convex in  $\mathbb{E} \times \mathbb{F}$ . This is equivalent to saying that  $\forall (x_0, x_1) \in \mathbb{E}^2$  and  $\forall t \in [0, 1]$ :

$$T((1-t)x_0 + tx_1) \supset (1-t)T(x_0) + tT(x_1).$$

Note that

$$T \text{ convex and } C \text{ convex in } \mathbb{E} \implies T(C) \text{ convex in } \mathbb{F}.$$

# Background

## Nonsmooth analysis (Lipschitz continuity)

- Let  $\mathbb{E}$  and  $\mathbb{F}$  be two normed spaces and  $F : \mathbb{E} \rightarrow \mathbb{F}$  be a function.
- $F$  is **Lipschitz on a set**  $U \subset \mathbb{E}$  if

$$\exists L \geq 0, \quad \forall (x, x') \in U^2 : \quad \|F(x) - F(x')\| \leq L\|x - x'\|.$$

- $F$  is **Lipschitz near**  $x \in \mathbb{E}$  if it is Lipschitz on some neighborhood of  $x$ .
- $F$  is **locally Lipschitz** on an open set  $\Omega \subset \mathbb{E}$  if it is Lipschitz near any point of  $\Omega$ .

## Background

Nonsmooth analysis (generalized differentiability I) [6]

For a function  $F : \mathbb{E} \rightarrow \mathbb{F}$  ( $\mathbb{E}, \mathbb{F}$  are finite dimensional vector spaces), denote

$$\mathcal{D}_F := \{x \in \mathbb{E} : F \text{ is Fréchet-differentiable at } x\}.$$

### Theorem (Rademacher, 1919)

If  $F$  is *Lipschitz* near any point of an open set  $\Omega \subset \mathbb{E}$ , then the Lebesgue measure of  $\Omega \setminus \mathcal{D}_F$  is zero; in particular  $\mathcal{D}_F$  is dense in  $\Omega$ , i.e.,  $\Omega \subset \overline{\mathcal{D}_F}$ .

- The **B-differential** ( $B$  honoring Bouligand) of  $F$  at  $x$  is the set ( $\mathcal{N}$  is an arbitrary negligible set, i.e., with zero Lebesgue measure)

$$\partial_B F(x) := \{J \in \mathcal{L}(\mathbb{E}, \mathbb{F}) : \exists \{x_k\} \subset \mathcal{D}_F \setminus \mathcal{N} \text{ such that } x_k \rightarrow x, F'(x_k) \rightarrow J\}.$$

- The **C-differential** ( $C$  for Clarke) of  $F$  at  $x$  is the set

$$\partial_C F(x) := \text{co } \partial_B F(x).$$



## Background

Nonsmooth analysis (generalized differentiability II) [6]

### Proposition (compactness and upper semicontinuity)

If  $F$  is  $L$ -Lipschitz near  $x$ , then

- 1  $\partial_C F(x)$  is nonempty compact ( $\subset L\bar{B}$ ) and convex,
- 2  $\partial_C F$  is upper semicontinuous at  $x$ .

$\partial_C F(x)$  is said to be **nonsingular** if any  $J \in \partial_C F(x)$  is nonsingular.

### Proposition (nonsingularity diffusion)

If  $F$  is  $L$ -Lipschitz near  $x$  and  $\partial_C F(x)$  is nonsingular, then there are constants  $C > 0$  and  $\delta > 0$  such that

$$\sup_{\substack{x' \in \bar{B}(x, \delta) \\ J \in \partial_C F(x')}} \max(\|J\|, \|J^{-1}\|) \leq C.$$

# Background

## Optimization (a generic problem)

One considers the generic optimization problem

$$(P_X) \quad \begin{cases} \min f(x) \\ x \in X. \end{cases}$$

where

- $f : \mathbb{E} \rightarrow \mathbb{R}$  ( $\mathbb{E}$  is a Euclidean vector space),
- $X$  is a set of  $\mathbb{E}$  (possibly nonconvex).

# Background

## Optimization (tangent cone to a nonconvex set)

- A direction  $d \in \mathbb{E}$  is **tangent to  $X \subset \mathbb{E}$  at  $x \in X$**  if

$$\exists \{x_k\} \subset X, \quad \exists \{t_k\} \downarrow 0 : \quad \frac{x_k - x}{t_k} \rightarrow d.$$

- The **tangent cone** to  $X$  at  $x$  is the set of tangent directions. It is denoted by

$$T_x X \quad \text{or} \quad T_X(x).$$

- **Properties** Let  $x \in X$ .

$T_x X$  is closed.

$$X \text{ is convex} \implies T_x X \text{ is convex and } T_x X = \overline{\mathbb{R}_+(X - x)}.$$

## Background

Optimization (Peano-Kantorovich NC1)

### Theorem (Peano-Kantorovich NC1)

If  $x_*$  is a local minimizer of  $(P_X)$  and  $f$  is differentiable at  $x_*$ , then

$$\nabla f(x_*) \in (T_{x_*} X)^+.$$

- The **gradient** of  $f$  at  $x$  is denoted by  $\nabla f(x) \in \mathbb{E}$  and is defined from the derivative  $f'(x)$  by

$$\forall d \in \mathbb{E} : \quad \langle \nabla f(x_*), d \rangle = f'(x) \cdot d.$$

## Background

Optimization (NSC1 for convex problems)

- Recall that a convex function  $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$  has directional derivatives  $f'(x; d) \in \overline{\mathbb{R}}$  for all  $x \in \text{dom } f$  and all  $d \in \mathbb{E}$ .

### Proposition (NSC1 for a convex problem)

Suppose that  $X$  is convex,  $f$  is convex on  $X$ , and  $x_* \in X$ . Then  $x_*$  is a **global solution** to  $(P_X)$  *if and only if*

$$\forall x \in X : \quad f'(x_*; x - x_*) \geq 0.$$

Proof. Straightforward, using the convexity inequality

$$\forall x \in X : \quad f(x) \geq f(x_*) + f'(x_*; x - x_*).$$

# Background

Optimization (problem  $(P_E)$ )

Let  $\mathbb{E}, \mathbb{F}$  be Euclidean vector spaces. The **equality constrained problem** is

$$(P_E) \quad \begin{cases} \inf_x f(x) \\ c(x) = 0, \end{cases}$$

where  $f : \mathbb{E} \rightarrow \mathbb{R}$ ,  $c : \mathbb{E} \rightarrow \mathbb{F}$  are smooth (possibly non convex) functions.

- The **feasible set** is denoted by

$$X_E := \{x \in \mathbb{E} : c(x) = 0\}.$$

- $(P_E)$  is said to be **convex** if  $f$  is convex and  $c_E$  is affine.



30 / 109

# Background

Optimization (problem  $(P_E)$  – Lagrange optimality conditions)

## Theorem (NC1 for $(P_E)$ , Lagrange)

If  $x_*$  is a local minimum of  $(P_E)$ , if  $f$  and  $c$  are differentiable at  $x_*$ , and if  $c$  is qualified for representing  $X_E$  at  $x_*$  in the sense (2) below, then there exists a multiplier  $\lambda_* \in \mathbb{F}$  such that

$$\nabla_x \ell(x_*, \lambda_*) = 0, \tag{1a}$$

$$c(x_*) = 0. \tag{1b}$$

Some explanations.

- The constraint  $c$  is **qualified** for representing  $X_E$  at  $x$  if

$$T_x X_E = T'_x X_E := \mathcal{N}(c'(x)). \tag{2}$$

Qualification holds if  $c'(x)$  is surjective.

- The **Lagrangian** of  $(P_E)$  is the function

$$\ell : (x, \lambda) \in \mathbb{E} \times \mathbb{F} \mapsto \ell(x, \lambda) = f(x) + \lambda^T c(x).$$



31 / 109

# Background

Optimization (problem  $(P_E)$  – second order optimality conditions)

## Theorem (NC2 for $(P_E)$ )

If  $x_*$  is a local minimum of  $(P_E)$ , if  $f$  and  $c$  are twice differentiable at  $x_*$ , and if (1a) holds for some  $\lambda_* \in \mathbb{F}$ , then

$$\forall d \in T_{x_*} X_E : d^T \nabla_{xx}^2 \ell(x_*, \lambda_*) d \geq 0. \quad (3)$$

- Inequality in (3) is not necessarily true for  $d \in \mathcal{N}(c'(x_*)) \setminus T_{x_*} X_E$ .
- $\nabla_{xx}^2 \ell(x_*, \lambda_*)$  is not necessarily positive semi-definite (even if qualification holds).

## Theorem (SC2 for $(P_E)$ )

If  $f$  and  $c$  are twice differentiable at  $x_*$ , if (1) holds for some  $\lambda_* \in \mathbb{F}$ , and if

$$\forall d \in T_{x_*} X_E \setminus \{0\} : d^T \nabla_{xx}^2 \ell(x_*, \lambda_*) d > 0, \quad (4)$$

then  $x_*$  is a strict local minimum of  $(P_E)$ .

- (4) stronger (hence conclusion holds) if inequality holds  $\forall d \in \mathcal{N}(c'(x_*)) \setminus \{0\}$ .

32 / 109

# Background

Optimization (problem  $(P_{EI})$ )

A generic form of the **nonlinear optimization problem**:

$$(P_{EI}) \quad \begin{cases} \inf_x f(x) \\ c_E(x) = 0 \\ c_I(x) \leq 0, \end{cases}$$

where  $f : \mathbb{E} \rightarrow \mathbb{R}$ ,  $E$  and  $I$  form a partition of  $[1 : m]$ ,  $c_E : \mathbb{E} \rightarrow \mathbb{R}^{m_E}$ , and  $c_I : \mathbb{E} \rightarrow \mathbb{R}^{m_I}$  are smooth (possibly non convex) functions.

- The **feasible set** is denoted by

$$X_{EI} := \{x \in \mathbb{E} : c_E(x) = 0, c_I(x) \leq 0\}.$$

- We say that an inequality constraint is **active** at  $x \in X_{EI}$  if  $c_i(x) = 0$ .
- The **set of indices of active inequality constraints** is denoted by

$$I^0(x) := \{i \in I : c_i(x) = 0\}.$$

- $(P_{EI})$  is said to be **convex** if  $f$  is convex,  $c_E$  is affine, and  $c_I$  is componentwise convex.

33 / 109

# Background

Optimization (problem  $(P_{EI})$  – NC1 or KKT conditions)

## Theorem (NC1 for $(P_{EI})$ , Karush-Kuhn-Tucker (KKT))

If  $x_*$  is a local minimum of  $(P_{EI})$ , if  $f$  and  $c = (c_E, c_I)$  are differentiable at  $x_*$ , and if  $c$  is qualified for representing  $X_{EI}$  at  $x_*$  in the sense (6) below, then there exists a multiplier  $\lambda_* \in \mathbb{R}^m$  such that

$$\nabla_x \ell(x_*, \lambda_*) = 0, \quad (5a)$$

$$c_E(x_*) = 0, \quad (5b)$$

$$0 \leq (\lambda_*)_I \perp c_I(x_*) \leq 0. \quad (5c)$$

Some explanations.

