Optimisation Différentiable

Théorie et Algorithmes

Partie I

Exemple de résumé du cours

J. Ch. GILBERT

14 novembre 2020

Informations pratiques

- 14 séances, échappatoire possible à mi-parcours.
- Objectif du cours : l'optimisation
  - aspects théoriques : convexité, CO, dualité, ...,
  - aspects pratiques : algorithmes.
- Organisation :
  - Partie I : 7 séances, dont 1/2 pour l'examen.
  - CM : 6 séances d'1h15++,
  - TD + TP : 4+3 séances d'2h00--,
  - TP : projet d'optimisation (Matlab),
  - travail personnel.
- Supports de cours
  - syllabus [site] : ne pas voir les sections avec $\ominus$,
  - planches [pdf] : points importants du cours [§],
  - notes manuscrites [SP] : 1 document par séance,
  - exercices : en TD (+sol), dans le syllabus (−sol).
- Contrôle des connaissances
  - TP : rapport et code incrémental à remettre,
  - Séance 7 : résolution de problèmes (1h30).

Plan du cours I

1. Introduction : optimisation et analyse convexe
   TD1 (rappels, concepts de base)
2. Conditions d'optimalité I : méthode et outils
   TP1
3. Conditions d'optimalité II : égalités et inégalités
   TD2 (conditions d'optimalité)
4. Conditions d'optimalité III : CO2 égalités et
   Méthode de descente : RL
   TP2
5. Méthodes newtoniennes : N et qN
   TD3 (recherche linéaire, moindres-carrés)
6. Dualité
   TD4 (dualité)
7. Contrôle des connaissances
   TP3

I Introduction

Vocabulaire de l'optimisation (§ 1.1)

Le problème à résoudre :

$$(P_X) \qquad \inf_{x \in X} f(x).$$

Quelques définitions et conventions :

- $f : X \to \overline{\mathbb{R}}$ est appelée critère ou fonction-coût ou fonction-objectif,
- X est appelé ensemble admissible,
- un point de X est dit admissible,
- $\text{val}(P_X) := \inf_{x \in X} f(x) \in \overline{\mathbb{R}}$ est la valeur optimale,
- (P_X) est dit réalisable si $X \neq \varnothing$,
- convention : $\inf_{x \in \varnothing} f(x) = +\infty$,
- $\sup_{x \in X} f(x) = -\inf_{x \in X} \big[-f(x)\big]$ (mêmes solutions),
- (P_X) est dit non borné si $\text{val}(P_X) = -\infty$, i.e., $\exists \{x_k\} \subset X$ telle que $f(x_k) \to -\infty$.

Si $X \neq \varnothing$, il existe une suite minimisante $\{x_k\}$, laquelle vérifie

- $\{x_k\} \subset X$,
- $f(x_k) \to \text{val}(P_X)$.

On dit que x_* est solution de (P_X) si

- $x_* \in X$,
- $\forall\, x \in X$: $f(x_*) \leqslant f(x)$.

On dit aussi minimum ou minimiseur.

On note l'ensemble des solutions

$$\text{Sol}(P_X) \qquad \text{ou} \qquad \arg\min_{x \in X} f(x).$$

On dit que x_* est solution stricte de (P_X) si

- $x_* \in X$,
- $\forall\, x \in X \setminus \{x_*\}$: $f(x_*) < f(x)$.

Si X topologique, on dit que x_* est solution locale de (P_X) s'il existe un voisinage V de x_* tel que

- $x_* \in X$,
- $\forall\, x \in X \cap V$: $f(x_*) \leqslant f(x)$.

Si X topologique, on dit que x_* est solution locale stricte de (P_X) s'il existe un voisinage V de x_* tel que

- $x_* \in X$,
- $\forall\, x \in (X \cap V) \setminus \{x_*\}$: $f(x_*) < f(x)$.

Existence de solution (§ 1.2)

Le problème à résoudre ($f : X \to \mathbb{R} \cup \{+\infty\}$) :

$$(P_X) \quad \left\{ \begin{array}{l} \inf f(x) \\ x \in X. \end{array} \right.$$

On dit que f est fermée si $(\text{epi } f)$ est fermé.

$$\boxed{\begin{array}{l} \underline{\text{Si}} \ \bullet \ f \text{ est } \underline{\text{fermée}} \text{ sur } X, \\ \quad\ \ \bullet \ X \text{ est } \underline{\text{compact}} \text{ et non vide,} \\ \underline{\text{alors}} \ (P_X) \text{ a (au moins) une solution.} \end{array}}$$

En dimension finie (c'est notre cas) :

- X compact $\iff$ X fermé borné.
- On peut remplacer l'hypothèse

$$X \text{ compact}$$

par

$$X \text{ fermé} \quad \text{et} \quad f \text{ coercive sur } X.$$

Unicité de solution (§ 3.1)

- Soient X un convexe de $\mathbb{E}$ et $f : X \to \mathbb{R}$.

  Définitions : f est convexe sur X si pour tout $x, y \in X$, $x \neq y$, et $t \in]0,1[$:

  $$f((1-t)x + ty) \leqslant (1-t)f(x) + tf(y).$$

  f est strictement convexe si on a inégalité stricte ci-dessus.

