Institut Polytechnique de Paris
Master Program in Optimization

Advanced Continuous Optimization

J. Ch. Gilbert (INRIA Paris)

— May 26, 2021 —

Lectures: November 16, 2020 → January 11, 2020
Examination: January 18, 2020

Keep a close eye to
http://who.rocq.inria.fr/Jean-Charles.Gilbert/paris-saclay/optim.html

Outline

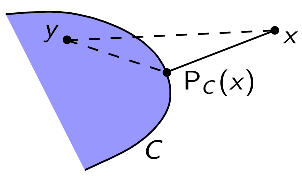
Background
Convex analysis (notation)

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

Background
Convex analysis (projection)

- *Definition*. For a *nonempty*, *closed*, *convex* set C in a *Euclidean* space $\mathbb{E}$, with scalar product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \|$, the problem

$$\inf_{y \in C} \|y - x\|$$



has a *unique* solution, called the projection of x on C and denoted $\mathrm{P}_C(x)$.

- *Characterization*. For $x \in \mathbb{E}$ and $\bar{x} \in C$, there hold

$$\begin{aligned} \bar{x} = \mathrm{P}_C(x) \quad &\iff \quad \langle y - \bar{x}, \bar{x} - x \rangle \geqslant 0, \quad \forall y \in C, \\ &\iff \quad \langle y - \bar{x}, y - x \rangle \geqslant 0, \quad \forall y \in C, \\ &\iff \quad \langle y - x, \bar{x} - x \rangle \geqslant \|\bar{x} - x\|^2, \quad \forall y \in C. \end{aligned}$$

Background
Convex analysis (relative interior I)

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

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

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

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

In finite dimension, the following holds

$$C \text{ convex and nonempty} \quad \Longrightarrow \quad \begin{cases} \operatorname{ri} C \neq \varnothing, \\ \operatorname{aff} C = \operatorname{aff}(\operatorname{ri} C). \end{cases}$$

Background
Convex analysis (relative interior II)

Proposition (relative interior criterion)

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

$$x \in \operatorname{ri} C \text{ and } y \in \overline{C} \quad \Longrightarrow \quad [x, y) \subseteq \operatorname{ri} C.$$

$$x \in \operatorname{ri} C \quad \Longleftrightarrow \quad \forall\, x_0 \in C \text{ (or aff } C),\ \exists\, t > 1 : \quad (1-t)x_0 + tx \in C.$$

Let C be a nonempty convex set. Then

- $\operatorname{ri} C$ is convex,
- $\overline{C}$ is convex and $\operatorname{aff} C = \operatorname{aff} \overline{C}$,
- $\overline{\operatorname{ri} C} = \overline{C}$ and $\operatorname{ri} \overline{C} = \operatorname{ri} C$ (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 $x + td \in C$.

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

Background

Convex analysis (dual cone and Farkas lemma)

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

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

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

Lemma (Farkas, generalized)

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

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

- $A^* : \mathbb{F} \to \mathbb{E}$ 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 $\mathcal{R}(A) = \mathcal{N}(A^*)^\perp$.

Background

Convex analysis (tangent and normal cones)

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

$$\mathrm{T}_x C \equiv \mathrm{T}_C(x) = \overline{\mathbb{R}_+(C - x)}.$$

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

$$\mathrm{N}_x C \equiv \mathrm{N}_C(x) = \{d \in \mathbb{E} : \langle x' - x, d \rangle \leqslant 0, \ \forall x' \in C\}.$$

  - There hold

$$\mathrm{N}_x C = (\mathrm{T}_x C)^- \qquad \text{and} \qquad \mathrm{T}_x C = (\mathrm{N}_x C)^-.$$

Background

Convex analysis (asymptotic cone I)

Let $\mathbb{E}$ be a vector space of finite dimension and C be a nonempty closed convex set of $\mathbb{E}$.

- The asymptotic cone of C is

$$C^\infty := \{d \in \mathbb{E} : C + \mathbb{R}_+ d \subseteq C\} = \{d \in \mathbb{E} : C + d \subseteq C\}.$$

- Properties
  - C^∞ is closed.
  - For any $x \in C$:

$$\begin{aligned} C^\infty &= \{d \in \mathbb{E} : x + \mathbb{R}_+ d \subseteq C\} = \bigcap_{t>0} \frac{C - x}{t} \\ &= \left\{ d \in \mathbb{E} : \exists \{x_k\} \subseteq C,\ \exists \{t_k\} \to \infty \text{ such that } \frac{x_k}{t_k} \to d \right\}. \end{aligned}$$

  - Boundedness by calculation:

$$C \text{ is bounded} \quad \iff \quad C^\infty = \{0\}.$$

Background

Convex analysis (asymptotic cone II)

- Two calculation rules (there are many more)
  - If $K \neq \varnothing$, then K is a closed convex cone $\iff K^\infty = K$.
  - For an arbitrary collection $\{C_i\}_{i \in I}$ of closed convex sets C_i with nonempty intersection:

$$(\cap_{i \in I} C_i)^\infty = \cap_{i \in I} C_i^\infty.$$

- Example
  - Let $A : \mathbb{E} \to \mathbb{F}$ and $B : \mathbb{E} \to \mathbb{G}$ be linear maps, $a \in \mathbb{F}$, $b \in \mathbb{G}$, K be a nonempty closed convex cone of $\mathbb{G}$, and

$$P := \{x \in \mathbb{E} : Ax = a,\ Bx \in b + K\} \neq \varnothing.$$

Then

$$P^\infty = \{d \in \mathbb{E} : Ad = 0,\ Bd \in K\}.$$

Background

Convex analysis (strict separation of convex sets)

The sets S_1 and S_2 in a Euclidean vector space $\mathbb{E}$ are said to be strictly separable if there exists a vector $\xi \in \mathbb{E}$ (necessarily nonzero) such that

$$\sup_{x_1 \in S_1} \langle \xi, x_1 \rangle < \inf_{x_2 \in S_2} \langle \xi, x_2 \rangle .$$

Proposition (strict separation of convex sets)

One can strictly separate two disjoint nonempty closed convex sets C_1 and $C_2 \subseteq \mathbb{E}$ in any of the following situations

1. $C_1 - C_2$ *is closed,*
2. $C_1^\infty \cap C_2^\infty = \{0\}$,
3. C_1 *or* C_2 *is compact,*
4. C_1 *and* C_2 *are polyhedral.*

Background

Convex analysis (convex polyhedron)

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

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

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

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

- $(\{x_k\} \to x) \implies I(x_k) \subseteq I(x)$ for large k.
- If $T : \mathbb{E} \to \mathbb{F}$ 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.
- $\mathrm{T}_x P = \mathrm{T}_x^f P = \{d \in \mathbb{E} : (Ad)_{I(x)} \leqslant 0\}$.
  $I(x_1) \subseteq I(x_2) \implies \mathrm{T}_{x_1} P \supseteq \mathrm{T}_{x_2} P$.
- $\mathrm{N}_x P = \mathrm{cone}\{A^* e_i : i \in I(x)\}$ (A^*: adjoint of A for the scalar product of $\mathbb{E}$).
  $I(x_1) \subseteq I(x_2) \implies \mathrm{N}_{x_1} P \subseteq \mathrm{N}_{x_2} P$.