- The **Lagrangian** of  $(P_{EI})$  is the function

$$\ell : (x, \lambda) \in \mathbb{E} \times \mathbb{R}^m \mapsto \ell(x, \lambda) = f(x) + \lambda^T c(x).$$

- The **complementarity condition** (5c) means

$$(\lambda_*)_I \geq 0, \quad (\lambda_*)_I^T c_I(x_*) = 0, \quad \text{and} \quad c_I(x_*) \leq 0.$$



# Background

Optimization (problem  $(P_{EI})$  – constraint qualification)

- The tangent cone  $T_x X_{EI}$  is always contained in the **linearizing cone**

$$T'_x X_{EI} := \{d \in \mathbb{E} : c'_E(x) \cdot d = 0, \quad c'_{I^0(x)}(x) \cdot d \leq 0\}.$$

- It is said that the constraint  $c$  is **qualified** for representing  $X_{EI}$  at  $x$  if

$$T_x X_{EI} = T'_x X_{EI}. \quad (6)$$

- Sufficient conditions of qualification: continuity and/or differentiability and **one** of the following

- ▶ (CQ-A)  $c_{E \cup I^0(x)}$  is affine near  $x$  (**Affinity**),
- ▶ (CQ-S)  $c_E$  is affine,  $c_{I^0(x)}$  is componentwise convex,  $\exists \hat{x} \in X_{EI}$  such that  $c_{I^0(x)}(\hat{x}) < 0$  (**Slater**),
- ▶ (CQ-LI)  $\sum_{i \in E \cup I^0(x)} \alpha_i \nabla c_i(x) = 0 \implies \alpha = 0$  (**Linear Independence**),
- ▶ (CQ-MF)  $\sum_{i \in E \cup I^0(x)} \alpha_i \nabla c_i(x) = 0$  and  $\alpha_{I^0(x)} \geq 0 \implies \alpha = 0$  (**Mangasarian-Fromovitz**).



## Background

Optimization (problem  $(P_{EI})$  – more on constraint qualification)

### Proposition (other forms of (CQ-MF))

Suppose that  $c_{E \cup I^0(x)}$  is differentiable at  $x \in X_{EI}$ . Then the following properties are equivalent:

- i (CQ-MF) holds at  $x$ ,
- ii  $\forall v \in \mathbb{R}^m, \exists d \in \mathbb{E}: c'_E(x) \cdot d = v_E$  and  $c'_{I^0(x)}(x) \cdot d \leq v_{I^0(x)}$ ,
- iii  $c'_E(x)$  is surjective and  $\exists d \in \mathbb{E}: c'_E(x) \cdot d = 0$  and  $c'_{I^0(x)}(x) \cdot d < 0$ .

## Background

Optimization (problem  $(P_{EI})$  – SC1 for convex problem)

### Theorem (SC1 for convex $(P_{EI})$ )

If  $(P_{EI})$  is **convex**, if  $f$  and  $c$  are differentiable at  $x_* \in X_{EI}$ , and if there is a  $\lambda_* \in \mathbb{F}$  such that  $(x_*, \lambda_*)$  satisfies (5a)-(5b), then  $x_*$  is a **global** minimum of  $(P_{EI})$ .

No need of constraint qualification.

The goal of the first part of this course is to extend the previous NC1 and SC1 to a more general problem and to derive second order optimality conditions.

# Background

Optimization (problem  $(P_{EI})$  – augmented Lagrangian)

The **augmented Lagrangian**  $\ell_r : \mathbb{E} \times \mathbb{R}^m \rightarrow \mathbb{R}$  of problem  $(P_{EI})$  reads

$$\begin{aligned} \ell_r(x, \lambda) &:= \inf_{s \geq 0} f(x) + \lambda_E^T c_E(x) + \frac{r}{2} \|c_E(x)\|_2^2 + \lambda_I^T (c_I(x) + s) + \frac{r}{2} \|c_I(x) + s\|_2^2 \\ &= f(x) + \lambda_E^T c_E(x) + \frac{r}{2} \|c_E(x)\|_2^2 + \frac{1}{2r} \sum_{i \in I} [\max(0, \lambda_i + r c_i(x))^2 - \lambda_i^2] \\ &= f(x) + \lambda^T \tilde{c}_{\lambda, r}(x) + \frac{r}{2} \|\tilde{c}_{\lambda, r}(x)\|_2^2, \end{aligned}$$

where  $[\tilde{c}_{\lambda, r}(x)]_i = c_i(x)$  if  $i \in E$ ,  $= \max(c_i(x), \lambda_i/r)$  if  $i \in I$ .

## Theorem (exactness of the augmented Lagrangian)

Suppose that  $f$  and  $c_{E \cup I^0(x_*)}$  are twice differentiable near a local minimum  $x_*$  of  $(P_{EI})$ , that  $(x_*, \lambda_*)$  satisfies KKT, and that the semi-strong SC2 holds. Then,  $\exists r_{\min} > 0$  and a neighborhood  $V$  of  $x_*$  such that  $\forall r \geq r_{\min}$ ,  $x_*$  is a strict minimizer of  $\ell_r(\cdot, \lambda_*)$  on  $V$ .

38 / 109

# Background

Optimization (linear optimization duality)

- Let  $c \in \mathbb{E}$  (a Euclidean vector space),  $A : \mathbb{E} \rightarrow \mathbb{R}^m$  and  $B : \mathbb{E} \rightarrow \mathbb{R}^p$  linear,  $a \in \mathbb{R}^m$ , and  $b \in \mathbb{R}^p$ .
- A **linear optimization problem**  $(P_L)$  and its **dual**  $(D_L)$  read

$$(P_L) \quad \begin{cases} \inf_{x \in \mathbb{E}} \langle c, x \rangle \\ Ax = a \\ Bx \leq b \end{cases} \quad \text{and} \quad (D_L) \quad \begin{cases} \sup_{(y, s) \in \mathbb{R}^m \times \mathbb{R}^p} a^T y - b^T s \\ A^* y - B^* s = c \\ s \geq 0. \end{cases}$$

- Properties**

$$\begin{aligned} (P_L) \text{ has a solution} &\iff \text{val}(P_L) \in \mathbb{R}, \\ \text{val}(D_L) \leq \text{val}(P_L) &\quad \text{[named weak duality]}, \\ (P_L), (D_L) \text{ feasible} &\iff \text{Sol}(P_L) \neq \emptyset \iff \text{Sol}(D_L) \neq \emptyset. \quad (7) \end{aligned}$$

When (7) holds [named **strong duality**],  $\text{val}(D_L) = \text{val}(P_L)$ .

39 / 109

# Background

Algorithmics (speeds of convergence)

Let  $\mathbb{E}$  be a normed space and  $\{x_k\} \subset \mathbb{E}$  be a sequence converging to  $\bar{x}$ .

- $\{x_k\}$  is said to converge **linearly**, if  $\exists r \in [0, 1[$  and  $K \in \mathbb{N}$  such that  $\forall k \geq K$ , there holds  $\|x_{k+1} - \bar{x}\| \leq r\|x_k - \bar{x}\|$ .
  - ▶ Depends on the norm of  $\mathbb{E}$ .
- $\{x_k\}$  is said to converge **superlinearly**, if  $x_{k+1} - \bar{x} = o(\|x_k - \bar{x}\|)$ .
  - ▶ Independent of the norm of  $\mathbb{E}$ .
  - ▶ Faster than linear convergence.
  - ▶ Typical of the quasi-Newton methods.
- $\{x_k\}$  is said to converge **quadratically**, if  $x_{k+1} - \bar{x} = O(\|x_k - \bar{x}\|^2)$ .
  - ▶ Independent of the norm of  $\mathbb{E}$ .
  - ▶ Faster than superlinear convergence.
  - ▶ Typical of Newton's method.



41 / 109

# Background

Algorithmics (global convergence in unconstrained optimization - line-search)

Consider the problem of minimizing a function  $f : \mathbb{E} \rightarrow \mathbb{R}$  (no constraint):

$$\min_{x \in \mathbb{E}} f(x).$$

**Global convergence** means “only”  $g_k := \nabla f(x_k) \rightarrow 0$  (or even something weaker). It can be obtained with line-search or trust-region algorithms.

- A **line-search algorithm** generates a sequence  $\{x_k\}$  as follows.
  - ▶ Choice of a **descent direction** at  $x_k$ , i.e.,  $d_k \in \mathbb{E}$  such that  $f'(x_k) \cdot d_k < 0$  (standard examples: gradient, conjugate gradient, Newton, quasi-Newton, Gauss-Newton).
  - ▶ Determination of a step-size  $\alpha_k > 0$  by a **line-search rule** to force the decrease of  $f$  along  $d_k$  (standard examples: Cauchy, Armijo, Wolfe).
  - ▶  $x_{k+1} := x_k + \alpha_k d_k$ .

## Proposition (global convergence by line-search)

*If  $\{f(x_k)\}$  is bounded below, then  $\sum_{k \geq 1} \|g_k\|^2 \cos^2 \theta_k < +\infty$ , where  $\theta_k := \arccos \langle -g_k, d_k \rangle / (\|g_k\| \|d_k\|)$  is the angle between  $d_k$  and  $-g_k$ .*

42 / 109

# Background

Algorithmics (global convergence in unconstrained optimization - trust-region)

- A **trust-region algorithm** generates a sequence  $\{x_k\}$  as follows.
  - ▶ Choice of a **model**  $\varphi$  of  $f$  around  $x_k$ , usually quadratic:  
 $\varphi(s) := \langle g_k, s \rangle + \frac{1}{2} \langle M_k s, s \rangle$  (standard models: gradient, Newton, quasi-Newton, Gauss-Newton).
  - ▶ Determination of a trust-radius  $\Delta_k > 0$  such that  $f(x_k + s_k)$  is sufficiently smaller than  $f(x_k)$ , where  $s_k \in \arg \min \{\varphi(s) : \|s\| \leq \Delta_k\}$ .
  - ▶  $x_{k+1} := x_k + s_k$ .

## Proposition (global convergence by trust-region)

If  $\{f(x_k)\}$  is bounded below and  $\{M_k\}$  is bounded, then  $g_k \rightarrow 0$ .



43 / 109

# Background

Algorithmics (global convergence for nonlinear systems)

Let  $F : \mathbb{E} \rightarrow \mathbb{E}$  and consider the problem of finding  $x$  such that

$$F(x) = 0.$$

- **No algorithm** with global convergence (exception, e.g., when  $F$  is polynomial, but the methods are expensive or for rather small dimensions).
- A less ambitious goal: minimizing the **least-squares objective**

$$\min_{x \in \mathbb{E}} \left( \varphi(x) := \frac{1}{2} \|F(x)\|_2^2 \right).$$

- **Amazing but misleading fact:** if  $\nabla \varphi(x_k) \neq 0$ , the Gauss-Newton direction

$$d_k^{\text{GN}} \in \arg \min_{d \in \mathbb{E}} \|F(x_k) + F'(x_k)d\|_2^2$$

is a descent direction of  $\varphi$  at  $x_k$ . May yield false convergence. It is **better to use the trust-region approach**:  $x_{k+1} = x_k + s_k$ , where

$$s_k \in \arg \min_{\|s\|_2 \leq \Delta_k} \|F(x_k) + F'(x_k)s\|_2^2,$$

with well adapted trust-radius  $\Delta_k > 0$ .