- Le problème à résoudre :

$$(P_X) \quad \left\{ \begin{array}{l} \min f(x) \\ x \in X. \end{array} \right.$$

$$\boxed{\begin{array}{l} \underline{\text{Si}} \ \bullet \ X \text{ est convexe,} \\ \quad\ \ \bullet \ f \text{ est } \underline{\text{strictement}} \text{ convexe sur } X, \\ \underline{\text{alors}} \ (P_X) \text{ a au plus une solution.} \end{array}}$$

Différentiabilité première (§§ C.1, C.2.1)

Soient $\mathbb{E}$ et $\mathbb{F}$ deux espaces normés, Ω un ouvert de $\mathbb{E}$ et $f : \Omega \to \mathbb{F}$.

1. Différentiabilité directionnelle suivant $h \in \mathbb{E}$:

   $$f'(x;h) := \lim_{t \to 0+} \frac{1}{t}\Big(f(x+th) - f(x)\Big) \text{ existe.}$$

2. Différentiabilité au sens de Gâteaux :
   - $f'(x;h)$ existe pour tout $h \in \mathbb{E}$ et
   - $h \mapsto f'(x;h)$ est linéaire (continue).

   On note $f'(x)$ l'application linéaire (continue).

3. Différentiabilité au sens de Fréchet : il existe $L : \mathbb{E} \to \mathbb{F}$, linéaire (continue) :

   $$f(x+h) = f(x) + Lh + o(\|h\|).$$

   On note $f'(x) := L$ (même opérateur qu'en 2).

Soit $\langle \cdot, \cdot \rangle$ un produit scalaire sur $\mathbb{E}$ et $\mathbb{F} = \mathbb{R}$. On définit le gradient de f en x comme l'unique vecteur $\nabla f(x) \in \mathbb{E}$:

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

Différentiabilité seconde (§ C.2.2)

Supposons que $f : \Omega \to \mathbb{F}$ soit 2 fois différentiable (pour une définition rigoureuse, voir le syllabus).

Propriétés :

- $f''(x) \cdot (h,k)$ est la dérivée directionnelle de $x \mapsto f'(x) \cdot h$ dans la direction k:

  $$f''(x) \cdot (h,k) = \lim_{t \to 0+} \frac{1}{t} \left(f'(x+tk) \cdot h - f'(x) \cdot h\right).$$

- l'application

  $$(h,k) \mapsto f''(x) \cdot (h,k)$$

  est bilinéaire symétrique.

Soit $\langle \cdot, \cdot \rangle$ un produit scalaire sur $\mathbb{E}$ et $\mathbb{F} = \mathbb{R}$. On définit le hessien de f en x comme l'unique opérateur linéaire symétrique $\nabla^2 f(x)$ sur $\mathbb{E}$ tel que

$$\langle \nabla^2 f(x)h, k \rangle = f''(x) \cdot (h,k), \quad \forall (h,k) \in \mathbb{E}^2.$$

II Analyse convexe

Ensemble convexe (§ 2.1)

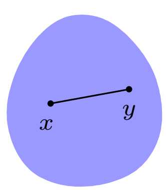
Soit $\mathbb{E}$ un espace vectoriel.

- **Dfn**. Soient x, $y \in \mathbb{E}$. Un segment de $\mathbb{E}$:

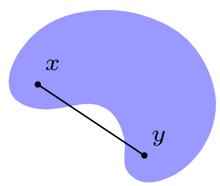
$$[x, y] := \{(1-t)x + ty : t \in [0, 1]\}.$$

- **Dfn**. Un ensemble $C \subset \mathbb{E}$ est convexe si

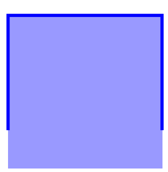
$$\forall\, x, y \in C \quad \Longrightarrow \quad [x, y] \subset C.$$



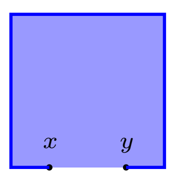
convexe



non convexe



convexe



non convexe

Aspects topologiques (§ 2.3)

- **Dfn**. L'enveloppe affine d'une partie $P \subset \mathbb{E}$ est le *plus petit espace affine* contenant P. On le note

$$\text{aff } P = \bigcap \{A : A = \text{ espace affine contenant } P\}.$$

- **Dfn**. L'intérieur relatif d'une partie $P \subset \mathbb{E}$ est son intérieur dans aff A (muni de la topologie induite de celle de $\mathbb{E}$). On le note

$$\text{intr}\, P \qquad \text{ou} \qquad P^{\ominus}.$$

Polyèdre convexe (§ 2.4)

Soient $\mathbb{E}$ et $\mathbb{F}$ des espaces vectoriels ($\mathbb{E} = \mathbb{R}^n$ et $\mathbb{F} = \mathbb{R}^m$ parfois).

- **Dfn**. Description primale d'un polyèdre convexe :

  $$P := \mathrm{co}\{x_1, \ldots, x_p\} + \mathrm{cone}\{y_1, \ldots, y_q\},$$

  où les x_i et $y_j \in \mathbb{E}$.
  Description duale d'un polyèdre convexe :

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

  où $A : \mathbb{F} \to \mathbb{F}$ linéaire et $b \in \mathbb{F}$.

- **Prop**.

  Si P polyèdre convexe et $L : \mathbb{F} \to \mathbb{F}$ linéaire, alors $L(P)$ polyèdre convexe.

Projection sur un convexe fermé (§ 2.5.2)

$\mathbb{E}$ muni d'un produit scalaire $\langle \cdot, \cdot \rangle = \| \cdot \|^2$.

Si $C \subset \mathbb{E}$ convexe fermé non vide et $x \in \mathbb{E}$, alors le problème

$$\min \{\|y - x\| : y \in C\} \qquad (1)$$

a une unique solution.

- **Dfn** : l'unique solution de (1) est appelée projection/projeté de x sur C et est notée $P_C x$.