Background

Convex analysis (asymptotic function I)

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

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

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

Proposition (asymptotic function f^∞)

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

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

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

3. $\operatorname{dom} f^\infty \subseteq (\operatorname{dom} f)^\infty$,
4. $f^\infty \in \overline{\operatorname{Conv}}(\mathbb{E})$.

Background

Convex analysis (asymptotic function II)

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

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

Proposition (existence of a bounded set of minimizers)

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

1. $\forall \nu \in \mathbb{R}$ *such that* $L_\nu(f) \neq \varnothing$, *the following holds*

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

2. *the following properties are equivalent:*
   - i. $\exists \nu \in \mathbb{R}$: $L_\nu(f)$ *is nonvoid and bounded,*
   - ii. $\forall \nu \in \mathbb{R}$: $L_\nu(f)$ *is bounded,*
   - iii. $\operatorname{Arg\,min} f$ *is nonvoid and bounded,*
   - iv. $\forall d \in \mathbb{E} \setminus \{0\}$: $f^\infty(d) > 0$.

Background

Convex analysis (subdifferential)

A subgradient at $x \in \mathbb{E}$ of $f \in \text{Conv}(\mathbb{E})$ is a vector $x^* \in \mathbb{E}$ such that

$$f(y) \geqslant f(x) + \langle x^*, y - x \rangle, \quad \forall\, y \in \mathbb{E}.$$

The subdifferential $\partial f(x)$ of f at x is the set of its subgradients at x. f is said to be subdifferentiable at x if $\partial f(x) \neq \varnothing$.

Proposition (characterization of subgradients)

For $f \in \overline{\text{Conv}}(\mathbb{E})$, $x \in \text{dom}\, f$, $x^* \in \mathbb{E}$, here are equivalent propreties

1. $x^* \in \partial f(x)$,
2. $f'(x; d) \geqslant \langle x^*, d \rangle, \quad \forall\, d \in \mathbb{E}$,
3. $x \in \text{Arg min}_{y \in \mathbb{E}} \big(f(y) - \langle x^*, y \rangle\big) = \text{Arg max}_{y \in \mathbb{E}} \big(\langle x^*, y \rangle - f(y)\big)$.

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:

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

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

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

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

- The image of a part $P \subseteq \mathbb{E}$ 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

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

- 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 semi-continuous at $x \in \mathbb{E}$ if $\forall\, \varepsilon > 0$, $\exists\, \delta > 0$, $\forall\, x' \in x + \delta B$, one has $T(x') \subseteq T(x) + \varepsilon B$ (in this definition, B may be the open or closed ball at any place).

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 $\forall\, t \in [0, 1]$:

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

Note that

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

Background

Nonsmooth analysis (Lipschitz continuity)

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

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

- 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 $\Omega \subseteq \mathbb{E}$ if it is Lipschitz near any point of Ω.

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

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

Définitions:

- solution or (global) minimum $x_* \in X$ if $\forall\, x \in X$, $f(x_*) \leqslant f(x)$,
- local minimum $x_* \in X$ if $\exists V \in \mathcal{N}(x_*)$, $\forall\, x \in X \cap V$, $f(x_*) \leqslant f(x)$,
- strict local/global minimum x_* if $f(x_*) < f(x)$ above when $x \neq x_*$.

Background

Optimization (tangent cone to a nonconvex set)

- A direction $d \in \mathbb{E}$ is tangent to $X \subseteq \mathbb{E}$ at $x \in X$ (in the sense of Bouligand) if

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

- The tangent cone to X at x (in the sense of Bouligand) is the set of tangent directions. It is denoted by

$$\mathrm{T}_x X \qquad \text{or} \qquad \mathrm{T}_X(x).$$

- Properties Let $x \in X$.

$$\mathrm{T}_x X \text{ is closed.}$$

$$X \text{ is convex} \quad \Longrightarrow \quad \mathrm{T}_x X \text{ is convex and } \mathrm{T}_x X = \overline{\mathbb{R}_+(X - x)}.$$

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

$$\nabla f(x_*) \in (\mathrm{T}_{x_*} X)^+.$$

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

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

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

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

Proof. Straightforward, using the convexity inequality

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

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

$$(P_E) \quad \left\{ \begin{array}{l} \inf_x \; f(x) \\ c(x) = 0, \end{array} \right.$$

where $f : \mathbb{E} \to \mathbb{R}$, $c : \mathbb{E} \to \mathbb{F}$ 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 X_E is convex.

Theorem (NC1 for (P_E), Lagrange, XVIIIth)

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*

$$\nabla_x \ell(x_*, \lambda_*) = 0, \qquad (1a)$$

$$c(x_*) = 0. \qquad (1b)$$

Some explanations.

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

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

  Qualification holds of $c'(x)$ is surjective (sufficient condition of CQ).
- The Lagrangian of (P_E) is the function

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

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*

$$\forall\, d \in \mathrm{T}_{x_*} X_E : \quad \left\langle \nabla^2_{xx} \ell(x_*, \lambda_*) d, d \right\rangle \geqslant 0. \qquad (3)$$

- Inequality in (3) is not necessarily true for $d \in \mathcal{N}(c'(x_*)) \setminus \mathrm{T}_{x_*} X_E$.
- $\nabla^2_{xx} \ell(x_*, \lambda_*)$ 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*

$$\forall\, d \in \mathrm{T}_{x_*} X_E \setminus \{0\} : \quad d^{\mathsf{T}} \nabla^2_{xx} \ell(x_*, \lambda_*) d > 0, \qquad (4)$$

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

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

A generic form of the nonlinear optimization problem:

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

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

- The feasible set is denoted by

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

- 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 and X_{EI} is convex.

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

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

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

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

Some explanations.

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

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

- The complementarity condition (5c) means

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

- The tangent cone $\mathrm{T}_x X_{EI}$ is always contained in the linearizing cone

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

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

$$\mathrm{T}_x X_{EI} = \mathrm{T}'_x X_{EI}. \tag{6}$$

- 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) $\sum_{i\in E\cup I^0(x)} \alpha_i \nabla c_i(x) = 0 \Longrightarrow \alpha = 0$ (Linear Independence),
  - (CQ-MF) $\sum_{i\in E\cup I^0(x)} \alpha_i \nabla c_i(x) = 0$ and $\alpha_{I^0(x)} \geqslant 0 \Longrightarrow \alpha = 0$ (Mangasarian-Fromovitz).