44 / 109

# Background

Algorithmics (Dennis & Moré criterion for superlinear convergence)

Let  $F : \mathbb{E} \rightarrow \mathbb{F}$  and consider the nonlinear system to solve in  $x \in \mathbb{E}$ :

$$F(x) = 0.$$

A [quasi-Newton algorithm](#) locally generates a sequence  $\{x_k\}$  by the recurrence

$$F(x_k) + M_k(x_{k+1} - x_k) = 0, \quad (8)$$

where  $M_k \in \mathcal{L}(\mathbb{E}, \mathbb{F})$  is an approximation of  $F'(x_k)$ , generated by the algorithm.

## Proposition (Dennis & Moré criterion for superlinear convergence)

If

- $F$  is differentiable at a zero  $x_*$  of  $F$ ,

- $F'(x_*)$  is nonsingular,
- $\{x_k\}$  generated by (8) converges to  $x_*$ ,

then the convergence is superlinear if and only if

$$(M_k - F'(x_*))(x_{k+1} - x_k) = o(\|x_{k+1} - x_k\|).$$

45 / 109

# Background

Algorithmics (augmented Lagrangian algorithm for  $(P_{EI})$ )

The [augmented Lagrangian algorithm](#) or [method of multipliers](#) is a dual algorithm: it generates a sequence of multipliers  $\{\lambda_k\} \subset \mathbb{R}^m$  and penalty parameters  $\{r_k\} \subset \mathbb{R}_{++}$  by

- 1 If  $\tilde{c}_{\lambda_k, r_k}(x) \simeq 0$ , stop.
- 2  $x_{k+1} \in \arg \min \{l_{r_k}(x, \lambda_k) : x \in \mathbb{E}\}$  (approximately is fine).
- 3  $\lambda_{k+1} := \lambda_k + r_k \tilde{c}_{\lambda_k, r_k}(x)$  (sometimes  $r_k \curvearrowright \alpha_k$ , a stepsize).
- 4  $r_{k+1} = ?$  (heuristics).

## Remarks

- Tries to find  $\lambda$  and  $r$  such that the  $l_r(\cdot, \lambda)$  has  $x_*$  for local minimizer.
- When it stops in step 1, a stationary point is (approximately) found.
- $l_r(\cdot, \lambda)$  is differentiable, so that it can be minimized by quasi-Newton algorithms.
- Step 3 is motivated by the formula

$$\nabla_x l_r(x, \lambda) = \nabla f(x) + c'(x)^T (\lambda + r \tilde{c}_{\lambda, r}(x)).$$

# Optimality conditions

First order optimality conditions (the problem  $(P_A)$ )

- We consider the problem

$$(P_A) \quad \begin{cases} \min f(x) \\ c(x) \in K, \end{cases}$$

where

- ▶  $f : \mathbb{E} \rightarrow \mathbb{R}$  ( $\mathbb{E}$  is a Euclidean vector space),
  - ▶  $c : \mathbb{E} \rightarrow \mathbb{F}$  ( $\mathbb{F}$  is another Euclidean vector space),
  - ▶  $K$  is nonempty **closed convex** set in  $\mathbb{F}$ ,
- The **feasible set** is denoted by
$$X_A := \{x \in \mathbb{E} : c(x) \in K\} = c^{-1}(K).$$
- $(P_A)$  is said to be **convex** if
    - ▶  $f$  is convex,
    - ▶  $T : \mathbb{E} \rightarrow \mathbb{F} : x \mapsto c(x) - K$  is convex (then  $X_A$  is convex).



48 / 109

# Optimality conditions

First order optimality conditions (tangent and linearizing cones, qualification)

## Proposition (tangent and linearizing cones)

If  $c$  is differentiable at  $x \in X_A$ , then

$$T_x X_A \subset T'_x X_A := \{d \in \mathbb{E} : c'(x) \cdot d \in T_{c(x)} K\}.$$

- $T'_x X_A$  is called the **linearizing cone** to  $X$  at  $x$ .
- The equality  $T_x X_A = T'_x X_A$  is not guaranteed.
- The constraint function  $c$  is said to be **qualified** for representing  $X_A$  at  $x$  if

$$T_x X_A = T'_x X_A, \tag{9a}$$

$$c'(x)^* [(T_{c(x)} K)^+] \text{ is closed.} \tag{9b}$$



49 / 109

# Optimality conditions

First order optimality conditions (NC1 for  $(P_A)$ )

## Theorem (NC1 for $(P_A)$ )

If  $x_*$  is a local minimum of  $(P_A)$ , if  $f$  and  $c$  are differentiable at  $x_*$ , and if  $c$  is qualified for representing  $X_A$  at  $x_*$  in the sense (9a)-(9b), then

- 1 there exists a multiplier  $\lambda_* \in \mathbb{F}$  such that

$$\nabla f(x_*) + c'(x_*)^* \lambda_* = 0, \quad (10a)$$

$$\lambda_* \in (\mathbb{T}_{c(x_*)} K)^-, \quad (10b)$$

- 2 if, furthermore,  $K$  is a convex cone, then (10b) can be written

$$K^- \ni \lambda_* \perp c(x_*) \in K. \quad (10c)$$

One recognizes in (10a) the gradient of the Lagrangian wrt  $x$

$$\ell : \mathbb{E} \times \mathbb{F} \rightarrow \mathbb{R} : x \mapsto \ell(x, \lambda) = f(x) + \langle \lambda, c(x) \rangle$$

and in (10c) the complementarity conditions.



# Optimality conditions

First order optimality conditions (SC1 for convex  $(P_A)$ )

## Theorem (SC1 for convex $(P_A)$ )

If  $(P_A)$  is convex, if  $f$  and  $c$  are differentiable at  $x_* \in X_A$ , and if there is a  $\lambda_* \in \mathbb{F}$  such that  $(x_*, \lambda_*)$  satisfies (10a)-(10b), then  $x_*$  is a global minimum of  $(P_A)$ .

The goal of the next slides is to highlight and analyze a condition (Robinson's condition) that

- 1 provides an error bound for the feasible set  $X_A$ ,
- 2 ensures the stability of the feasible set  $X_A$ ,
- 3 ensures that  $c$  is qualified for representing  $X_A$ .



# Optimality conditions

First order optimality conditions (Robinson's condition)

We say that **Robinson's condition** holds at  $x \in X_A$  if

$$(CQ-R) \quad \boxed{0 \in \text{int}\left(c(x) + c'(x) \cdot \mathbb{E} - K\right)}. \quad (11)$$

We will see below that

- it is useful for
  - ▶ getting an error bound for  $X_A$ ,
  - ▶ ensuring the stability of the feasible set  $X_A$  with respect to small perturbations,
  - ▶ showing that the constraint function  $c$  is qualified for representing  $X_A$  at  $x$ , in the sense (9a)-(9b),
- it generalizes to  $(P_A)$  the Mangasarian-Fromovitz constraint qualification (CQ-MF) for  $(P_{EI})$ .



52 / 109

# Optimality conditions

First order optimality conditions (Robinson's error bound I) [21]

Link between Robinson's condition (CQ-R) and distances to the perturbed set  $X_A$ .

## Theorem (Robinson's error bounds)

*If  $c$  is continuously differentiable near  $x_0 \in X_A$ , then the following properties are equivalent:*

- 1 (CQ-R) holds at  $x = x_0$ ,
- 2 there exists a constant  $\mu \geq 0$ , such that  $\forall (x, y)$  near  $(x_0, 0)$ :

$$\text{dist}(x, c^{-1}(y + K)) \leq \mu \text{dist}(c(x), y + K). \quad (12)$$



53 / 109

## Optimality conditions

First order optimality conditions (Robinson's error bound II) [21]

An **error bound** is an estimation of the distance to a set by a quantity easier to compute.

### Corollary (Robinson's error bound)

If  $c$  is continuously differentiable near  $x_0 \in X_A$  and if (CQ-R) holds at  $x = x_0$ , then there exists a constant  $\mu \geq 0$ , such that

$$\forall x \text{ near } x_0 : \quad \text{dist}(x, X_A) \leq \mu \text{dist}(c(x), K). \quad (13)$$

- $\text{dist}(x, X_A)$  is often difficult to evaluate in  $\mathbb{E}$ ,
- $\text{dist}(c(x), K)$  is often easier to evaluate in  $\mathbb{F}$  (it is the case if  $c(x)$  is easy to evaluate and  $K$  is simple),
- useful in *theory* (e.g., for proving that (CQ-R) implies constraint qualification), in *algorithmics* (e.g., for proving [speed of] converge) or in *practice* (for estimating  $\text{dist}(x, X_A)$ ).



54 / 109

## Optimality conditions

First order optimality conditions (stability wrt small perturbations)

The estimate (12) readily implies the following stability result.

### Corollary (stability wrt small perturbations)

If  $c$  is continuously differentiable near some  $x_0 \in X_A$  and if (CQ-R) holds at  $x_0$ , then, for all  $p \in \mathbb{F}$  near 0:

$$\{x \in \mathbb{E} : c(x) + p \in K\} \neq \emptyset.$$



55 / 109

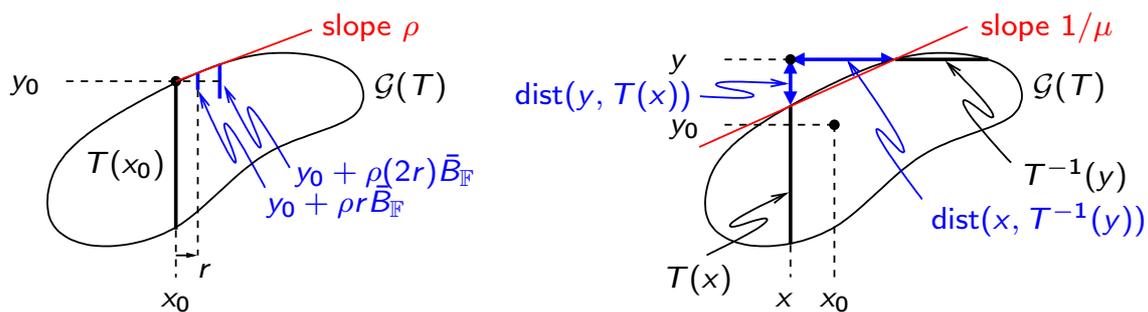
# Optimality conditions

First order optimality conditions (open multifunction theorem I) [27, 23, 5]

Let  $T : \mathbb{E} \multimap \mathbb{F}$  be a multifunction and  $\bar{B}_{\mathbb{E}}$  and  $\bar{B}_{\mathbb{F}}$  be the closed balls of  $\mathbb{E}$  and  $\mathbb{F}$ .

- $T$  is **open** at  $(x_0, y_0) \in \mathcal{G}(T)$  with **ratio**  $\rho > 0$  if there exist a neighborhood  $W$  of  $(x_0, y_0)$  and a radius  $r_{\max} > 0$  such that for all  $(x, y) \in W \cap \mathcal{G}(T)$  and  $r \in [0, r_{\max}]$ :  $y + \rho r \bar{B}_{\mathbb{F}} \subset T(x + r \bar{B}_{\mathbb{E}})$ .
- $T$  is **metric regular** at  $(x_0, y_0) \in \mathcal{G}(T)$  with **modulus**  $\mu > 0$  if for all  $(x, y)$  near  $(x_0, y_0)$ :  $\text{dist}(x, T^{-1}(y)) \leq \mu \text{dist}(y, T(x))$ .