- **Prop** : Soit $\bar{x} \in C$. Alors

$$\begin{aligned} \bar{x} = P_C x \quad &\Longleftrightarrow \quad \forall y \in C, \quad \langle y - \bar{x}, \bar{x} - x \rangle \geqslant 0 \\ &\Longleftrightarrow \quad \forall y \in C, \quad \langle y - \bar{x}, y - x \rangle \geqslant 0 \\ &\Longleftrightarrow \quad \forall y \in C, \quad \langle y - x, \bar{x} - x \rangle \geqslant \|\bar{x} - x\|^2. \end{aligned}$$

Séparation des convexes (§ 2.5.4)

$\mathbb{E}$ muni d'un produit scalaire $\langle\cdot,\cdot\rangle$.

- **Dfn** : On peut séparer $C_1, C_2 \subset \mathbb{E}$ s'il existe $\xi \in \mathbb{E}$ non nul tel que
$$\sup_{x_1 \in C_1} \langle \xi, x_1 \rangle \leqslant \inf_{x_2 \in C_2} \langle \xi, x_2 \rangle \,.$$
La séparation est stricte si l'inégalité ci-dessus est stricte (alors ξ est nécessairement non nul).
- **Théor (Hahn-Banach)** :

Si • C_1 et C_2 convexes, non vides, disjoints,
• $\dim \mathbb{E} < \infty$,
alors on peut séparer C_1 et C_2.

Si • C_1 et C_2 convexes, non vides, disjoints,
• C_1 ou C_2 est d'intérieur non vide,
alors on peut séparer C_1 et C_2.

Si • C_1 et C_2 convexes, non vides, disjoints,
• l'un est fermé, l'autre est compact,
alors on peut séparer C_1 et C_2 strictement.

Cône dual (§ 2.5.6)

$\mathbb{E}$ muni d'un produit scalaire $\langle\cdot,\cdot\rangle$.

- **Dfn** : Le cône dual de $P \subset \mathbb{E}$ est défini par
$$P^+ := \{y \in \mathbb{E} : \langle y, x\rangle \geqslant 0, \ \forall x \in P\}.$$
C'est un cône, convexe, fermé, non vide.
- **Lemme de Farkas (généralisé)**

Si • $\mathbb{E}$ et $\mathbb{F}$ deux espaces euclidiens,
• K un cône convexe $\neq \varnothing$ de $\mathbb{E}$,
• $A : \mathbb{E} \to \mathbb{F}$ linéaire,
alors $\{y \in \mathbb{F} : A^* y \in K^+\}^+ = \overline{A(K)}$.

- **Cas particulier** : Soit A une matrice. Alors
$$\begin{aligned} \{Ax : x \geqslant 0\} &= \text{cône, convexe, fermé, } \neq \varnothing \\ \{y : A^\top y \geqslant 0\}^+ &= \{Ax : x \geqslant 0\}. \end{aligned}$$
$(\cdot)^+$ = dual pour le produit scalaire euclidien.
[c'est une généralisation de $N(A^\top)^\perp = R(A)$]

Fonction convexe (§ 3.1)

Soient $\mathbb{E}$ un espace vectoriel et $f : \mathbb{E} \to \overline{\mathbb{R}}$.

- **Dfn**. Le domaine de f est l'ensemble

$$\operatorname{dom} f := \{x \in \mathbb{E} : f(x) < +\infty\}.$$

  On peut avoir $f(x) = -\infty$ pour $x \in \operatorname{dom} f$.

- **Dfn**. L'épigraphe de f est l'ensemble

$$\operatorname{epi} f := \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} : f(x) \leqslant \alpha\}.$$

- **Dfn**. f est convexe $\iff$ epi f est convexe.

f est convexe $\iff$ $\forall\, x, y \in \operatorname{dom} f, \forall\, t \in \,]0, 1[\,:$

$$f\Big((1-t)x + ty\Big) \leqslant (1-t)f(x) + tf(y).$$

Différentiabilité directionnelle (§ 3.3.2)

Soient $f \in \operatorname{Conv}(\mathbb{E})$, $x \in \operatorname{dom} f$ et $d \in \mathbb{E}$.

1) $t \in \,]0, +\infty[\, \mapsto \frac{f(x+td)-f(x)}{t}$ est croissante,
2) $f'(x; d)$ existe dans $\overline{\mathbb{R}}$,
3) $f'(x; d) = +\infty \iff f(x + td) = +\infty, \forall\, t > 0$,
4) $f'(x; -d) \geqslant -f'(x; d)$,
5) $f'(x; \cdot)$ est convexe,
6) $x \in (\operatorname{dom} f)^{\ominus} \implies \begin{cases} f'(x; \cdot) \in \mathbb{R}, \\ f'(x; \cdot) \text{ Lipschitz}, \\ f'(x; \cdot) \in \overline{\operatorname{Conv}}(\mathbb{E}). \end{cases}$

Enveloppe supérieure (§ 3.4.2)

- Enveloppe supérieure d'une famille de $f_i : \mathbb{E} \to \overline{\mathbb{R}}$, $i \in I$ (quelconque) :

$$\left(\sup_{i\in I} f_i\right)(x) := \sup_{i\in I} \Big(f_i(x)\Big).$$

> - $\operatorname{epi}(\sup_{i\in I} f_i) = \bigcap_{i\in I} (\operatorname{epi} f_i)$.
> - f_i convexes $\Longrightarrow$ $\sup_{i\in I} f_i$ convexe.
> - f_i fermées $\Longrightarrow$ $\sup_{i\in I} f_i$ fermée.