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:

1. *(CQ-MF) holds at x,*
2. $\forall\, v \in \mathbb{R}^m$, $\exists\, d \in \mathbb{E}$: $c'_E(x)\cdot d = v_E$ *and* $c'_{I^0(x)}(x)\cdot d \leqslant v_{I^0(x)}$,
3. $c'_E(x)$ *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
- *f is a convex function and X_{EI} is a convex set,*
- *f and c are differentiable at $x_* \in X_{EI}$,*
- *there is a $\lambda_* \in \mathbb{F}$ such that (x_*, λ_*) satisfies (5),*

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 (linear optimization duality)

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

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

- **Properties**

$$(P_L) \text{ has a solution} \iff \operatorname{val}(P_L) \in \mathbb{R},$$

$$\operatorname{val}(D_L) \leqslant \operatorname{val}(P_L) \qquad \text{[named weak duality]},$$

$$(P_L),\ (D_L) \text{ feasible} \iff \operatorname{Sol}(P_L) \neq \varnothing \iff \operatorname{Sol}(D_L) \neq \varnothing. \qquad (7)$$

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

Background

Algorithmics (speeds of convergence)

Let $\mathbb{E}$ be a normed space and $\{x_k\} \subseteq \mathbb{E}$ 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 $\forall\, k \geqslant K$, one has $\|x_{k+1} - \bar{x}\| \leqslant r\|x_k - \bar{x}\|$.
  - 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.

Lemma

If $\{x_k\} \to x_*$ superlinearly, then $\{x_{k+1} - x_k\} \sim \{x_k - x_*\}$.

Background

Algorithmics (Dennis & Moré criterion for superlinear convergence)

Let $F : \mathbb{E} \to \mathbb{F}$ 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

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

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

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

Optimality conditions

First order optimality conditions for (P_G) (the problem I)

- We consider the problem

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

  where
  - $f : \mathbb{E} \to \mathbb{R}$ ($\mathbb{E}$ is a Euclidean vector space),
  - $c : \mathbb{E} \to \mathbb{F}$ ($\mathbb{F}$ is another Euclidean vector space),
  - G is nonempty closed convex set in $\mathbb{F}$.
- The feasible set is denoted by

$$X_G := \{x \in \mathbb{E} : c(x) \in G\} = c^{-1}(G).$$

- (P_G) is said to be convex if f is convex and X_G is convex.

$$T : \mathbb{E} \multimap \mathbb{F} : x \mapsto c(x) - G \text{ is convex} \quad \Longrightarrow \quad X_G \text{ is convex.}$$

Optimality conditions

First order optimality conditions for (P_G) (the problem II)

Some examples of optimization problems that can be written in the form (P_G).

- The nonlinear optimization problem

$$(P_{EI}) \quad \begin{cases} \inf_{x \in \mathbb{R}} \ f(x) \\ c_E(x) = 0 \text{ and } c_I(x) \leqslant 0, \end{cases}$$

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

$$(P_{\text{SDO}}) \quad \begin{cases} \inf_{X \in \mathcal{S}^n} \ \langle C, X \rangle \\ A(X) = b \text{ and } X \succcurlyeq 0, \end{cases}$$

  where $C \in \mathcal{S}^n$, $A : \mathcal{S}^n \to \mathbb{R}^m$ is linear, and $b \in \mathbb{R}^m$.
- The composite optimization problem

$$\inf_{x \in \mathbb{E}} (g \circ f)(x),$$

  where $f : \mathbb{E} \to \mathbb{F}$ and $g : \mathbb{F} \to \mathbb{R}$.

Proposition (tangent and linearizing cones)

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

$$\mathrm{T}_x X_G \subseteq \mathrm{T}'_x X_G := \{d \in \mathbb{E} : c'(x) \cdot d \in \mathrm{T}_{c(x)} G\}.$$

- $\mathrm{T}'_x X_G$ is called the linearizing cone to X at x.
- The equality $\mathrm{T}_x X_G = \mathrm{T}'_x X_G$ is not guaranteed.
- The constraint function c is said to be qualified for representing X_G at x if

$$\mathrm{T}_x X_G = \mathrm{T}'_x X_G, \tag{9a}$$

$$c'(x)^*[(\mathrm{T}_{c(x)} G)^+] \text{ is closed.} \tag{9b}$$

Theorem (NC1 for (P_G))

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

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

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

$$\lambda_* \in \mathrm{N}_{c(x_*)} G. \tag{10b}$$

2. *if, furthermore, $G \equiv K$ is a convex cone, then (10b) can be written*

$$K^- \ni \lambda_* \perp c(x_*) \in K \qquad \text{or} \qquad c(x_*) \in \mathrm{N}_{\lambda_*} K^-. \tag{10c}$$

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

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

and in (10c) the complementarity conditions.

Theorem (SC1 for convex (P_G))

If
- *f is a convex function and X_G is a convex set,*
- *f and c are differentiable at $x_* \in X_G$,*
- *there is a $\lambda_* \in \mathbb{F}$ such that (x_*, λ_*) satisfies (10a)-(10b),*

then x_* is a global minimum of (P_G).

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 $y + X_G$ (y small),
2. claims the *stability* of the feasible set X_G,
3. ensures that c is *qualified* for representing X_G.

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

$$\text{(CQ-R)} \qquad \boxed{0 \in \operatorname{int}\Big(c(x) + c'(x) \cdot \mathbb{E} - G\Big).} \qquad (11)$$

We will see below that

- it is useful since it
  - provides an error bound for small perturbations $y + X_G$ of X_G,
  - claims the stability of the feasible set X_G with respect to small perturbations,
  - shows that the constraint function c is qualified for representing X_G at x, in the sense (9a)-(9b),
- it generalizes to (P_G) the Mangasarian-Fromovitz constraint qualification (CQ-MF) for (P_{EI}).

Theorem ((CQ-R) and metric regularity)

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

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

$$\operatorname{dist}(x, c^{-1}(y + G)) \leqslant \mu \operatorname{dist}(c(x), y + G). \tag{12}$$

Condition 2 is named *metric regularity* since it is equivalent to that property (see its definition below) for the multifunction

$$T : \mathbb{E} \multimap \mathbb{F} : x \mapsto c(x) - G.$$

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 $x_0 \in X_G$ and if (CQ-R) holds at $x = x_0$, then there exists a constant $\mu \geqslant 0$, such that

$$\forall\, x \text{ near } x_0 : \quad \operatorname{dist}(x, X_G) \leqslant \mu \operatorname{dist}(c(x), G). \tag{13}$$

- $\operatorname{dist}(x, X_G)$ is often difficult to evaluate in $\mathbb{E}$,
- $\operatorname{dist}(c(x), G)$ is often easier to evaluate in $\mathbb{F}$ (it is the case if $c(x)$ is easy to evaluate and G 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 $\operatorname{dist}(x, X_G)$).