Difference:  $(x, y) \in \mathcal{G}(T)$  for the openness, but not for the metric regularity.



Both estimate the change in  $T(x)$  with  $x$  (but these are not infinitesimal notions). either from inside  $G(T)$  (openness) or outside (metric regularity).

*Inria*

56 / 109

# Optimality conditions

First order optimality conditions (open multifunction theorem II) [27, 23, 5]

Extention of the *open mapping theorem* for linear (continuous) maps to (nonlinear) convex multifunctions.

## Theorem (open multifunction theorem, finite dimension)

If  $T : \mathbb{E} \multimap \mathbb{F}$  is **convex** and  $(x_0, y_0) \in \mathcal{G}(T)$ , then the following properties are equivalent:

- 1  $y_0 \in \text{int } \mathcal{R}(T)$ ,
- 2 for all  $r > 0$ ,  $y_0 \in \text{int } T(x_0 + r \bar{B}_{\mathbb{E}})$ ,
- 3  $T$  is open at  $(x_0, y_0)$  with rate  $\rho > 0$ ,
- 4  $T$  is metric regular at  $(x_0, y_0)$  with modulus  $\mu > 0$ .

One can take  $\mu = 1/\rho$  in point 4 if  $\rho$  is given by point 3.

*Inria*

57 / 109

# Optimality conditions

First order optimality conditions (metric regularity diffusion I) [7]

- $\varphi : \mathbb{E} \rightarrow \bar{\mathbb{R}}_+$  is **subcontinuous** if for all sequences  $\{x_k\} \rightarrow x$  such that  $\varphi(x_k) \rightarrow 0$ , there holds  $\varphi(x) = 0$ .
- $\varphi : \mathbb{E} \rightarrow \bar{\mathbb{R}}_+$  is  **$r$ -steep** at  $x \in \mathbb{E}$ , with  $r \in [0, 1[$ , if for all  $x' \in \bar{B}(x, \varphi(x)/(1-r))$ , there exists  $x'' \in \bar{B}(x', \varphi(x'))$ , such that  $\varphi(x'') \leq r \varphi(x')$ .

## Lemma (Lyusternik [18, 7])

If  $\varphi : \mathbb{E} \rightarrow \bar{\mathbb{R}}_+$  is **subcontinuous** and  **$r$ -steep** at  $x \in \text{dom } \varphi$ , with  $r \in [0, 1[$ , then  $\varphi$  has a zero in  $\bar{B}(x, \varphi(x)/(1-r))$ .

# Optimality conditions

First order optimality conditions (metric regularity diffusion II) [7]

## Theorem (metric regularity diffusion)

Suppose that

- $c : \mathbb{E} \rightarrow \mathbb{F}$  a continuous function,  $K \neq \emptyset$  a closed convex set of  $\mathbb{F}$ ,
- $T : \mathbb{E} \rightarrow \mathbb{F} : x \mapsto c(x) - K$  is  $\mu$ -metric regular at  $(x_0, y_0) \in \mathcal{G}(T)$ ,
- $\delta : \mathbb{E} \rightarrow \mathbb{F}$  is Lipschitz near  $x_0$  with modulus  $L < 1/\mu$ ,
- $\tilde{T} : \mathbb{E} \rightarrow \mathbb{F} : x \mapsto c(x) + \delta(x) - K$ .

Then  $\tilde{T}$  is also metric regular at  $(x_0, y_0 + \delta(x_0)) \in \mathcal{G}(\tilde{T})$  with modulus  $\mu/(1-L\mu)$ : for all  $(x, y)$  near  $(x_0, y_0 + \delta(x_0))$ , there holds

$$\text{dist}(x, \tilde{T}^{-1}(y)) \leq \frac{\mu}{1-L\mu} \text{dist}(y, \tilde{T}(x)). \quad (14)$$

No need of convexity.

# Optimality conditions

First order optimality conditions (qualification with (CQ-R) I)

## Proposition (other forms of (CQ-R))

Suppose that  $c$  is differentiable at  $x \in X_A$ . Then the following properties are equivalent

- $0 \in \text{int}(c(x) + c'(x)\mathbb{E} - K)$  [this is (CQ-R)],
- $c'(x)\mathbb{E} - T_{c(x)}^f K = \mathbb{F}$ ,
- $c'(x)\mathbb{E} - T_{c(x)} K = \mathbb{F}$  [corresponds to (CQ-MF) (ii)],
- $\overline{c'(x)\mathbb{E} - T_{c(x)} K} = \mathbb{F}$ .

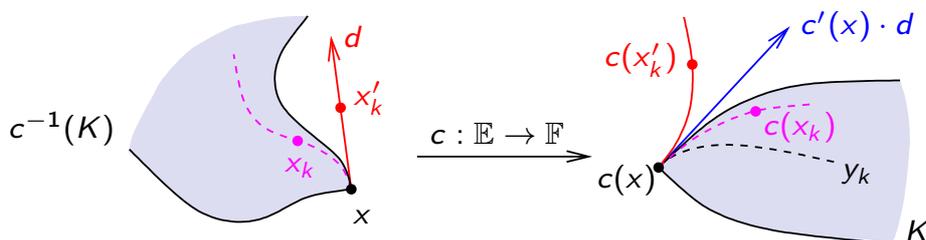
# Optimality conditions

First order optimality conditions (qualification with (CQ-R) II)

## Proposition (qualification with (CQ-R))

Suppose that  $c$  is continuously differentiable near  $x \in X_A$  and that (CQ-R) holds at  $x$ . Then  $c$  is qualified for representing  $X_A$  at  $x$ .

The figure below shows how (CQ-R) is used to create an appropriate sequence  $\{x_k\}$  in  $X_A := c^{-1}(K)$  to get qualification (the figure requires some oral explanations, though ...).



# Optimality conditions

First order optimality conditions (qualification with (CQ-R) III)

## Proposition (Gauvin's boundedness property)

Suppose that  $f$  and  $c$  are differentiable at  $x_* \in X_A$  and that the set  $\Lambda_*$  of multipliers  $\lambda_* \in \mathbb{F}$  satisfying (10a)-(10b) is nonempty. Then

- 1  $\Lambda_*^\infty = [c'(x_*)\mathbb{E} - T_{c(x_*)} K]^+$ ,
- 2  $\Lambda_*$  is bounded if and only if (CQ-R) holds.

The property was originally established for problem  $(P_{EI})$  and (CQ-MF) [9].

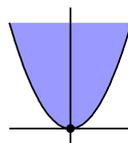
# Optimality conditions

Second order optimality conditions for  $(P_{EI})$  (three instructive examples)

## Three instructive examples

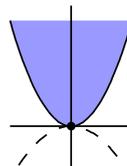
- 1 Strong NC2 (not always true):  $\forall \lambda_* \in \Lambda_*, \forall d \in C_*: \langle L_* d, d \rangle \geq 0$ .

$$\begin{cases} \min x_2 \\ x_2 \geq x_1^2, \end{cases}$$



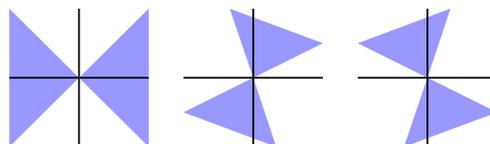
- 2 Semi-strong NC2 (not always true):  $\exists \lambda_* \in \Lambda_*, \forall d \in C_*: \langle L_* d, d \rangle \geq 0$ .

$$\begin{cases} \min x_2 \\ x_2 \geq x_1^2 \\ x_2 \geq -\frac{1}{2}x_1^2. \end{cases}$$



- 3 Weak NC2 (always true):  $\forall d \in C_*, \exists \lambda_* \in \Lambda_*: \langle L_* d, d \rangle \geq 0$ .

$$\begin{cases} \min x_3 \\ x_3 \geq (x_1 + x_2)(x_1 - x_2) \\ x_3 \geq (x_2 + 3x_1)(2x_2 - x_1) \\ x_3 \geq (2x_2 + x_1)(x_2 - 3x_1). \end{cases}$$



# Optimality conditions

Second order optimality conditions for  $(P_{EI})$  (critical cone)

- **Critical cone** at  $x \in X_{EI}$  (it is a part of  $T'_x X_{EI}$ )

$$C(x) := \{d \in \mathbb{E} : c'_E(x) \cdot d = 0, c'_{I^0}(x) \cdot d \leq 0, f'(x) \cdot d \leq 0\}.$$

- Short notation for index sets

$$\begin{aligned} I_*^0 &:= \{i \in I : c_i(x_*) = 0\} := I^0(x_*), \\ I_*^{0+} &:= \{i \in I : c_i(x_*) = 0, (\lambda_*)_i > 0\}, \\ I_*^{00} &:= \{i \in I : c_i(x_*) = 0, (\lambda_*)_i = 0\}. \end{aligned}$$

- Other forms of the critical cone at a stationary pair  $(x_*, \lambda_*)$ :

$$\begin{aligned} C_* &= \{d \in \mathbb{E} : c'_E(x_*) \cdot d = 0, c'_{I^0}(x_*) \cdot d \leq 0, f'(x_*) \cdot d = 0\}, \\ &= \{d \in \mathbb{E} : c'_{E \cup I_*^{0+}}(x_*) \cdot d = 0, c'_{I_*^{00}}(x_*) \cdot d \leq 0\}. \end{aligned}$$



65 / 109

# Optimality conditions

Second order optimality conditions (NC2)

Notation at a stationary point  $x_*$ :

- $\Lambda_* := \{\lambda_* \in \mathbb{R}^m : (x_*, \lambda_*) \text{ satisfies the KKT system (5)}\}$ .
- $L_* := \nabla_{xx}^2 \ell(x_*, \lambda_*)$  for some specified  $\lambda_* \in \Lambda_*$ .

## Theorem (NC2 for $(P_{EI})$ )

Suppose that

- $x_*$  is a local minimum of  $(P_{EI})$ ,
- $f$  and  $c_E$  are  $C^2$  near  $x_*$ ,  $c_{I^0}$  is twice differentiable at  $x_*$ ,  $c_{I \setminus I^0}$  is continuous at  $x_*$ ,
- (CQ-MF) holds at  $x_*$ ,

then  $\forall d \in C_*, \exists \lambda_* \in \Lambda_*$  such that  $\langle L_* d, d \rangle \geq 0$ .

These conditions are named **weak second order necessary conditions** and also read

$$\forall d \in C_* : \max_{\lambda_* \in \Lambda_*} \langle L_* d, d \rangle \geq 0. \quad (15)$$



66 / 109

# Optimality conditions

## Second order optimality conditions (SC2)

### Theorem (SC2 for $(P_{EI})$ )

Suppose that

- $f$  and  $c_{E \cup I_0^*}$  are twice differentiable at  $x_*$ ,
- $(x_*, \lambda_*)$  verifies the KKT conditions (5),
- the following equivalent conditions hold

$$\forall d \in C_* \setminus \{0\}, \quad \exists \lambda_* \in \Lambda_* : \quad \langle L_* d, d \rangle > 0, \quad (16)$$

$$\exists \bar{\gamma} > 0, \quad \forall d \in C_*, \quad \exists \lambda_* \in \Lambda_* : \quad \langle L_* d, d \rangle \geq \bar{\gamma} \|d\|^2, \quad (17)$$

then  $\forall \gamma \in [0, \bar{\gamma}[$ ,  $\exists$  a neighborhood  $V$  of  $x_*$ ,  $\forall x \in (V \setminus \{x_*\}) \cap X_{EI}$ :

$$f(x) > f(x_*) + \frac{\gamma}{2} \|x - x_*\|^2. \quad (18)$$

In particular,  $x_*$  is a strict local minimum of  $(P_{EI})$ .