Reconnaître une fonction convexe par ses dérivées (§ 3.3.3)

Soient X un convexe de $\mathbb{E}$ et $f : X \to \mathbb{R}$.

- Si f est 1 fois dérivable et X ouvert

> Les propriétés suivantes sont équivalentes :
>
> – f est convexe sur X
> [resp. strictement convexe],
>
> – $\forall x, y \in X, x \neq y$:
>
> $$f(y) \geqslant f(x) + f'(x) \cdot (y - x)$$
>
> [resp. $>$],
>
> – $\forall x, y \in X, x \neq y$:
>
> $$(f'(y) - f'(x)) \cdot (y - x) \geqslant 0$$
>
> [resp. $>$].

- Si f est 2 fois dérivable et X ouvert :

  - f est convexe sur $X \Longleftrightarrow$
    $\forall x \in X, \forall h \in \mathbb{E}$, $f''(x) \cdot h^2 \geqslant 0$,
  - f est strictement convexe sur $X \Longleftarrow$
    $\forall x \in X, \forall h \in \mathbb{E}$ non nul, $f''(x) \cdot h^2 > 0$.

  Contre-exemple: $f(x) = x^4$.

III Conditions d'optimalité (CO)

Le problème à résoudre :

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

où $X \subset E$ (espace euclidien, produit scalaire $\langle \cdot, \cdot \rangle$).

- **But**. L'optimalité s'exprime par un nombre infini de conditions $x_* \in X$ et $\forall\, x \in X : f(x_*) \leqslant f(x)$. Peut-on exprimer cela avec un nombre fini de conditions ?
- **Qu'est-ce ?** Des $=$ et $\leqslant$ décrivant les solutions de (P_X).
- **Utilité** des CO :
  - donner des renseignements sur (P_X),
  - vérifier qu'un point est solution,
  - calculer la solution analytiquement (parfois),
  - définir des algorithmes de résolution.
- Il y a des CO nécessaires (notées CN) et des CO suffisantes (notées CS).
- Il y a des CO du 1$^{\text{er}}$ ordre (CN1, CS1) et des CO du 2$^{\text{ième}}$ ordre (CN2, CS2).

CO sans contrainte (rappel, § 4.2)

Le problème à résoudre :

$$\begin{cases} \min f(x) \\ x \in \mathbb{E}. \end{cases}$$

On note $\nabla f(x)$ et $\nabla^2 f(x)$ les gradient et hessien de f en x pour $\langle \cdot, \cdot \rangle$.

- **CN1** :

  $$x_* \text{ min local} \quad \Longrightarrow \quad \nabla f(x_*) = 0.$$

  (Si f est convexe, c'est une **CS1** globale.)

- **CN2** :

  $$x_* \text{ min local} \quad \Longrightarrow \quad \begin{cases} \nabla f(x_*) = 0 \\ \nabla^2 f(x_*) \succcurlyeq 0. \end{cases}$$

- **CS2** pour un minimum local strict :

  $$\begin{cases} \nabla f(x_*) = 0 \\ \nabla^2 f(x_*) \succ 0 \end{cases} \quad \Longrightarrow \quad x_* \text{ min local strict}.$$

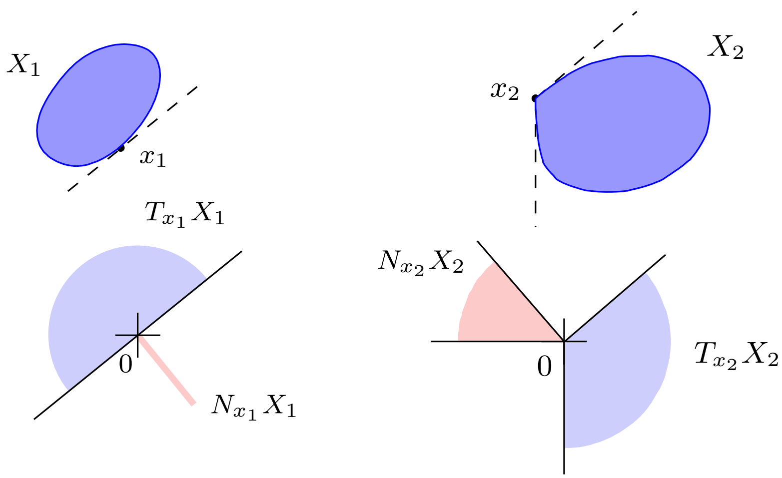
CN1 de Peano-Kantorovitch (§ 4.1)

Le problème à résoudre :

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

Dfn. Cône tangent : $d \in \mathrm{T}_x X \iff$

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



Prop

- $\mathrm{T}_x X$ est fermé.
- X convexe et $x \in X \implies \mathrm{T}_x X$ convexe.

Conditions d'optimalité

- **CN1 de Peano-Kantorovitch**. On exprime plus ou moins le fait que f croît si on se déplace "vers l'intérieur" de X :

$$f'(x_*) \cdot d \geqslant 0, \quad \forall d \in \mathrm{T}_{x_*} X, \tag{2}$$

  où $\mathrm{T}_{x_*} X$ est le cône tangent à X en x_*, ce qui s'écrit aussi

$$\nabla f(x_*) \in (\mathrm{T}_{x_*} X)^+, \tag{3}$$

- **CN1**. Lorsque X est convexe, la relation (2) se simplifie en :

$$f'(x_*) \cdot (x - x_*) \geqslant 0, \quad \forall x \in X. \tag{4}$$

- **CS1**. Si X est convexe, f est convexe et (4), alors x_* est un minimum global.