Optimality conditions

First order optimality conditions for (P_G) (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 $x_0 \in X_G$ and if (CQ-R) holds at x_0, then, for all small $y \in \mathbb{F}$:

$$\{x \in \mathbb{E} : c(x) \in y + G\} \neq \varnothing.$$

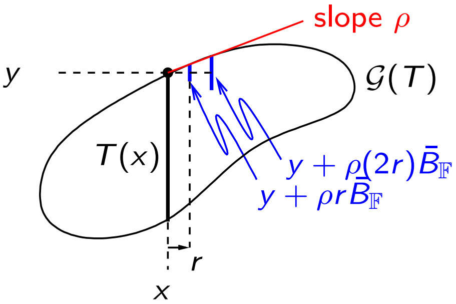
Optimality conditions

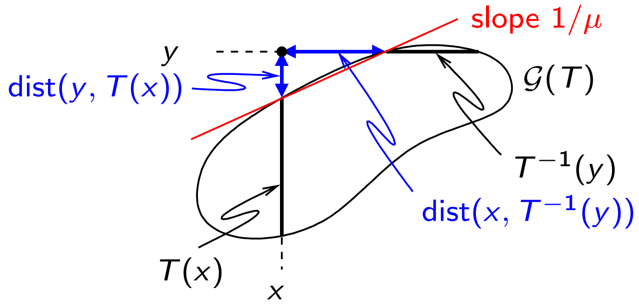
First order optimality conditions for (P_G) (open multifunction theorem I) [26, 23, 5]

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 $r \in [0, r_{\max}]$: $y + \rho\, r\, \bar{B}_{\mathbb{F}} \subseteq T(x + r\, \bar{B}_{\mathbb{E}})$.
- 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): $\operatorname{dist}(x, T^{-1}(y)) \leqslant \mu \operatorname{dist}(y, T(x))$.

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).

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. $y_0 \in \operatorname{int} \mathcal{R}(T)$,
2. *for all* $r > 0$, $y_0 \in \operatorname{int} T(x + r\bar{B}_{\mathbb{E}})$,
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* ρ *is given by point 3.*

Theorem (metric regularity diffusion)

Suppose that

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

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* $(x_0, y_0 + \delta(x_0))$, *the following holds*

$$\operatorname{dist}(x, \tilde{T}^{-1}(y)) \leqslant \frac{\mu}{1 - L\mu} \operatorname{dist}(y, \tilde{T}(x)). \tag{14}$$

No need of convexity.

Optimality conditions

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

Proposition (other forms of (CQ-R))

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

1. $0 \in \operatorname{int}(c(x) + c'(x)\mathbb{E} - G)$ *[this is (CQ-R)],*
2. $c'(x)\mathbb{E} - \mathrm{T}^f_{c(x)}\, G = \mathbb{F}$,
3. $c'(x)\mathbb{E} - \mathrm{T}_{c(x)}\, G = \mathbb{F}$,
4. $\overline{c'(x)\mathbb{E} - \mathrm{T}_{c(x)}\, G} = \mathbb{F}$.

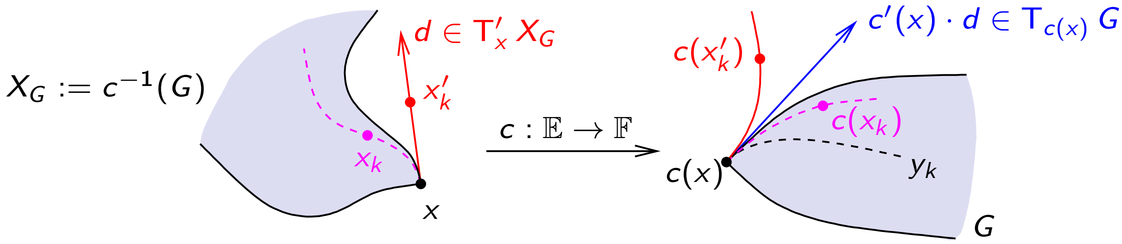
Optimality conditions

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

Proposition (qualification with (CQ-R))

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

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



Proposition (Gauvin's boundedness property)

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

1. $\Lambda_*^\infty = [c'(x_*)\mathbb{E} - \mathrm{T}_{c(x_*)}\, G]^+$,
2. *Λ_* is bounded if and only if (CQ-R) holds.*

The property was originally established for problem (P_{EI}) and (CQ-MF) [10].

Let x_* be a local solution to (P_{EI}), λ_* be an associated optimal multiplier, and $L_* := \nabla^2_{xx}\ell(x_*, \lambda_*)$.

In view of the second order optimality conditions of the equality constrained optimization problem (P_E), it is tempting to claim that

$$\forall\, d \in \mathrm{T}_{x_*} X_{EI} : \quad \langle L_* d, d\rangle \geqslant 0.$$

But this is not guaranteed!

- The good cone is not $\mathrm{T}_{x_*} X_{EI}$ but the critical cone C_* (to be defined).
- The multiplier λ_* must be chosen, depending on $d \in C_*$.

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

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

- Short notation for index sets

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

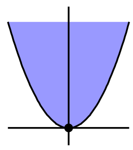
- Other forms of the critical cone at a stationary pair (x_*, λ_*):

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

Three instructive examples

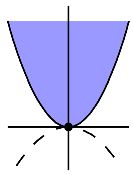
1. Strong NC2 (not always true): $\forall\, \lambda_* \in \Lambda_*$, $\forall\, d \in C_*$: $\langle L_* d, d\rangle \geqslant 0$.

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



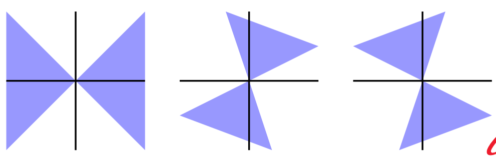
2. Semi-strong NC2 (not always true): $\exists\, \lambda_* \in \Lambda_*$, $\forall\, d \in C_*$: $\langle L_* d, d\rangle \geqslant 0$.

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



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

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



Optimality conditions

Second order optimality conditions (NC2)

Notation at a stationary point x_*:

- $\Lambda_* := \{\lambda_* \in \mathbb{R}^m : (x_*, \lambda_*)$ satisfies the KKT system (5)$\}$.
- $L_* := \nabla^2_{xx}\ell(x_*, \lambda_*)$ 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* $\langle L_* d, d\rangle \geqslant 0$.

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

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

Optimality conditions

Second order optimality conditions (SC2 I)

Theorem (SC2 for (P_{EI}))

Suppose that

- f *and* $c_{E\cup I^0_*}$ *are twice differentiable at* x_*,
- (x_*, λ_*) *verifies the KKT conditions* (5),
- *the following equivalent conditions hold*

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

$$\exists\, \bar\gamma > 0, \quad \forall\, d \in C_*, \quad \exists\, \lambda_* \in \Lambda_* : \quad \langle L_* d, d\rangle \geqslant \bar\gamma\, \|d\|^2, \tag{16b}$$

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

$$f(x) > f(x_*) + \frac{\gamma}{2}\, \|x - x_*\|^2. \tag{17}$$

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

Optimality conditions

Second order optimality conditions (SC2 II)

- Condition (17) is called the quadratic growth property.
- No need of a constraint qualification assumption.
- The property

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

is stronger than (16) 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.

Perturbation analysis

Stability result for (P_G) with a polyhedral cone G I [22]

- Let $\mathbb{P}$ be a vector space. For $p \in \mathbb{P}$, consider the perturbed problem

$$(P_K^p) \quad \begin{cases} \min_x \; f(x,p) \\ c(x,p) \in K, \end{cases}$$

where $f : \mathbb{E} \times \mathbb{P} \to \mathbb{R}$ smooth, $c : \mathbb{E} \times \mathbb{P} \to \mathbb{F}$ smooth, $K \subseteq \mathbb{F}$ convex polyhedral cone.