67 / 109

# Optimality conditions

## Second order optimality conditions (SC2)

- Condition (18) is called the **quadratic growth property**.
- The property

$$\exists \lambda_* \in \Lambda_* : \quad \forall d \in C_* \setminus \{0\}, \quad \langle L_* d, d \rangle > 0$$

is stronger than (16)-(17) and is called the **semi-strong SC2**.

- The even stronger property

$$\forall \lambda_* \in \Lambda_* : \quad \forall d \in C_* \setminus \{0\}, \quad \langle L_* d, d \rangle > 0$$

is called the **strong SC2**.

# Linearization methods

## Overview

Two classes of linearization methods that are used to solve systems with nonsmoothness.

- 1 Methods that capture much of the local behavior of the system.

### Examples

- ▶ The [Josephy-Newton algorithm](#) for inclusions.
- ▶ The [SQP algorithm](#) for  $(P_{EI})$ .

### Features

- ▶ Expensive iteration, easy to globalize.

- 2 Methods that use a single piece of the local behavior of the system.

### Example

- ▶ The [semismooth Newton algorithm](#) for nonsmooth system of equations.

### Features

- ▶ Cheap iteration, difficult to globalize.



# Linearization methods

## Josephy-Newton algorithm for inclusions (the problem to solve I)

Let  $\mathbb{E}$  and  $\mathbb{F}$  be Euclidean vector spaces of the same dimension,  $F : \mathbb{E} \rightarrow \mathbb{F}$  be a smooth function, and  $N : \mathbb{E} \multimap \mathbb{F}$  be a multifunction.

We consider the [inclusion problem](#)

$$(P_{IN}) \quad \boxed{F(x) + N(x) \ni 0.} \quad (19)$$

Interested in algorithmic issues (not theoretical ones, like existence of solution).

### Examples

Let  $X \subset \mathbb{E}$  and  $N_X(x) := T_X(x)^-$  its [normal cone](#) at  $x$  ( $= \emptyset$  if  $x \notin X$ ).

- 1 The [variational problem](#) reads

$$(P_V) \quad F(x) + N_X(x) \ni 0. \quad (20)$$

- 2 The [variational inequality problem](#) is  $(P_V)$  with  $X = C$  (a convex):

$$(P_{VI}) \quad \begin{cases} x \in C \\ \langle F(x), y - x \rangle \geq 0, \quad \forall y \in C. \end{cases} \quad (21)$$

- 3 The [complementarity problem](#) is  $(P_V)$  with  $X = K$  (a convex cone):

$$(P_{NCP}) \quad K \ni x \perp F(x) \in K^+. \quad (22)$$

# Linearization methods

Josephy-Newton algorithm for inclusions (the problem to solve II)

## Examples (continued)

- 4 The **Peano-Kantorovitch NC1** of problem  $\min\{f(x) : x \in X\}$  reads

$$\nabla f(x) + N_X(x) \ni 0.$$

- 5 The **first order optimality conditions** for  $(P_A)$  with  $K \curvearrowright K_A$  can be written like  $(P_{NCP})$  with the variable  $x \curvearrowright (x, \lambda) \in \mathbb{E} \times \mathbb{F}$ ,

$$F(x) \curvearrowright F(x, \lambda) = \begin{pmatrix} \nabla f(x) + c'(x)^* \lambda \\ -c(x) \end{pmatrix}, \quad \text{and} \quad K = \mathbb{E} \times K_A^-. \quad (23)$$

- 6 If  $N$  is the constant multifunction  $x \mapsto \{0_{\mathbb{R}^E}\} \times \mathbb{R}_+^{m_I} \subset \mathbb{R}^m \equiv \mathbb{F}$  where  $E$  and  $I$  make a partition of  $[1 : m]$ ,  $(P_{IN})$  becomes the **system of equalities and inequalities**

$$F_E(x) = 0 \quad \text{and} \quad F_I(x) \leq 0.$$



72 / 109

# Linearization methods

Josephy-Newton algorithm for inclusions (the JN algorithm) [14, 15]

## Algorithm (Josephy-Newton algorithm for solving $(P_{IN})$ )

Given  $x_k$ , compute  $x_{k+1}$  as a solution to the problem in  $x$ :

$$F(x_k) + M_k(x - x_k) + N(x) \ni 0, \quad (24)$$

where  $M_k = F'(x_k)$  or an approximation to it.

## Remarks

- Only  $F$  is linearized, not  $N$  (is the reason for the chosen structure of  $(P_{IN})$ ).
- (24) captures more information from  $(P_{IN})$  than a “simple” linearization.
- (24) is often a **nonlinear** problem, hence yielding an expensive iteration.
- Makes sense computationally if  $N$  is sufficiently simple.



73 / 109

# Linearization methods

Josephy-Newton algorithm for inclusions (semistability) [3]

A solution  $x_*$  to  $(P_{IN})$  is said **semistable** if  $\exists \sigma_1 > 0$  and  $\exists \sigma_2 > 0$  such that for any  $(x, p)$  satisfying

$$F(x) + N(x) \ni p \quad \text{and} \quad \|x - x_*\| \leq \sigma_1$$

there holds

$$\|x - x_*\| \leq \sigma_2 \|p\|.$$

## Remarks

- The perturbed inclusion  $F(x) + N(x) \ni p$  is not required to have a solution.
- The property is demanding only for small perturbations  $p$  (i.e.,  $\|p\| < \sigma_1/\sigma_2$ ).
- Semistability implies that  $x_*$  is the unique solution to  $(P_{IN})$  on  $\bar{B}(x_*, \sigma_1)$ .
- If  $N \equiv \{0\}$ , then semistability of  $x_* \Leftrightarrow$  nonsingularity of  $F'(x_*)$ .



74 / 109

# Linearization methods

Josephy-Newton algorithm for inclusions (speed of convergence) [3]

## Proposition (speed of convergence of quasi-Newton methods)

Suppose that  $F$  is  $C^1$  near a **semistable** solution  $x_*$  to  $(P_{IN})$ . Let  $\{x_k\}$  be a sequence generated by algorithm (24), converging to  $x_*$ .

- 1 If  $(M_k - F'(x_*))(x_{k+1} - x_k) = o(\|x_{k+1} - x_k\|)$ , then  $\{x_k\}$  converges superlinearly.
- 2 If  $(M_k - F'(x_*))(x_{k+1} - x_k) = O(\|x_{k+1} - x_k\|^2)$  and  $F$  is  $C^{1,1}$  near  $x_*$ , then  $\{x_k\}$  converges quadratically.

## Corollary (speed of convergence of Newton's method)

Suppose that  $F$  is  $C^1$  near a **semistable** solution  $x_*$  to  $(P_{IN})$ . Let  $\{x_k\}$  be a sequence generated by algorithm (24) with  $M_k = F'(x_k)$ , converging to  $x_*$ . Then

- 1  $\{x_k\}$  converges superlinearly,
- 2 if, furthermore  $F$  is  $C^{1,1}$  near  $x_*$ , then  $\{x_k\}$  converges quadratically.

75 / 109

## Linearization methods

Josephy-Newton algorithm for inclusions (semistability for polyhedral VI) [3]

### Proposition (semistability characterization for polyhedral VI)

Consider problem  $(P_{VI})$  with a closed convex *polyhedron*  $C$  and  $F$  being  $C^1$  near a solution  $x_*$ . Then the following properties are equivalent:

- 1  $x_*$  is semistable,
- 2  $x_*$  is an isolated solution to

$$F(x_*) + F'(x_*)(x - x_*) + N_C(x) \ni 0,$$

- 3 there holds  $\langle F'(x_*)(x - x_*), x - x_* \rangle > 0$  when  $x \in C \setminus \{x_*\}$  satisfies

$$\langle F(x_*), x - x_* \rangle = 0 \quad \text{and} \quad F(x_*) + F'(x_*)(x - x_*) + N_C(x_*) \ni 0,$$

- 4  $x_*$  is the unique solution to

$$N_C(x) \subset N_C(x_*), \quad \langle F(x_*), x - x_* \rangle = 0, \quad \mathbb{R}_+ F(x_*) + F'(x_*)(x - x_*) + N_C(x) \ni 0.$$

76 / 109

## Linearization methods

Josephy-Newton algorithm for inclusions (local convergence) [3]

A solution  $x_*$  to  $(P_{IN})$  is said *hemistable* if  $\forall \alpha > 0, \exists \beta > 0, \forall x_0 \in \bar{B}(x_*, \beta)$ , the “linearized” inclusion in  $x$

$$F(x_0) + F'(x_0)(x - x_0) + N(x) \ni 0$$

has a solution in  $\bar{B}(x_*, \alpha)$ .

### Theorem (local convergence of JN)

Suppose that  $F$  is  $C^1$  near a *semistable* and *hemistable* solution  $x_*$  to  $(P_{IN})$ . Then  $\exists \varepsilon > 0$  such that if  $x_1 \in \bar{B}(x_*, \varepsilon)$ , then

- 1 the JN algorithm (24) with  $M_k = F'(x_k)$  can generate  $\{x_k\} \subset \bar{B}(x_*, \varepsilon)$ ,
- 2 any sequence  $\{x_k\}$  generated in  $\bar{B}(x_*, \varepsilon)$  by the JN algorithm converges superlinearly to  $x_*$  (quadratically if  $F$  is  $C^{1,1}$ ).

77 / 109

# Linearization methods

## The SQP algorithm (overview)

Recall the equality and inequality constrained problem

$$(P_{EI}) \quad \begin{cases} \inf_x f(x) \\ c_E(x) = 0 \\ c_I(x) \leq 0. \end{cases}$$

### Three popular methods to solve $(P_{EI})$

- **Augmented Lagrangian methods:** a dual method that generates

$\{(\lambda_k, r_k)\} \subset \mathbb{R}^m \times \mathbb{R}_{++}$  by

- $\inf_{x, s \geq 0} \left( f(x) + (\lambda_k)_E^T c_E(x) + \frac{r_k}{2} \|c_E(x)\|_2^2 + (\lambda_k)_I^T (c_I(x) + s) + \frac{r_k}{2} \|c_I(x) + s\|_2^2 \right),$
- $\lambda_{k+1} := (\lambda_k + r_k c(x_k))^\#$  and  $r_{k+1} = ?$  (heuristics for nonlinear problems).

- **SQP methods:** it is a linearization method on the KKT system (see below).