CO avec contraintes d'$=$ (§ 4.3)

Le problème en $x \in \mathbb{E}$ (e.v. euclidien) à résoudre :

$$(P_E) \quad \begin{cases} \min f(x) \\ c(x) = 0 \in \mathbb{F} \quad \text{(e.v. euclidien)}. \end{cases}$$

Ensemble admissible noté X_E.

Le lagrangien du problème :

$$\ell(x, \lambda) = f(x) + \langle \lambda, c(x) \rangle \, .$$

- **CN1** : si $c'(x_*)$ est surjective, il existe $\lambda_* \in \mathbb{F}$, unique, tel que

$$\begin{cases} \nabla_x \ell(x_*, \lambda_*) = 0 \\ c(x_*) = 0. \end{cases} \tag{5}$$

  (Si c affine, λ_* existe, pas néc. unique.)
  (Si f est convexe et X_E est convexe, ce sont des **CS1** globales.)

Si $\mathbb{F} = \mathbb{R}^m$, la première condition de (5) s'écrit

$$\nabla f(x_*) + \sum_{i=1}^{m} (\lambda_*)_i \nabla c_i(x_*) = 0.$$

- **CN2** : si x_* minimum local et $c'(x_*)$ est surjective, alors il existe λ_* tel que

$$\begin{cases} \nabla_x \ell(x_*, \lambda_*) = 0 \\ c(x_*) = 0 \\ \nabla^2_{xx} \ell(x_*, \lambda_*) \succcurlyeq 0 \text{ sur } N(c'(x_*)). \end{cases}$$

- **CS2** : si (x_*, λ_*) vérifie

$$\begin{cases} \nabla_x \ell(x_*, \lambda_*) = 0 \\ c(x_*) = 0 \\ \nabla^2_{xx} \ell(x_*, \lambda_*) \succ 0 \text{ sur } N(c'(x_*)), \end{cases}$$

  alors x_* est un minimum local strict.

CO avec contraintes d'$=$ et d'$\leqslant$ (§ 4.4)

Le problème à résoudre en $x \in \mathbb{E}$:

$$(P_{EI}) \quad \begin{cases} \min f(x) \\ c_E(x) = 0 \in \mathbb{R}^{m_E} \\ c_I(x) \leqslant 0 \in \mathbb{R}^{m_I}. \end{cases}$$

Ensemble admissible noté X_{EI}.

Le lagrangien du problème ($c := (c_E, c_I)$) :

$$\ell(x, \lambda) = f(x) + \lambda^\top c(x).$$

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

- **CN1** : si les contraintes sont qualifiées en x_*, il existe $\lambda_* \in \mathbb{R}^m$ tel que

$$(\mathrm{KKT}) \quad \begin{cases} \nabla_x \ell(x_*, \lambda_*) = 0 \\ c_E(x_*) = 0 \\ 0 \leqslant (\lambda_*)_I \perp c_I(x_*) \leqslant 0. \end{cases}$$

(Si f est convexe et X_{EI} est convexe, ce sont des **CS1** globales.)

Qualification des contraintes (§ 4.4.3)

- **Dfn** : on dit que les contraintes de (P_{EI}) sont qualifiées en x si

$$\mathrm{T}_x X = \mathrm{T}'_x X, \tag{6}$$

où

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

On a toujours : $\mathrm{T}_x X \subset \mathrm{T}'_x X$.

- **Conditions suffisantes** de qualification des contraintes.
Régularité + l'une des conditions suivantes :

(QC-A) $c_{E \cup I^0(x)}$ est affine dans un voisinage de x.

(QC-S) c_E est *affine*,
$c_{I^0(x)}$ *convexe*,
$\exists\, \check{x} \in X$ tel que $c_{I^0(x)}(\check{x}) < 0$.

(QC-IL) les gradients $\{\nabla c_i(x)\}_{i \in E \cup I^0(x)}$ sont linéairement indépendants.

(QC-MF) $\sum_{i \in E \cup I^0(x)} \alpha_i \nabla c_i(x) = 0$ et $\alpha_{I^0(x)} \geqslant 0 \implies \alpha_{E \cup I^0(x)} = 0$.

(QC-MF') $c'_E(x)$ surjective et $\exists\, d \in \mathbb{E}$ tel que $c'_E(x) \cdot d = 0$ et $c'_{I^0_x}(x) \cdot d < 0$.

Démarche suivie pour obtenir (KKT)

- On part de (2)
  [i.e., f croît de x_* vers l'intérieur de X].

- On suppose que les contraintes sont qualifiées en x_* (on a (6) avec $x = x_*$). Dès lors

$$\nabla f(x_*) \in \left(\mathrm{T}'_{x_*} X\right)^+. \qquad (7)$$

- <u>Lemme</u> <u>de</u> <u>Farkas</u> :
  Données : $A : \mathbb{E} \to \mathbb{F}$ linéaire et K cône de $\mathbb{E}$.

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

  C'est une généralisation de $N(A)^\perp = R(A^\top)$.

- Le lemme de Farkas permet d'exprimer (7) autrement : $\exists \lambda_* \in \mathbb{R}^m$ tel que l'on ait (KKT).