- Optimality system at $x \in \mathbb{E}$: $\exists\, \lambda \in \mathbb{F}$ such that

$$\begin{cases} \nabla_x f(x,p) + c'_x(x,p)^* \lambda = 0 \\ K^- \ni \lambda \perp c(x,p) \in K. \end{cases} \tag{18}$$

- The multiplier multifunction $\Lambda : \mathbb{E} \times \mathbb{P} \multimap \mathbb{F}$ is defined at $(x,p) \in \mathbb{E} \times \mathbb{P}$ by

$$\Lambda(x,p) := \{\lambda \in \mathbb{F} : (x,p,\lambda) \text{ satisfies } (18)\}.$$

- The stationary multifunction $\Sigma : \mathbb{P} \multimap \mathbb{E}$ is defined at $p \in \mathbb{P}$ by

$$\Sigma(p) := \{x \in \mathbb{E} : \Lambda(x,p) \neq \varnothing\}.$$

Assume the framework defined above.

Proposition (stability of (P_K))

If
- $f(\cdot, p_0)$ and $c(\cdot, p_0)$ are C^2 near $x_0 \in \mathbb{E}$ for some $p_0 \in \mathbb{P}$,
- f'_x, c and c'_x are Lipschitz continuous near (x_0, p_0),
- $0 \in \operatorname{int}\big(c(x_0, p_0) + c'_x(x_0, p_0)\mathbb{E} - K\big)$,
- $x_0 \in \Sigma(p_0)$,
- $\exists\, \lambda_0 \in \Lambda(x_0, p_0)$ such that strong SC2 holds for $(P_K^{p_0})$,

then $\exists\, L \geqslant 0$, such that $\forall\, p$ near p_0:

1. $\Sigma(p) \neq \varnothing$,
2. $\forall\, x \in \Sigma(p)$ near x_0, $\forall\, \lambda \in \Lambda(x, p)$:

$$\operatorname{dist}\big((x, \lambda), \{x_0\} \times \Lambda(x_0, p_0)\big) \leqslant L\, \|p - p_0\|.$$

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.

   Features
   - Expensive iteration (nonlinear), fast convergence, easy to globalize.

   Examples
   - The Josephy-Newton algorithm for functional inclusions.
   - The SQP algorithm for (P_{EI}).

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

   Features
   - Cheap iteration (linear), fast convergence, difficult to globalize.

   Example
   - The semismooth Newton algorithm for nonsmooth system of equations.

Linearization methods

Josephy-Newton algorithm for functional inclusions (functional inclusion I)

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

We consider the functional inclusion problem

$$(P_{FI}) \qquad \boxed{F(x) + N(x) \ni 0.} \tag{19}$$

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

Examples

1. The variational problem if $N = \mathrm{N}_X$ the normal cone to $X \subseteq \mathbb{E}$ at x ($= \varnothing$ if $x \notin X$):
$$(P_V) \qquad F(x) + \mathrm{N}_X(x) \ni 0. \tag{20}$$
2. The variational inequality problem is (P_V) with $X = C$ (a closed convex set):
$$(P_{VI}) \qquad \begin{cases} x \in C \\ \langle F(x), y - x \rangle \geqslant 0, \quad \forall y \in C. \end{cases} \tag{21}$$
3. The complementarity problem is (P_{FI}) with $N = N_{K^+} \circ G$ (closed convex cone $K \subseteq \mathbb{F}$, $G : \mathbb{E} \to \mathbb{F}$)
$$(P_{CP}) \qquad K^+ \ni G(x) \perp F(x) \in K. \tag{22}$$

Linearization methods

Josephy-Newton algorithm for inclusions (functional inclusion II)

Examples (continued)

4. The Peano-Kantorovitch NC1 of problem $\min\{f(x) : x \in X\}$ reads
$$\nabla f(x) + \mathrm{N}_X(x) \ni 0.$$
5. The first order optimality conditions for (P_G), when $G \equiv K$ is a closed convex *cone*, can be written
$$\tilde{K}^- \ni \tilde{x} \perp \tilde{F}(\tilde{x}) \in \tilde{K},$$
with the variable $\tilde{x} := (x, \lambda) \in \mathbb{E} \times \mathbb{F}$,
$$\tilde{F}(\tilde{x}) := \begin{pmatrix} \nabla f(x) + c'(x)^* \lambda \\ -c(x) \end{pmatrix}, \quad \text{and} \quad \tilde{K} := \{0_{\mathbb{E}}\} \times (-K). \tag{23}$$
6. If N is the constant multifunction $x \multimap \{0_{\mathbb{R}}^{m_E}\} \times \mathbb{R}_+^{m_I} \subseteq \mathbb{R}^m \equiv \mathbb{F}$ where E and I make a partition of $[1:m]$, (P_{FI}) becomes the system of equalitites and inequalitites
$$F_E(x) = 0 \qquad \text{and} \qquad F_I(x) \leqslant 0.$$

Linearization methods

Josephy-Newton algorithm for inclusions (the JN algorithm) [15, 16]

Algorithm (Josephy-Newton algorithm for solving (P_{FI}))

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

$$F(x_k) + M_k(x - x_k) + N(x) \ni 0, \tag{24}$$

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_{FI})).
- (24) captures more information from (P_{FI}) than a "simple" linearization.
- (24) is often a nonlinear problem, hence yielding an expensive iteration.
- Makes sense computationally if N is sufficiently simple.

Linearization methods

Josephy-Newton algorithm for inclusions (semi-stability) [3]

A solution x_* to (P_{FI}) is said semi-stable if $\exists\, \sigma_1 > 0$ and $\exists\, \sigma_2 > 0$ such that

$$\left.\begin{array}{r}(x,p) \in \mathbb{E} \times \mathbb{F} \\ F(x) + N(x) \ni p \\ \|x - x_*\| \leqslant \sigma_1\end{array}\right\} \quad \Longrightarrow \quad \|x - x_*\| \leqslant \sigma_2 \|p\|.$$

Remarks

- The perturbed inclusion $F(x) + N(x) \ni p$ is not required to have a solution.
- Semistability implies that x_* is the unique solution to (P_{FI}) on $\bar{B}(x_*, \sigma_1)$.
- If $N \equiv \{0\}$, then semi-stability of x_* $\Leftrightarrow$ injectivity of $F'(x_*)$ (hence, nonsingularity if $\dim \mathbb{E} = \dim \mathbb{F}$).