- **Interior point methods,** which can be viewed as a penalization method solving (approximately) a sequence of problems (25) below with  $\mu \downarrow 0$ , thanks to the SQP algorithm:

$$\begin{cases} \min_{(x,s)} f(x) - \mu \sum_{i \in I} \log s_i \\ c_E(x) = 0, \quad c_I(x) + s = 0. \end{cases} \quad (25)$$

*Inria*

79 / 109

# Linearization methods

## The SQP algorithm (definition - KKT is a nonlinear complementarity problem)

- Similarly to Newton's method in unconstrained optimization, the SQP algorithm is conceptually interested in solutions of the first order optimality (KKT) system in  $(x, \lambda)$  of  $(P_{EI})$ :

$$\nabla_x \ell(x, \lambda) = 0, \quad (26a)$$

$$c_E(x) = 0, \quad (26b)$$

$$0 \leq \lambda_I \perp c_I(x) \leq 0. \quad (26c)$$

- This system in  $(x, \lambda)$  can be written like  $(P_{IN})$  or  $(P_{NCP})$ , namely

$$F(x, \lambda) + N_K(x, \lambda) \ni 0 \quad \text{or} \quad K \ni (x, \lambda) \perp F(x, \lambda) \in K^+, \quad (27)$$

with the data

$$F(x, \lambda) = \begin{pmatrix} \nabla_x \ell(x, \lambda) \\ -c(x) \end{pmatrix} \quad \text{and} \quad K = \mathbb{E} \times (\mathbb{R}^{m_E} \times \mathbb{R}_+^{m_I}). \quad (28)$$

$$\Rightarrow N_K(x, \lambda) = \{0_{\mathbb{E}}\} \times (\{0_{\mathbb{R}^{m_E}}\} \times N_{\mathbb{R}_+^{m_I}}(\lambda_I)), \quad K^+ = \{0_{\mathbb{E}}\} \times (\{0_{\mathbb{R}^{m_E}}\} \times \mathbb{R}_+^{m_I}).$$

*Inria*

80 / 109

## Linearization methods

The SQP algorithm (definition - the SQP algorithm viewed as a JN method) [22]

- The **SQP algorithm** (SQP for *Sequential Quadratic Programming*) for solving  $(P_{EI})$  is the JN algorithm (24) on the inclusion (27)-(28):

$$\nabla_x \ell(x_k, \lambda_k) + M_k(x - x_k) + c'(x_k)^*(\lambda - \lambda_k) = 0, \quad (29a)$$

$$c_E(x_k) + c'_E(x_k) \cdot (x - x_k) = 0, \quad (29b)$$

$$0 \leq \lambda_I \perp (c_I(x_k) + c'_I(x_k) \cdot (x - x_k)) \leq 0, \quad (29c)$$

where  $M_k = L_k := \nabla_{xx}^2 \ell(x_k, \lambda_k)$  or an approximation to it ( $M_k \neq \nabla^2 f(x_k)$ !).

- (29) is formed of the KKT conditions of the **osculating quadratic problem**

$$(OQP) \quad \begin{cases} \min_x \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle M_k(x - x_k), x - x_k \rangle \\ c_E(x_k) + c'_E(x_k) \cdot (x - x_k) = 0, \\ c_I(x_k) + c'_I(x_k) \cdot (x - x_k) \leq 0. \end{cases} \quad (30)$$

The primal-dual solution  $(x_{k+1}, \lambda_{k+1})$  to the OQP is the new iterate.



81 / 109

## Linearization methods

The SQP algorithm (definition - the local SQP algorithm) [11, 12]

### Algorithm (local SQP)

From  $(x_k, \lambda_k)$  to  $(x_{k+1}, \lambda_{k+1})$ :

- 1 If (26) holds at  $(x, \lambda) = (x_k, \lambda_k)$ , stop.
- 2 Compute a primal-dual stationary point  $(x_{k+1}, \lambda_{k+1})$  of the OQP (30).

### Remarks

- The OQP's are still hard to solve (not just a linear system, expensive iteration):
  - ▶ If  $M_k = L_k \not\geq 0$ , the OQP is NP-hard.
  - ▶ One of the good reasons for taking  $M_k \simeq L_k$  with  $M_k \succ 0$ , updated by a quasi-Newton method; the OQP is then polynomial, but still difficult.
- Other (non local) difficulties to overcome:
  - ▶ What if the linearized constraints are incompatible?
  - ▶ What if the OQP is unbounded?
- Nothing is done for forcing the convergence from remote starting points.



82 / 109

## Linearization methods

The SQP algorithm (local convergence - semistability and hemistability of a KKT pair) [3]

### Proposition (semistability and SC2)

If  $(x_*, \lambda_*)$  satisfies KKT, then the following properties are equivalent:

- 1  $x_*$  is a local minimizer and  $(x_*, \lambda_*)$  is semistable,
- 2  $\Lambda_* = \{\lambda_*\}$  and  $x_*$  satisfies SC2.

At a local solution, semistability implies hemistability:

### Proposition (SC for hemistability)

If

- $x_*$  is a local minimizer of  $(P_{EI})$ ,
- $(x_*, \lambda_*)$  satisfies the KKT conditions,
- $(x_*, \lambda_*)$  is semistable,

then  $(x_*, \lambda_*)$  is hemistable.



83 / 109

## Linearization methods

The SQP algorithm (local convergence - the result) [3]

### Theorem (local convergence of SQP)

If

- $f$  and  $c$  are  $C^{2,1}$  near a local minimizer  $x_*$  of  $(P_{EI})$ ,
- there is a unique multiplier  $\lambda_*$  associated with  $x_*$ ,
- the SC2 are satisfied,

then there exists a neighborhood  $V$  of  $(x_*, \lambda_*)$  such that if the first iterate  $(x_1, \lambda_1) \in V$ , then

- 1 the SQP algorithm **can** generate  $\{(x_k, \lambda_k)\}$  in  $V$ ,
- 2 **any** sequence  $\{(x_k, \lambda_k)\}$  generated **in  $V$**  by the SQP algorithm converges quadratically to  $(x_*, \lambda_*)$ .



84 / 109

# Linearization methods

The SQP algorithm (exact penalization)

- Consider the nonsmooth penalty function  $\Theta_\sigma : \mathbb{E} \rightarrow \mathbb{R}$  associated with  $(P_{EI})$ :

$$\Theta_\sigma(x) := f(x) + \|c(x)^\#\|,$$

where for  $v \in \mathbb{R}^m$ ,  $v^\# \in \mathbb{R}^m$  is defined by:  $(v^\#)_i = v_i$  when  $i \in E$  and  $(v^\#)_i = v_i^+ = \max(0, v_i)$  when  $i \in I$ .

- The **dual norm** of  $\|\cdot\|$  is defined by

$$\|u\|_D := \sup_{\|v\| \leq 1} u^\top v.$$

## Proposition (exact penalty property)

If •  $f$  and  $c$  are  $C^2$  near a local minimizer  $x_*$  of  $(P_{EI})$ ,

- the set  $\Lambda_*$  of associated multipliers is nonempty,
- weak SC2 holds,
- $\sigma > \sup\{\|\lambda_*\|_D : \lambda_* \in \Lambda_*\}$ ,

then  $x_*$  is a strict local minimum of  $\Theta_\sigma$ .

85 / 109

# Linearization methods

The SQP algorithm (globalization - descent property of the convex OQP solution)

Recall the **osculating quadratic problem** at  $(x_k, \lambda_k)$ :  $M_k \simeq L(x_k, \lambda_k)$  and

$$(OQP)_k \quad \begin{cases} \min_d \langle \nabla f(x_k), d \rangle + \frac{1}{2} \langle M_k d, d \rangle \\ (c(x_k) + c'(x_k)d)^\# = 0. \end{cases} \quad (31)$$

We often make the following assumption (true for  $\ell_p$  norms)

$$\|\cdot^\#\| : v \in \mathbb{R}^m \mapsto \|v^\#\| \text{ is convex.}$$

## Proposition (descent direction)

If •  $(d_k, \lambda_k^{\text{QP}})$  is a stationary pair of (31),

- $\|\cdot^\#\|$  is convex,

then

- 1  $\Theta'_\sigma(x_k; d_k) \leq -\langle M_k d_k, d_k \rangle + (\lambda_k^{\text{QP}})^\top c(x_k) - \sigma \|c(x_k)^\#\|,$
- 2  $\Theta'_\sigma(x_k; d_k) < 0$ , if  $\sigma \geq \|\lambda_k^{\text{QP}}\|_D$ ,  $M_k \succ 0$ , and  $x_k$  is not stationary.

86 / 109

## Linearization methods

### The SQP algorithm (globalization I)

The SQP algorithm with linesearch forces  $M_k \succ 0$  and minimizes the **changing nondifferentiable** merit function  $\Theta_{\sigma_k}$  along the SQP directions  $d_k$ .

#### Algorithm (global SQP)

Given  $(x_k, \lambda_k, M_k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{S}_{+++}^n$ , compute  $(x_{k+1}, \lambda_{k+1}, M_{k+1})$  by

- solve (31) to get a PD solution  $(d_k, \lambda_k^{\text{QP}})$  (if any),
- impose  $\sigma_k \geq \|\lambda_k^{\text{QP}}\|_D + \bar{\sigma}$  keeping  $\sigma_k$  constant if  $\{\lambda_k^{\text{QP}}\}$  is bounded,
- linesearch:  $\alpha_k > 0$  such that  $\Theta_{\sigma_k}(x_k + \alpha_k d_k) \leq \Theta_{\sigma_k}(x_k) + \omega \alpha_k \Delta_k$ ,
- $x_{k+1} := x_k + \alpha_k d_k$  and  $\lambda_{k+1} := \lambda_k + \alpha_k (\lambda_k^{\text{QP}} - \lambda_k)$ ,
- update  $M_k \curvearrowright M_{k+1}$ .

We have used  $\Delta_k := \langle \nabla f(x_k), d_k \rangle - \sigma_k \|c(x_k)^\# \|$  as a negative over-estimate of  $\Theta'_{\sigma_k}(x_k; d_k)$ .

## Linearization methods

### The SQP algorithm (globalization II)

#### Theorem (global convergence of SQP)

If •  $f$  and  $c$  are  $C^{1,1}$ ,

- $\|\cdot\|^\#$  is convex,
- $\{M_k\}$  and  $\{M_k^{-1}\}$  are  $\succ 0$  and bounded,
- (31) has a PD solution  $(d_k, \lambda_k^{\text{QP}})$  for all  $k \geq 1$ ,
- $\{\lambda_k^{\text{QP}}\}$  is bounded,
- $\Theta_{\sigma_k}(x_k)$  is bounded below,

then the KKT conditions are satisfied asymptotically, meaning that  $\nabla_x \ell(x_k, \lambda_k^{\text{QP}}) \rightarrow 0$ ,  $c(x_k)^\# \rightarrow 0$ ,  $(\lambda_k^{\text{QP}})_I \geq 0$ , and  $(\lambda_k^{\text{QP}})_I^\top c_I(x_k) \rightarrow 0$ .