<u>Signification des multiplicateurs optimaux (§ 4.6.1)</u>

- Problème perturbé : pour $p \in \mathbb{R}^m$, on définit

$$(P^p_{EI}) \quad \begin{cases} \min\ f(x) \\ c_E(x) + p_E = 0 \\ c_I(x) + p_I \leqslant 0. \end{cases}$$

- **Dfn**. La fonction valeur associée à (P^p_{EI}) est $v : p \in \mathbb{R}^m \to \overline{\mathbb{R}}$ définie par

$$v(p) = \inf_{x \in X^p} f(x),$$

  où X^p est l'ensemble admissible de (P^p_{EI}).

$$(P_{EI}) \text{ convexe } \Longrightarrow v \text{ convexe}.$$

- Cas différentiable régulier.

<u>Si</u> • (x_*, λ_*) solution PD de (P_{EI}), · $(\bar{x}(p), \bar{\lambda}(p))$ solution PD de (P^p_{EI}), · $p \mapsto \bar{x}(p)$ différentiable en 0, $\bar{x}(0) = x_*$, · $p \mapsto \bar{\lambda}(p)$ continue en 0, $\bar{\lambda}(0) = \lambda_*$, <u>alors</u> $\lambda_* = \nabla v(0) = \nabla(f \circ \bar{x})(0)$.

- On note
$$\Lambda := \{\lambda \in \mathbb{R}^m : \lambda_I \geqslant 0\}.$$
- **Dfn**. On dit que $(x_*, \lambda_*) \in \mathbb{R}^n \times \Lambda$ est un point-selle de ℓ sur $\mathbb{R}^n \times \Lambda$, si $\forall (x, \lambda) \in \mathbb{R}^n \times \Lambda$:
$$\ell(x_*, \lambda) \leqslant \ell(x_*, \lambda_*) \leqslant \ell(x, \lambda_*).$$
- Cas convexe non différentiable.

<u>Si</u> • x_* est solution de (P_{EI}), • $v \in \mathrm{Conv}(\mathbb{R}^m)$, <u>alors</u> $\partial v(0)$ $= \{\lambda_* : (x_*, \lambda_*)$ est point-selle de ℓ sur $\mathbb{R}^m \times \Lambda\}$.

<u>Remarque</u> : Ci-dessus, $\partial v(0)$ peut être vide ! Avec qualification de Slater : $\partial v(0) \neq \varnothing$.

- CN et CS d'existence de solution PD globale.

<u>CN d'optimalité</u> (cas convexe non diff.).

<u>Si</u> • (P_{EI}) convexe (avec f et c finies), • (Slater) : c'_E surjective, $\exists \hat{x} \in X$ t.q. $c_I(\hat{x}) < 0$, • x_* solution de (P_{EI}), <u>alors</u> 1) v est loc. lipschitzienne dans un vois. de 0, 2) $\partial v(0) \neq \varnothing$.

<u>CS d'optimalité globale</u>.

Peu de chance d'être applicable si (P_{EI}) non convexe.

<u>Si</u> • $(x_*, \lambda_*) \in \mathbb{R}^n \times \Lambda$ est un point-selle de ℓ sur $\mathbb{R}^n \times \Lambda$, <u>alors</u> x_* solution (globale) de (P_{EI}).

IV $\boxed{\textbf{Méthodes à directions de descente}}$

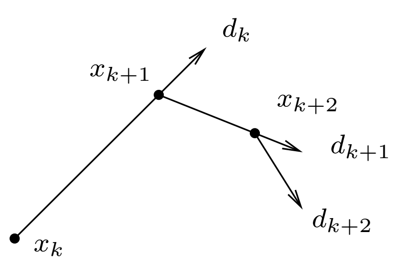
Schéma des algorithmes (§ 6.1)

- **Dfn** : d est direction de descente de f en x si

$$f'(x) \cdot d < 0.$$

$\Longrightarrow$ f décroît en x le long de d.

- **Algorithme à directions de descente** : il génère une suite $\{x_k\} \subset \mathbb{E}$ comme suit
  - *Calcul d'une direction de descente* d_k;
  - *Recherche linéaire* : on détermine un pas $\alpha_k > 0$ le long de d_k;
  - *Nouvel itéré* : $x_{k+1} := x_k + \alpha_k d_k$.



Exemples d'algorithmes à DD (§ 6.2)

On note $g_k := \nabla f(x_k)$.

- **Algorithme du gradient**.

$$d_k = -g_k.$$

- **Algorithme du gradient conjugué**.

$$d_k = \begin{cases} -g_1 & \text{si} \quad k = 1 \\ -g_k + \beta_k d_{k-1} & \text{si} \quad k \geqslant 2. \end{cases}$$

- **Algorithme de Newton**.

$$d_k = -\nabla^2 f(x_k)^{-1} g_k.$$

- **Algorithme de quasi-Newton**.

$$d_k = -M_k^{-1} g_k.$$

- **Algorithme de Gauss-Newton** pour $f(x) = \frac{1}{2}\|r(x)\|_2^2$ et $J(x) := r'(x)$ injective :

$$d_k = -\left(J(x_k)^* J(x_k)\right)^{-1} J(x_k)^* r(x_k).$$

La recherche linéaire (§ 6.3)

Deux techniques souvent utilisées : RL d'Armijo et RL de Wolfe.