Proposition (semi-stability characterization for polyhedral VI)

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

1. x_* *is semi-stable,*
2. x_* *is an isolated solution to*
$$F(x_*) + F'(x_*)(x - x_*) + \mathrm{N}_C(x) \ni 0,$$
3. *one has* $\langle F'(x_*)(x - x_*), x - x_* \rangle > 0$ *when* $x \in C \setminus \{x_*\}$ *satisfies*
$$\langle F(x_*), x - x_* \rangle = 0 \quad \text{and} \quad F(x_*) + F'(x_*)(x - x_*) + \mathrm{N}_C(x_*) \ni 0,$$
4. x_* *is the unique solution to*
$$\mathrm{N}_C(x) \subseteq \mathrm{N}_C(x_*),\ \langle F(x_*), x - x_* \rangle = 0,\ \mathbb{R}_+ F(x_*) + F'(x_*)(x - x_*) + \mathrm{N}_C(x) \ni 0.$$

Proposition (speed of convergence of quasi-Newton methods)

Suppose that F *is* C^1 *near a* semi-stable *solution* x_* *to* (P_{FI}). *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* semi-stable *solution* x_* *to* (P_{FI}). *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.*

Linearization methods

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

A solution x_* to (P_{FI}) is said hemi-stable 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 semi-stable and hemi-stable solution x_* to (P_{FI}). 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 $\{x_k\} \subseteq \bar{B}(x_*, \varepsilon)$,*
2. *any sequence $\{x_k\}$ generated in $\bar{B}(x_*, \varepsilon)$ by the JN algorithm converges superlinearly to x_* (quadratically if F is $C^{1,1}$).*

Linearization methods

The SQP algorithm (overview)

Recall the equality and inequality constrained problem

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

Three popular methods to solve (P_{EI})

- Augmented Lagrangian methods: a dual method that generates $\{(\lambda_k, r_k)\} \subseteq \mathbb{R}^m \times \mathbb{R}_{++}$ by
  - $\inf\limits_{x,\; s\geqslant 0} \left(f(x) + (\lambda_k)_E^{\mathsf{T}} c_E(x) + \frac{r_k}{2}\|c_E(x)\|_2^2 + (\lambda_k)_I^{\mathsf{T}}(c_I(x){+}s) + \frac{r_k}{2}\|c_I(x){+}s\|_2^2 \right)$,
  - $\lambda_{k+1} := (\lambda_k + r_k c(x_k))^{\#}$ 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:

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

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, λ) of (P_{EI}):

$$\nabla_x \ell(x, \lambda) = 0, \tag{26a}$$
$$c_E(x) = 0, \tag{26b}$$
$$0 \leqslant \lambda_I \perp c_I(x) \leqslant 0. \tag{26c}$$

- This system in (x, λ) can be written like (P_{FI}) or (P_{CP}), namely

$$F(x, \lambda) + \mathrm{N}_{K^+}(x, \lambda) \ni 0 \qquad \text{or} \qquad K^+ \ni (x, \lambda) \perp F(x, \lambda) \in K, \tag{27}$$

with the data

$$F(x, \lambda) = \begin{pmatrix} \nabla_x \ell(x, \lambda) \\ -c(x) \end{pmatrix} \qquad \text{and} \qquad K = \{0_{\mathbb{E}}\} \times (\{0_{\mathbb{R}^{m_E}}\} \times \mathbb{R}^{m_I}_+). \tag{28}$$

Hence, $K^+ = \mathbb{E} \times \left(\mathbb{R}^{m_E} \times \mathbb{R}^{m_I}_+\right)$.

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 functional inclusion (27)-(28):

$$\nabla_x \ell(x_k, \lambda_k) + M_k(x - x_k) + c'(x_k)^*(\lambda - \lambda_k) = 0, \tag{29a}$$
$$c_E(x_k) + c_E'(x_k) \cdot (x - x_k) = 0, \tag{29b}$$
$$0 \leqslant \lambda_I \perp \left(c_I(x_k) + c_I'(x_k) \cdot (x - x_k)\right) \leqslant 0, \tag{29c}$$

where $M_k = L_k := \nabla^2_{xx} \ell(x_k, \lambda_k)$ or an approximation to it ($M_k \not\simeq \nabla^2 f(x_k)$!).

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

$$(\text{OQP}) \qquad \begin{cases} \min_x \ \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle M_k(x - x_k), x - x_k \rangle \\ c_E(x_k) + c_E'(x_k) \cdot (x - x_k) = 0, \\ c_I(x_k) + c_I'(x_k) \cdot (x - x_k) \leqslant 0. \end{cases} \tag{30}$$

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

Algorithm (local SQP)

From (x_k, λ_k) *to* (x_{k+1}, λ_{k+1})*:*

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

Remarks

- The OQP's are still hard to solve (not just a linear system, expensive iteration):
  - If $M_k = L_k \not\succcurlyeq 0$, 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.

Proposition (semi-stability and SC2)

If x_* *is a local minimum of* (P_{EI}) *and* λ_* *is an associated mutiplier, then the following properties are equivalent:*

1. (x_*, λ_*) *is semi-stable,*
2. $\Lambda_* = \{\lambda_*\}$ *and* x_* *satisfies* $SC2$*.*

At a local solution, semi-stability implies hemi-stability:

Proposition (SC for hemi-stability)

If
- x_* *is a local minimum of* (P_{EI})*,*
- (x_*, λ_*) *satisfies the KKT conditions,*
- (x_*, λ_*) *is semi-stable,*

then (x_*, λ_*) *is hemi-stable.*

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 λ_* associated with x_*,*
- *SC2 is satified,*

then there exists a neighborhood V of (x_*, λ_*) 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_*, λ_*).*

- Consider the nonsmooth penalty function $\Theta_\sigma : \mathbb{E} \to \mathbb{R}$ associated with (P_{EI}):
$$\Theta_\sigma(x) := f(x) + \sigma \|c(x)^\#\|,$$
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
$$\|u\|_D := \sup_{\|v\| \leqslant 1} u^\top v.$$

Proposition (exact penalty property)

If
- *f and c are C^2 near a local minimizer x_* of (P_{EI}),*
- *the set Λ_* of associated optimal multipliers is nonempty,*
- *weak SC2 holds,*
- *$\sigma \geqslant \sup\{\|\lambda_*\|_D : \lambda_* \in \Lambda_*\}$ and $\sigma > \|\hat{\lambda}_*\|_D$ for some $\hat{\lambda}_* \in \Lambda_*$,*

then x_* is a strict local minimum of Θ_σ.

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

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

We often make the following assumption (true for ℓ_p norms)

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

Proposition (descent direction)

If
- $(d_k, \lambda_k^{\mathrm{QP}})$ *is a stationary pair of* (31),
- $\| \cdot^\# \|$ *is convex,*

then

1. $\Theta'_\sigma(x_k; d_k) \leqslant -\langle M_k d_k, d_k \rangle + (\lambda_k^{\mathrm{QP}})^\mathsf{T} c(x_k) - \sigma \|c(x_k)^\#\|$,
2. $\Theta'_\sigma(x_k; d_k) < 0$, *if* $\sigma \geqslant \|\lambda_k^{\mathrm{QP}}\|_D$, $M_k \succ 0$, *and* x_k *is not stationary.*

The SQP algorithm with linesearch forces $M_k \succ 0$ and minimizes the changing nondifferentiable merit function Θ_{σ_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* $(d_k, \lambda_k^{\mathrm{QP}})$ *(if any),*
- *impose* $\sigma_k \geqslant \|\lambda_k^{\mathrm{QP}}\|_D + \bar{\sigma}$ *keeping* σ_k *constant if* $\{\lambda_k^{\mathrm{QP}}\}$ *is bounded,*
- *linesearch:* $\alpha_k > 0$ *such that* $\Theta_{\sigma_k}(x_k + \alpha_k d_k) \leqslant \Theta_{\sigma_k}(x_k) + \omega \alpha_k \Delta_k$,
- $x_{k+1} := x_k + \alpha_k d_k$ *and* $\lambda_{k+1} := \lambda_k + \alpha_k(\lambda_k^{\mathrm{QP}} - \lambda_k)$,
- *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 $\Theta'_{\sigma_k}(x_k; d_k)$.