# Linearization methods

## The semismooth Newton method (motivation)

- Let  $\mathbb{E}$  and  $\mathbb{F}$  be (finite dimensional) normed spaces and  $\Omega \subset \mathbb{E}$  be **open**.
- We consider the problem of finding a zero of a **nonsmooth** function  $F : \Omega \rightarrow \mathbb{F}$ :

$$F(x) = 0. \quad (32)$$

### • Examples

- 1 The NCP ( $0 \leq x \perp \Phi(x) \geq 0$ ) can be represented by (32):

$$\Psi(x, \Phi(x)) = 0,$$

where  $\Psi(u, v) = \{\psi(u_i, v_i)\}_i$  and  $\psi$  is a **C-function**, meaning that  $\psi(a, b) = 0$  iff  $a \geq 0$ ,  $b \geq 0$ , and  $ab = 0$ . Examples (FB = Fischer-Burmeister):

$$\psi_{\min}(a, b) = \min(a, b) \quad \text{and} \quad \psi_{FB}(a, b) = \sqrt{a^2 + b^2} - (a + b).$$

- 2 The VI problem (find  $x \in C$  s.t.  $\langle \Phi(x), y - x \rangle \geq 0$  for all  $y \in C$ ) can be written like (32):

$$P_C(x - \Phi(x)) - x = 0.$$



# Linearization methods

## The semismooth Newton method (motivation)

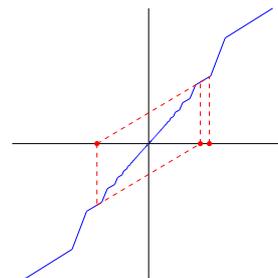
- Suppose that  $F : \mathbb{E} \rightarrow \mathbb{F}$  is **Lipschitz continuous** near  $x_* \in \Omega$ .
- Then  $\partial_C F(x) \neq \emptyset$  for  $x$  near  $x_*$  and one could think of the Newton-like algorithm generating  $\{x_k\}$  by

$$x_{k+1} = x_k - J_k^{-1} F(x_k),$$

provided some nonsingular  $J_k$  can be found in  $\partial_C F(x_k)$ .

- Actually, this algorithm may not converge **locally** as in the example below [17]:

$$\begin{aligned} F : \mathbb{R} \rightarrow \mathbb{R} \text{ is Lipschitz,} \\ F(0) = 0, \\ \partial_C F(0) = [\tfrac{1}{2}, 2] \not\ni 0, \\ \lim_{k \rightarrow \infty} \frac{F(x_k) - F(0) - F'(x_k)(x_k - 0)}{|x_k - 0|} \geq 1. \end{aligned}$$



- **Semismoothness** is an assumption on  $F$  that prevents such a cycling (arbitrary close to the zero).



## Linearization methods

The semismooth Newton method (semismoothness - definition) [19, 20]

Let  $\mathbb{E}, \mathbb{F}, \dots$  be normed spaces of finite dimension and  $\Omega$  be open in  $\mathbb{E}$ .

- $F : \Omega \rightarrow \mathbb{F}$  is **semismooth** at  $x \in \Omega$  if

(S<sub>1</sub>)  $F$  is Lipschitz near  $x$ ,

(S<sub>2</sub>)  $F'(x; h)$  exists for all  $h \in \mathbb{E}$ ,

(S<sub>3</sub>) for  $h \rightarrow 0$  in  $\mathbb{E}$ , there holds

$$\sup_{J \in \partial_C F(x+h)} \frac{\|F(x+h) - F(x) - Jh\|}{\|h\|} \rightarrow 0.$$

- $F : \Omega \rightarrow \mathbb{F}$  is **strongly semismooth** at  $x \in \Omega$  if it is semismooth at  $x$  with (S<sub>3</sub>) strengthened by

(S<sub>4</sub>) for  $h$  small in  $\mathbb{E}$ , there holds

$$\sup_{J \in \partial_C F(x+h)} \|F(x+h) - F(x) - Jh\| = O(\|h\|^2).$$



92 / 109

## Linearization methods

The semismooth Newton method (semismoothness - properties I) [8, 13]

### Proposition (differentiability and semismoothness)

If  $F : \Omega \rightarrow \mathbb{F}$  is  $C^1$  ( $C^{1,1}$ ) near  $x \in \Omega$ , then  $F$  is (strongly) semismooth at  $x$ .

### Proposition (a convex function is semismooth)

If  $f : \Omega \rightarrow \mathbb{R}$  is convex in a convex neighborhood of  $x \in \Omega$ , then  $f$  is semismooth at  $x$ .

$F : \Omega \rightarrow \mathbb{F}$  is said to be **piecewise semismooth** at  $x \in \Omega$  if there exist a neighborhood  $V$  of  $x$  and functions  $F_i : V \rightarrow \mathbb{F}$ , with  $i \in I$  ( $I$  finite), which are semismooth at  $x$ , such that for all  $y \in V$ ,  $F(y) = F_i(y)$  for some  $i \in I$ .

### Proposition (piecewise semismoothness)

If  $F : \Omega \rightarrow \mathbb{F}$  is piecewise semismooth at  $x \in \Omega$ , then  $F$  is semismooth at  $x$ .



93 / 109

## Linearization methods

The semismooth Newton method (semismoothness - properties II) [8, 13]

### Proposition (components)

Let  $F_1 : \Omega \rightarrow \mathbb{F}_1$ ,  $F_2 : \Omega \rightarrow \mathbb{F}_2$ , and  $x \in \Omega$ . Then  $(F_1, F_2) : \Omega \rightarrow \mathbb{F}_1 \times \mathbb{F}_2$  is (strongly) semismooth at  $x \in \Omega$  if and only if  $F_1$  and  $F_2$  are (strongly) semismooth at  $x$ .

### Proposition (composition)

If  $F : \Omega \rightarrow \mathbb{F}$  is (strongly) semismooth at  $x \in \Omega$ ,  $V$  is a neighborhood of  $F(x)$ , and  $G : V \rightarrow \mathbb{G}$  is (strongly) semismooth at  $F(x)$ , then  $G \circ F$  is (strongly) semismooth at  $x$ .

### Proposition (calculus)

If  $F_1 : \Omega \rightarrow \mathbb{F}$  and  $F_2 : \Omega \rightarrow \mathbb{F}$  are (strongly) semismooth at  $x \in \Omega$ , then the following functions are (strongly) semismooth at  $x$  (for the last two,  $\mathbb{F} = \mathbb{R}^m$ ):

$$F_1 + F_2, \quad \langle F_1, F_2 \rangle, \quad \max(F_1, F_2), \quad \text{and} \quad \min(F_1, F_2).$$

94 / 109

## Linearization methods

The semismooth Newton method (semismoothness - examples) [8, 13]

### Examples

- The  $\ell_p$  norm, for  $1 \leq p \leq \infty$ , is *strongly* semismooth.
- The **min C-function**  $\psi_{\min} : (a, b) \in \mathbb{R}^2 \mapsto \min(a, b)$  is *strongly* semismooth.
- The **Fischer-Burmeister C-function**  $\psi_{FB} : (a, b) \in \mathbb{R}^2 \mapsto \sqrt{a^2 + b^2} - (a + b)$  is *strongly* semismooth.
- The **projector**  $P_K$  on the convex set  $K := \{x \in \mathbb{E} : c(x) \leq 0\}$  is semismooth at  $x$ , provided
  - ▶  $c : \mathbb{E} \rightarrow \mathbb{R}^m$  is  $C^2$  and componentwise convex,
  - ▶ the **constant rank constraint qualification** (CQ-CR) holds at  $P_K(x)$  [(CQ-LI) is certainly fine].

# Linearization methods

## The semismooth Newton method (the algorithm)

### Algorithm (semismooth Newton for equations)

Given  $x_k \in \mathbb{E}$ , compute  $x_{k+1} \in \mathbb{E}$  as a solution to

$$F(x_k) + J_k(x - x_k) = 0,$$

for some nonsingular  $J_k \in \partial_C F(x_k)$  (if any).

### Remarks

- To work well the algorithm needs **smoothness** and **regularity** assumptions.
- There is a **single** linear system to solve per iteration (i.e., cheap iteration).



# Linearization methods

## The semismooth Newton method (local convergence)

### Theorem (local convergence of the semismooth Newton method)

If •  $F(x_*) = 0$ ,

- $F$  is semismooth at  $x_*$ ,
- $\partial_C F(x_*)$  is nonsingular,

then

- 1 there is a neighborhood  $V$  of  $x_*$  such that the semismooth Newton algorithm starting at  $x_1 \in V$  is well defined and generates a sequence in  $V$ , converging to  $x_*$  superlinearly,
- 2 the convergence is quadratic if  $F$  is strongly semismooth.



# Semidefinite optimization

Problem definition (cones  $\mathcal{S}_+^n$  and  $\mathcal{S}_{++}^n$ )

Notation and first properties

- $\mathcal{S}^n$  is the Euclidean space of symmetric  $n \times n$  real matrices, equipped with the **scalar product**

$$\langle \cdot, \cdot \rangle : (A, B) \in (\mathcal{S}^n)^2 \mapsto \langle A, B \rangle = \text{tr}(AB) = \sum_{ij} A_{ij}B_{ij} \in \mathbb{R}.$$

- $\mathcal{S}_+^n$  is the cone of  $\mathcal{S}^n$  made of the **positive semidefinite matrices**:

$$\begin{aligned} A \succcurlyeq 0 &\iff A \in \mathcal{S}_+^n \iff \lambda(A) \subset [0, +\infty[, \\ A \succcurlyeq 0 &\iff \forall B \succcurlyeq 0 : \langle A, B \rangle \geq 0, \end{aligned} \tag{33}$$

if  $A \succcurlyeq 0$  and  $B \succcurlyeq 0$ , then  $\langle A, B \rangle = 0 \iff AB = 0$ ,

$$\text{T}_A \mathcal{S}_+^n = \{D \in \mathcal{S}^n : v^T D v \geq 0, \text{ for all } v \in \mathcal{N}(A)\}.$$

By (33),  $\mathcal{S}_+^n$  is **self-dual**, meaning that  $(\mathcal{S}_+^n)^+ = \mathcal{S}_+^n$ .

- $\mathcal{S}_{++}^n$  is the cone of  $\mathcal{S}^n$  made of the **positive definite matrices**:

$$\begin{aligned} A \succ 0 &\iff A \in \mathcal{S}_{++}^n \iff \lambda(A) \subset ]0, +\infty[, \\ A \succ 0 &\iff \forall B \in \mathcal{S}_+^n \setminus \{0\} : \langle A, B \rangle > 0, \end{aligned}$$

$$A \succcurlyeq 0 \text{ and } [v^T M v > 0, \forall v \in \mathcal{N}(A) \setminus \{0\}] \implies M + rA \succ 0 \text{ for large } r$$

# Semidefinite optimization

Problem definition (primal and dual problems)

The primal and (Lagrange) dual of the SDO problem read

$$(P) \begin{cases} \inf \langle C, X \rangle \\ A(X) = b \\ X \succcurlyeq 0 \end{cases} \quad \text{and} \quad (D) \begin{cases} \sup \langle b, y \rangle \\ A^*(y) + S = C \\ S \succcurlyeq 0, \end{cases} \tag{34}$$

where

- $C \in \mathcal{S}^n$  and  $b \in \mathbb{R}^m$ ,
- $A : \mathcal{S}^n \rightarrow \mathbb{R}^m$  is linear ( $A^* : \mathbb{R}^m \rightarrow \mathcal{S}^n$  is its adjoint).