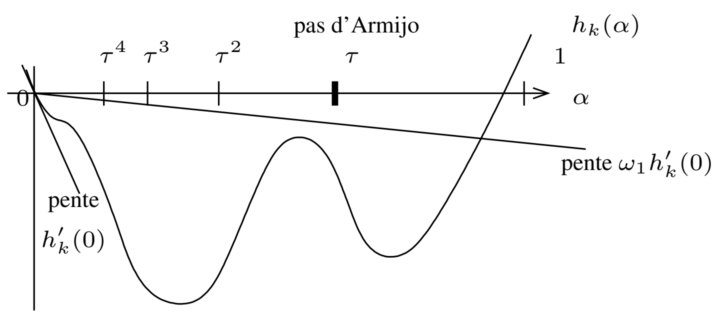
Soient d_k une direction de descente et

$$h_k(\alpha) := f(x_k + \alpha d_k).$$

- **RL d'Armijo** $\quad (0 < \omega_1 < \frac{1}{2}, \quad 0 < \tau < 1)$

$$h_k(\alpha_k) \leqslant h(0) + \omega_1 \alpha_k h_k'(0), \quad \alpha_k = \tau^{i_k},$$

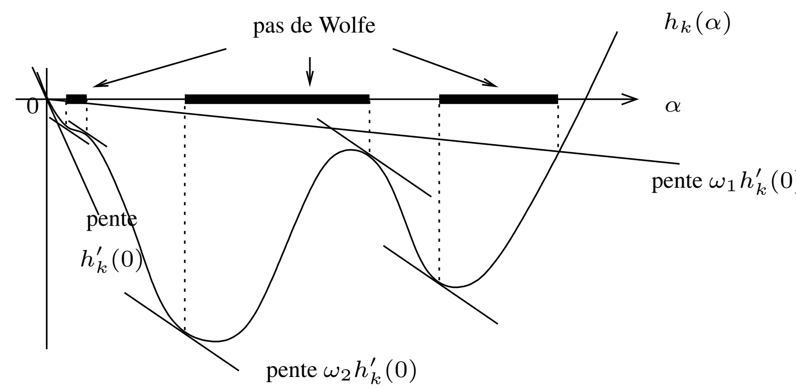
où i_k est le plus petit dans $\{0, 1, 2, \ldots\}$.



Valeurs typiques : $\omega_1 = 10^{-4}$ et $\tau = \frac{1}{2}$.

- **RL de Wolfe** $\quad (0 < \omega_1 < \frac{1}{2}, \quad \omega_1 < \omega_2 < 1)$

$$h_k(\alpha_k) \leqslant h(0) + \omega_1 \alpha_k h_k'(0),$$
$$h_k'(\alpha_k) \geqslant \omega_2 h_k'(0).$$



Valeurs typiques : $\omega_1 = 10^{-4}$ et $\omega_2 = 0.99$.

Convergence avec la RL de Wolfe (§ 6.3.4)

- **Dfn** :

$$\cos \theta_k := \frac{-\langle g_k, d_k \rangle}{\|g_k\| \, \|d_k\|}.$$

- **Théor** :

Si • $f \in C^{1,1}$,
• RL de Wolfe
• $\exists\, C, \quad \forall\, k \geqslant 0, \quad f(x_k) \geqslant C$,
alors

$$\sum_{k \geqslant 0} \|g_k\|^2 \cos^2 \theta_k < +\infty.$$

- **Convergence** :
  - ∗ Algo du gradient : $\theta_k = 0$, donc $g_k \to 0$.
  - ∗ Plus généralement : $\cos \theta_k \geqslant c > 0$, donc $g_k \to 0$.

V Méthodes newtoniennes pour équations

Vitesse de convergence des suites (§ 5.1.1)

Soit $\{x_k\}$ une suite convergeant vers $x_* \in \mathbb{E}$.
On suppose que $x_k \neq x_*$, pour tout $k \geqslant 1$.

- Convergence linéaire : il existe une norme $\|\cdot\|$, un indice k_0 et $\tau \in [0,1[$ tels que $\forall\, k \geqslant k_0$:
$$\frac{\|x_{k+1}-x_*\|}{\|x_k-x_*\|} \leqslant \tau.$$
- Convergence superlinéaire :
$$\frac{\|x_{k+1}-x_*\|}{\|x_k-x_*\|} \to 0.$$
- Convergence quadratique : il existe une constante $C>0$ telle que $\forall\, k \geqslant 1$:
$$\frac{\|x_{k+1}-x_*\|}{\|x_k-x_*\|^2} \leqslant C.$$

$\sigma_k =$ nombre de chiffres significatifs corrects.

	superlinéaire		quadratique	
k	x_k	σ_k	x_k	σ_k
1	2.00000000000000	0	2.00000000000000	0
2	1.50000000000000	0	0.86666666666667	1
3	0.61224489795918	1	-0.32323745064862	1
4	-0.16202797536640	1	-0.92578663808031	1
5	-0.92209500449059	1	-0.82332584261905	2
6	-0.78540447895661	1	-0.81774699537697	5
7	-0.81609056319699	3	-0.81773167400186	9
8	-0.81775774021392	5	-0.81773167388682	15
9	-0.81773165292101	8		
10	-0.81773167388656	13		
11	-0.81773167388682	15		

Linéaire $\Longrightarrow$ $\begin{cases} \exists \underline{\sigma} > 0,\ \forall k \text{ grand :} \\ \sigma_{k+1} - \sigma_k \geqslant \underline{\sigma}. \end{cases}$

Superlinéaire $\Longrightarrow$ $\sigma_{k+1} - \sigma_k \to \infty.$

Quadratique $\Longrightarrow$ $\liminf\limits_{k\to\infty} \dfrac{\sigma_{k+1}}{\sigma_k} \geqslant 2.$

Algorithme de Newton pour systèmes non linéaires

(§ 9.1.1)