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* $(d_k, \lambda_k^{\rm QP})$ *for all* $k \geqslant 1$,
- $\{\lambda_k^{\rm QP}\}$ *is bounded,*
- $\Theta_{\sigma_k}(x_k)$ *is bounded below,*

then the KKT conditions are satisfied asymptotically, meaning that $\nabla_x \ell(x_k, \lambda_k^{\rm QP}) \to 0$, $c(x_k)^{\#} \to 0$, $(\lambda_k^{\rm QP})_I \geqslant 0$, *and* $(\lambda_k^{\rm QP})_I^{\mathsf T} c_I(x_k) \to 0$.

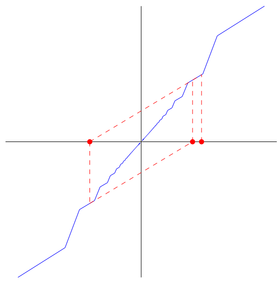
- Let $\mathbb{E}$ and $\mathbb{F}$ be (finite dimensional) normed spaces and $\Omega \subseteq \mathbb{E}$ be open.
- We consider the problem of fiding a zero of a nonsmooth function $F : \Omega \to \mathbb{F}$:
$$F(x) = 0. \tag{32}$$
- **Examples**
  1. The CP $(0 \leqslant x \perp \Phi(x) \geqslant 0)$ can be represented by (32):
  $$\Psi(x, \Phi(x)) = 0,$$
  where $\Psi(u, v) = \{\psi(u_i, v_i)\}_i$ and ψ is a C-function, meaning that $\psi(a, b) = 0$ iff $a \geqslant 0$, $b \geqslant 0$, and $ab = 0$. Examples (F = Fischer):
  $$\psi_{\min}(a, b) = \min(a, b) \quad \text{and} \quad \psi_F(a, b) = \sqrt{a^2 + b^2} - (a + b).$$
  2. The VI problem (find $x \in C$ s.t. $\langle \Phi(x), y - x\rangle \geqslant 0$ for all $y \in C$) can be written llike (32):
  $$\mathrm{P}_C(x - \Phi(x)) - x = 0.$$

Linearization methods

The semismooth Newton method (motivation)

- Suppose that $F : \mathbb{E} \to \mathbb{F}$ is Lipschitz continuous near $x_* \in \Omega$.
- Then $\partial_C F(x) \neq \varnothing$ 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 [18]:

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



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

Linearization methods

The semismooth Newton method (generalized differentiability I) [6]

For a function $F : \mathbb{E} \to \mathbb{F}$ ($\mathbb{E}$, $\mathbb{F}$ 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 $\Omega \subseteq \mathbb{E}$, then the Lebesgue measure of $\Omega \setminus \mathcal{D}_F$ is zero; in particular $\mathcal{D}_F$ is dense in Ω, i.e., $\Omega \subseteq \overline{\mathcal{D}_F}$.

- The B-differential (B honoring Bouligand) of F at x is the set
$$\partial_B F(x) := \{J \in \mathcal{L}(\mathbb{E}, \mathbb{F}) : \exists\, \{x_k\} \subseteq \mathcal{D}_F \text{ such that } x_k \to x,\ F'(x_k) \to J\}.$$
- The C-differential (C for Clarke) of F at x is the set
$$\partial_C F(x) := \operatorname{co} \partial_B F(x).$$

Proposition (compactness and upper semi-continuity)

If F is L-Lipschitz near x, then

1. $\partial_C F(x)$ *is nonempty compact* ($\subseteq L\bar{B}$) *and convex,*
2. $\partial_C F$ *is upper semi-continuous 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 Lipschitz near x and $\partial_C F(x)$ is nonsingular, then there are constants $C > 0$ and $\delta > 0$ such that

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

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

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

  (S_1) F is Lipschitz near x,
  (S_2) $F'(x; h)$ exists for all $h \in \mathbb{E}$,
  (S_3) for $h \to 0$ in $\mathbb{E}$, the following holds

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

- $F : \Omega \to \mathbb{F}$ is strongly semismooth at $x \in \Omega$ if it is semismooth at x with (S_3) strengthened by

  (S_4) for h small in $\mathbb{E}$, the following holds

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

Linearization methods

The semismooth Newton method (semismoothness - properties I) [9, 14]

Proposition (differentiable function)

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

Proposition (convex function)

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

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

Proposition (piecewise semismooth function)

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

When the pieces are affine, F is said to be piecewise affine at x.

Proposition (piecewise affine function)

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

Linearization methods

The semismooth Newton method (semismoothness - properties II) [9, 14]

Proposition (componentwise semismooth function)

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

Proposition (composition of functions)

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

Proposition (calculus)

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

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

Examples

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

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).

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. *if* F *is strongly semismooth, then the convergence is quadratic.*

Notation and first properties

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

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

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

$$\begin{aligned} A \succcurlyeq 0 \quad &\overset{\text{def}}{\Longleftrightarrow} \quad A \in \mathcal{S}^n_+ \quad \Longleftrightarrow \quad \lambda(A) \subseteq [0, +\infty), \\ A \succcurlyeq 0 \quad &\Longleftrightarrow \quad \forall\, B \succcurlyeq 0 : \ \langle A, B \rangle \geqslant 0, \\ \text{if } A \succcurlyeq 0 \text{ and } B \succcurlyeq 0, \text{ then} \quad &\langle A, B \rangle = 0 \quad \Longleftrightarrow \quad AB = 0, \\ \mathrm{T}_A \mathcal{S}^n_+ &= \{D \in \mathcal{S}^n : v^\mathsf{T} D v \geqslant 0, \ \text{for all } v \in \mathcal{N}(A)\}. \end{aligned} \tag{33}$$

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:

$$\begin{aligned} A \succ 0 \quad &\overset{\text{def}}{\Longleftrightarrow} \quad A \in \mathcal{S}^n_{++} \quad \Longleftrightarrow \quad \lambda(A) \subseteq (0, +\infty), \\ A \succ 0 \quad &\Longleftrightarrow \quad \forall\, B \in \mathcal{S}^n_+ \setminus \{0\} : \ \langle A, B \rangle > 0, \\ A \succcurlyeq 0 \text{ and } [v^\mathsf{T} M v > 0, \ \forall\, v \in \mathcal{N}(A) \setminus \{0\}] \quad &\Longrightarrow \quad M + rA \succ 0 \text{ for large} \end{aligned}$$

Semidefinite optimization

Problem definition (primal and dual problems)

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

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

where

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

Notation

- Feasible sets: $\mathcal{F}_P := \{X \in \mathcal{S}^n_+ : \mathcal{A}(X) = b\}$, $\mathcal{F}_D := \{(y, S) \in \mathbb{R}^m \times \mathcal{S}^n_+ : \mathcal{A}^*(y) + S = C\}$, and $\mathcal{F} := \mathcal{F}_D \times \mathcal{F}_D$.
- Strictly feasible sets: $\mathcal{F}^s_P := \{X \in \mathcal{S}^n_{++} : \mathcal{A}(X) = b\}$, $\mathcal{F}^s_D := \{(y, S) \in \mathbb{R}^m \times \mathcal{S}^n_{++} : \mathcal{A}^*(y) + S = C\}$, and $\mathcal{F}^s := \mathcal{F}^s_D \times \mathcal{F}^s_D$.
- Optimal values: $\mathrm{val}(P)$ and $\mathrm{val}(D)$.
- Solution sets: $\mathrm{Sol}(P)$ and $\mathrm{Sol}(D)$.