**Notation**

- Feasible sets:  $\mathcal{F}_p := \{X \in \mathcal{S}_+^n : A(X) = b\}$ ,  
 $\mathcal{F}_d := \{(y, S) \in \mathbb{R}^m \times \mathcal{S}_+^n : A^*(y) + S = C\}$ , and  $\mathcal{F} := \mathcal{F}_d \times \mathcal{F}_p$ .
- Strictly feasible sets:  $\mathcal{F}_p^s := \{X \in \mathcal{S}_{++}^n : A(X) = b\}$ ,  
 $\mathcal{F}_d^s := \{(y, S) \in \mathbb{R}^m \times \mathcal{S}_{++}^n : A^*(y) + S = C\}$ , and  $\mathcal{F}^s := \mathcal{F}_d^s \times \mathcal{F}_p^s$ .
- Optimal values:  $\text{val}(P)$  and  $\text{val}(D)$ . Duality gap:  $\text{val}(P) - \text{val}(D)$
- Solution sets:  $\text{Sol}(P)$  and  $\text{Sol}(D)$ .

# Semidefinite optimization

Problem definition (Lagrange dualization consequences)

## Proposition (consequences of the Lagrangian dualization)

- 1  $\text{val}(D) \leq \text{val}(P)$ .
- 2  $(X, y, S) \in \mathcal{F} \implies \langle C, X \rangle - \langle b, y \rangle = \langle X, S \rangle \geq 0$ .
- 3  $(X, y, S) \in \mathcal{F} \text{ et } \langle X, S \rangle = 0 \implies X \in \text{Sol}(P) \text{ et } (y, S) \in \text{Sol}(D)$ .

## Remark

The reverse implication in 3 is not necessarily true (not like in LO).

# Semidefinite optimization

Problem definition (examples of SDO formulations)

The **Schur complement** of  $A \succ 0$  in

$$K := \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

is  $A^s := C - B^T A^{-1} B$ . When  $A \succ 0$ , there hold

$$\begin{aligned} K \succeq 0 &\iff A^s \succeq 0, \\ K \succ 0 &\iff A^s \succ 0. \end{aligned}$$

## Examples of SDO modelling

- 1 Linear optimization.
- 2 Convex quadratic optimization.
- 3 Global minimization of polynomials (relaxation if  $n \geq 2$ ).
- 4 Shor relaxation of a QCQP.
- 5 Many more ...

# Semidefinite optimization

## Existence of solution

- Strong duality of linear optimization no longer holds (since the linear image of a closed convex cone is not necessarily closed).
- Here are conditions for having nonempty **compact** sets of solutions.

### Proposition (compact sets of solutions)

- 1  $\mathcal{F}_P \times \mathcal{F}_D^s \neq \emptyset \implies \text{Sol}(P) \neq \emptyset$  and compact.
- 2  $\mathcal{F}_P^s \times \mathcal{F}_D \neq \emptyset, A$  surjective  $\implies \text{Sol}(D) \neq \emptyset$  and compact.
- 3  $\mathcal{F}^s \neq \emptyset, A$  surjective  $\implies \text{Sol}(P) \times \text{Sol}(D) \neq \emptyset$  and compact.

In these cases, there is no duality gap:  $\text{val}(D) = \text{val}(P)$ .

- The sufficient condition in 1 is (CQ-R) for  $(D)$ .  
The sufficient conditions in 2 are (CQ-R) for  $(P)$ .



# Semidefinite optimization

## Optimality conditions

### Proposition (optimality conditions)

Suppose that  $\mathcal{F}^s \neq \emptyset$  and  $A$  is surjective, then  $(X, y, S)$  is a primal dual solution if and only if

$$\begin{cases} A^*(y) + S = C, & S \succcurlyeq 0, \\ A(X) = b, & X \succcurlyeq 0, \\ \langle X, S \rangle = 0. \end{cases} \quad (35)$$



# Semidefinite optimization

## An interior point algorithm (central path I)

- There are good reasons to generate iterates well inside  $\mathcal{F}_p^s$ . This is obtained analytically (not geometrically) by an **interior penalization**:

$$\begin{aligned} (P) & \quad \rightsquigarrow \quad (P_\mu) & \begin{cases} \inf_X \langle C, X \rangle + \mu \text{ld}(X) \\ A(X) = b, \end{cases} \\ (D) & \quad \rightsquigarrow \quad (D_\mu) & \begin{cases} \sup_{(y,S)} \langle b, y \rangle - \mu \text{ld}(S) \\ A^*(y) + S = C, \end{cases} \end{aligned}$$

where  $\text{ld} : \mathcal{S}^n \rightarrow \bar{\mathbb{R}}$  is the **strictly convex** and **closed** function defined at  $X$  by

$$\text{ld}(X) := \begin{cases} -\log \det(X) & \text{if } X \succ 0 \\ +\infty & \text{otherwise.} \end{cases}$$

- Three properties of  $\text{ld}$  (with  $X \succ 0$  and  $H, K \in \mathcal{S}^n$ ):

$$\begin{aligned} \text{ld}'(X) \cdot H &= -\langle X^{-1}, H \rangle, \\ \text{ld}''(X) \cdot (H, K) &= \langle X^{-1} H X^{-1}, K \rangle, \\ \text{ld}^\infty &= \mathcal{I}_{\mathcal{S}_+^n}. \end{aligned}$$



# Semidefinite optimization

## An interior point algorithm (central path II)

The **central path** is the smooth curve  $\mathcal{C} : \mu \in \mathbb{R}_{++} \mapsto$  the unique solution to

$$(O_\mu) \quad \begin{cases} A^*(y) + S = C, & S \succ 0, \\ A(X) = b, & X \succ 0, \\ XS = \mu I. \end{cases} \quad (36)$$

### Proposition (existence and smoothness of the central path)

If  $\mathcal{F}_D^s \neq \emptyset$  (or  $\mathcal{F}_p^s \neq \emptyset$  and  $A$  surjective) and  $\mu > 0$ , then

- 1 the system  $(O_\mu)$  has a solution  $(X_\mu, y_\mu, S_\mu)$ , unique in  $(X_\mu, y_\mu)$ ,
- 2 if  $A$  surjective, then also uniqueness in  $y_\mu$  and  $\mu \mapsto (X_\mu, y_\mu, S_\mu)$  is  $C^\infty$ .



# Semidefinite optimization

An interior point algorithm (an algorithmic scheme)

A primal-dual path-following interior-point algorithm generates iterates  $z_k := (X_k, y_k, S_k) \in \mathcal{F}^s$  in a neighborhood  $V(\theta)$  of the central path  $\mathcal{C}$  ( $\theta \in ]0, 1[$  is a parameter that determines the size of the neighborhood). Each iteration proceeds along a Newton direction aiming a moving point on  $\mathcal{C}$ , whose central parameter is  $\sigma \bar{\mu}(z)$  where  $\sigma \in ]0, 1[$  and  $\bar{\mu}(z) := \langle X, S \rangle / n$ .

## Algorithm (primal-dual path-following IP)

From one iterate  $z$  to the next one  $z_+$ .

- 1 Let  $d$  be the Newton direction on a symmetrized version of  $(O_{\sigma \bar{\mu}(z)})$ .
- 2 Determine a large stepsize  $\alpha > 0$  such that  $z + \alpha d \in V(\theta)$ .
- 3  $z_+ := z + \alpha d$ .



109 / 109

 M.F. Anjos, J.B. Lasserre (2012).  
*Handbook on Semidefinite, Conic and Polynomial Optimization*.  
Springer.

 A. Ben-Tal, A. Nemirovski (2001).  
*Lectures on Modern Convex Optimization – Analysis, Algorithms, and Engineering Applications*.  
MPS-SIAM Series on Optimization 2. SIAM.

 J.F. Bonnans (1994).  
Local analysis of Newton-type methods for variational inequalities and nonlinear programming.  
*Applied Mathematics and Optimization*, 29, 161–186.  
[doi].

 J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal (2006).  
*Numerical Optimization – Theoretical and Practical Aspects* (second edition).  
Universitext. Springer Verlag, Berlin.  
[authors] [editor] [google books].

 J.F. Bonnans, A. Shapiro (2000).  
*Perturbation Analysis of Optimization Problems*.  
Springer Verlag, New York.

 F.H. Clarke (1983).  
*Optimization and Nonsmooth Analysis*.  
John Wiley & Sons, New York.

 R. Cominetti (1990).

Metric regularity, tangent sets, and second-order optimality conditions.  
*Applied Mathematics and Optimization*, 21(1), 265–287.  
[doi].

 F. Facchinei, J.-S. Pang (2003).  
*Finite-Dimensional Variational Inequalities and Complementarity Problems* (two volumes).  
Springer Series in Operations Research. Springer.

 J. Gauvin (1977).  
A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming.  
*Mathematical Programming*, 12, 136–138.

 J.Ch. Gilbert (2014).  
*Éléments d'Optimisation Différentiable – Théorie et Algorithmes*.  
Syllabus de cours à l'ENSTA, Paris.  
[internet].

 S.-P. Han (1976).  
Superlinearly convergent variable metric algorithms for general nonlinear programming problems.  
*Mathematical Programming*, 11, 263–282.

 S.-P. Han (1977).  
A globally convergent method for nonlinear programming.  
*Journal of Optimization Theory and Applications*, 22, 297–309.

 A.F. Izmailov, M.V. Solodov (2014).



109 / 109

*Newton-Type Methods for Optimization and Variational Problems.*  
Springer Series in Operations Research and Financial Engineering. Springer.



N.H. Josephy (1979).

Newton's method for generalized equations.  
Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, WI, USA.



N.H. Josephy (1979).

Quasi-Newton's method for generalized equations.  
Summary Report 1966, Mathematics Research Center, University of Wisconsin, Madison, WI, USA.



D. Klatte, B. Kummer (2002).

*Nonsmooth Equations in Optimization - Regularity, Calculus, Methods and Applications.*  
Kluwer Academic Publishers.  
[doi].



B. Kummer (1988).

Newton's method for nondifferentiable functions.  
In J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzsch, L. Tammer, M. Vlach, K. Zimmerman (editors), *Advances in Mathematical Optimization*, pages 114–125. Akademie-Verlag, Berlin.



L. Lyusternik (1934).

Conditional extrema of functionals.  
*Math. USSR-Sb.*, 41, 390–401.



L. Qi (1993).

Convergence analysis of some algorithms for solving nonsmooth equations.  
*Mathematics of Operations Research*, 18, 227–244.



L. Qi, J. Sun (1993).

A nonsmooth version of Newton's method.  
*Mathematical Programming*, 58, 353–367.  
[doi].



S.M. Robinson (1976).

Stability theory for systems of inequalities, part II: differentiable nonlinear systems.  
*SIAM Journal on Numerical Analysis*, 13, 497–513.



S.M. Robinson (1982).

Generalized equations and their solutions, part II: applications to nonlinear programming.  
*Mathematical Programming Study*, 19, 200–221.



R.T. Rockafellar, R. Wets (1998).

*Variational Analysis.*  
Grundlehren der mathematischen Wissenschaften 317. Springer.



M. Ulbrich (2011).

*Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces.*  
MPS-SIAM Series on Optimization 11. SIAM Publications, Philadelphia, PA, USA.  
[doi].



Wikipédia.fr.



Algorithme de Josephy-Newton.



H. Wolkowicz, R. Saigal, L. Vandenberghe (editors) (2000).  
*Handbook of Semidefinite Programming – Theory, Algorithms, and Applications.*



Kluwer Academic Publishers.

J. Zowe, S. Kurcyusz (1979).

Regularity and stability for the mathematical programming problem in Banach spaces.  
*Applied Mathematics and Optimization*, 5(1), 49–62.