Semidefinite optimization

Problem definition (Lagrange dualization consequences)

The Lagrangian of problem (P) is the function $\ell : \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n \to \mathbb{R}$ defined at $(X, y, S) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ by

$$\ell(X, y, S) = \langle C, X\rangle - \langle y, A(X) - b\rangle - \langle S, X\rangle .$$

Proposition (consequences of the Lagrangian dualization)

1. $\mathrm{val}(D) \leqslant \mathrm{val}(P)$.
2. $(X, y, S) \in \mathcal{F} \implies \langle C, X\rangle - \langle b, y\rangle = \langle X, S\rangle \geqslant 0$.
3. $(X, y, S) \in \mathcal{F}$, $\langle X, S\rangle = 0$
   $\iff X \in \mathrm{Sol}(P)$, $(y, S) \in \mathrm{Sol}(D)$, $\mathrm{val}(D) = \mathrm{val}(P)$,
   $\iff (X, (y, S))$ *is a saddle-point of* ℓ *on* $\mathcal{S}^n \times (\mathbb{R}^m \times \mathcal{S}^n_+)$.

Remarks

- One says that there is a duality gap if $\mathrm{val}(D) < \mathrm{val}(P)$.
- $X \in \mathrm{Sol}(P)$, $(y, S) \in \mathrm{Sol}(D) \quad \not\Rightarrow \quad \mathrm{val}(D) = \mathrm{val}(P)$.

Semidefinite optimization
Problem definition (examples of SDO formulations)

The Schur complement of $A \succ 0$ in

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

is $(A\,|\,K) := C - B^\mathsf{T} A^{-1} B$. The following holds

$$K \succ 0 \quad \Longleftrightarrow \quad \begin{cases} A \succ 0 \\ (A\,|\,K) \succ 0. \end{cases}$$

Examples of SDO modelling

1. Linear optimization.
2. Convex quadratic optimization.
3. Global minimization of polynomials (relaxation if # variables $\geqslant 2$).
4. Rank 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. $\mathcal{F}_P \times \mathcal{F}_D^s \neq \varnothing \quad \Longrightarrow \quad \mathrm{Sol}(P) \neq \varnothing$ *and compact.*
2. $\mathcal{F}_P^s \times \mathcal{F}_D \neq \varnothing \quad \Longrightarrow \quad \mathrm{Sol}(D) \cap (\mathcal{R}(\mathcal{A}) \times \mathcal{S}^n) \neq \varnothing$ *and compact.*
3. $\mathcal{F}^s \neq \varnothing \quad \Longrightarrow \quad \mathrm{Sol}(P)$ *and* $\mathrm{Sol}(D) \cap (\mathcal{R}(\mathcal{A}) \times \mathcal{S}^n) \neq \varnothing$ *and compact.*

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

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

Proposition (optimality conditions)

Suppose that $\mathcal{F}^s \neq \varnothing$, then

$$(X,(y,S)) \in \text{Sol}(P) \times \text{Sol}(D) \quad \Longleftrightarrow \quad \left\{ \begin{array}{ll} \mathcal{A}^*(y) + S = C, & S \succcurlyeq 0, \\ \mathcal{A}(X) = b, & X \succcurlyeq 0, \\ \langle X, S \rangle = 0. & \end{array} \right. \tag{35}$$

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

$$(P) \quad \curvearrowright \quad (P_\mu) \quad \left\{ \begin{array}{l} \inf_X \ \langle C, X \rangle + \mu \, \text{ld}(X) \\ A(X) = b, \end{array} \right.$$

$$(D) \quad \curvearrowright \quad (D_\mu) \quad \left\{ \begin{array}{l} \sup_{(y,S)} \ \langle b, y \rangle - \mu \, \text{ld}(S) \\ A^*(y) + S = C, \end{array} \right.$$

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

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

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

$$\text{ld}'(X) \cdot H = -\left\langle X^{-1}, H \right\rangle,$$
$$\text{ld}''(X) \cdot (H,K) = \left\langle X^{-1} H X^{-1}, K \right\rangle,$$
$$\text{ld}^\infty = \mathcal{I}_{\mathcal{S}^n_+}.$$

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

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

Proposition (existence and smoothness of the central path)

Suppose that $\mathcal{F}^s \neq \varnothing$ and $\mu > 0$. Then,

1. *the system (O_μ) has a solution (X_μ, y_μ, S_μ), unique in $\mathcal{S}^n_{++} \times \mathcal{R}(\mathcal{A}) \times \mathcal{S}^n_{++}$,*
2. *the map $\mu \in \mathbb{R}_{++} \mapsto (X_\mu, y_\mu, S_\mu) \in \mathcal{S}^n_{++} \times \mathcal{R}(\mathcal{A}) \times \mathcal{S}^n_{++}$ is C^∞.*

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 $\mathscr{C}$ ($\theta \in (0,1)$ is a parameter that determines the size of the neighborhood). Each iteration proceeds along a Newton direction aiming a moving point on $\mathscr{C}$, whose central parameter if $\sigma\bar{\mu}(z)$ 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. $z_+ := z + \alpha d$.

M.F. Anjos, J.B. Lasserre (2012).
Handbook on Semidefinite, Conic and Polynomial Optimization, volume 166 of *International Series in Operations Research & Management Science*.
Springer.
[doi].

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] [doi].

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.
Reprinted in 1990 by SIAM, Classics in Applied Mathematics 5 [doi].

R. Cominetti (1990).
Metric regularity, tangent sets, and second-order optimality conditions.
Applied Mathematics and Optimization, 21(1), 265–287.
[doi].

A.L. Dontchev, R.T. Rockafellar (2009).
Implicit Functions and Solution Mappings – A View from Variational Analysis.
Springer Monographs in Mathematics. Springer.

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 (2018).
Fragments 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).
Newton-Type Methods for Optimization and Variational Problems.
Springer Series in Operations Research and Financial Engineering. Springer.
[doi].

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, volume 60 of *Nonconvex Optimization and Its Applications*.
Kluwer Academic Publishers, Dordrecht.
[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. Qi (1993).
Convergence analysis of some algorithms for solving nonsmooth equations.
Mathematics of Operations Research, 18, 227–244.
[doi].

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.
[doi].

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].

H. Wolkowicz, R. Saigal, L. Vandenberghe (editors) (2000).
Handbook of Semidefinite Programming – Theory, Algorithms, and Applications, volume 27 of *International Series in Operations Research & Management Science*.
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.