Soit $F : \mathbb{E} \to \mathbb{F}$, avec $\dim \mathbb{E} = \dim \mathbb{F} < \infty$. On cherche à résoudre en x :

$$F(x) = 0.$$

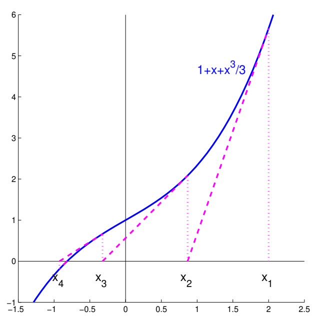
- **Algorithme de Newton.** De x_k à x_{k+1} :
  - *Résoudre* en d_k l'équation de Newton :

$$F'(x_k)\, d_k = -F(x_k). \tag{8}$$

  - *Nouvel itéré* :

$$x_{k+1} = x_k + d_k.$$

- **Exemple 1D.**



- **Propriétés de l'algorithme de Newton.**

⊕⊕ Convergence quadratique *locale* :

Si • x_* vérifie $F(x_*) = 0$,
• F est $C^{1,1}$ dans un voisinage de x_*,
• $F'(x_*)$ est inversible,

alors il existe un voisinage V de x_* tel que si $x_1 \in V$, l'algorithme de Newton (8) est bien défini et génère une suite $\{x_k\} \subset V$ qui converge *quadratiquement* vers x_*.

⊖ En général ne converge pas si x_1 n'est pas proche d'une solution.

⊖ Il faut calculer les dérivées premières de F.

Globalisation de l'algorithme de Newton par recherche linéaire (§ 9.3.1)

- **Dfn** : « globaliser » $=$ forcer la convergence lorsque x_1 n'est pas voisin d'une solution.

- Une solution miracle ?

  Si $F(x) \neq 0$, la direction de Newton en x,
  $$d^{\mathrm{N}} = -F'(x)^{-1}F(x),$$
  est une direction de descente de
  $$f(x) = \frac{1}{2}\|F(x)\|_2^2.$$
  On a $f'(x) \cdot d^{\mathrm{N}} = -\|F(x)\|_2^2 < 0$.

- RL sur f le long de d^{N} : $x_+ := x + \alpha d^{\mathrm{N}}$, avec $\alpha > 0$ tel que (ici $\omega_1 \in \left]0, \frac{1}{2}\right[$)
  $$f(x_+) \leqslant f(x) + \alpha\,\omega_1\, f'(x) \cdot d^{\mathrm{N}}.$$

- Un résultat de convergence :

  Si $\{F'(x_k)\}$ et $\{F'(x_k)^{-1}\}$ sont *bornées*,
  alors l'algorithme de Newton avec une RL « convenable » converge vers un point stationnaire de f : $\nabla f(x_k) \to 0$.

- Cette approche ne converge pas toujours !
  C'est la raison pour laquelle on a inventé la technique des régions de confiance (voir un cours plus avancé).

V'

Méthodes newtoniennes en optimisation (§ 9.1.2)

Soit le problème

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

- On se déclare satisfait avec x_* vérifiant

$$\nabla f(x_*) = 0.$$

La relation $F = \nabla f$ permet d'adapter l'algorithme de Newton ($F'(x) = \nabla^2 f(x)$ est symétrique).

- **Algorithme de Newton.** De x_k à x_{k+1} :

  – *Résoudre* en d_k l'équation de Newton :

$$\nabla^2 f(x_k)\, d_k = -\nabla f(x_k). \tag{9}$$

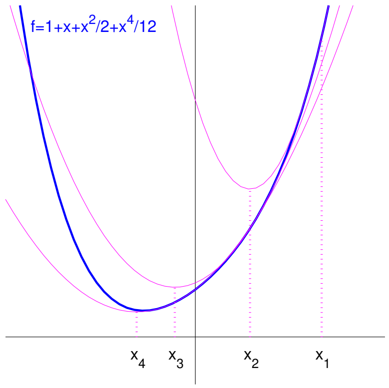
  – *Nouvel itéré* :

$$x_{k+1} = x_k + d_k.$$

- **Le problème quadratique osculateur.**

Le pas de Newton d_k est aussi un point *stationnaire* du problème quadratique

$$\min_{d\in\mathbb{E}} \left(f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla^2 f(x_k) d \right).$$



- **Propriétés de l'algorithme de Newton**.

$\oplus\oplus$ Convergence quadratique *locale* :

> Si • x_* vérifie $\nabla f(x_*) = 0$,
> • f est $C^{2,1}$ dans un voisinage de x_*,
> • $\nabla^2 f(x_*)$ est inversible,
>
> alors il existe un voisinage V de x_* tel que si $x_1 \in V$, l'algorithme de Newton est bien défini et génère une suite $\{x_k\} \subset V$ qui converge *quadratiquement* vers x_*.

$\ominus$ En général ne converge pas si x_1 n'est pas proche d'un point stationnaire.

$\ominus$ Pas de distinction entre min, max, point stationnaire.

$\ominus$ Les directions ne sont pas nécessairement de descente.

$\ominus$ Il faut calculer les dérivées secondes de f.

Algorithmes de quasi-Newton (§§ 10.1.1, 10.2.1, 10.2.2)

Soit le problème

$$\min_{x \in \mathbb{R}^n} f(x).$$

- Les **algorithmes de qN** génèrent 2 suites : $\{x_k\} \subset \mathbb{R}^n$ et $\{M_k\} \subset \mathbb{R}^{n\times n}$ sym. dfn. pos.

  1) $d_k := -M_k^{-1} g_k$;
  2) $\alpha_k > 0$ par recherche linéaire;
  3) $x_{k+1} := x_k + \alpha_k d_k$;
  4) $M_{k+1} := U(M_k, y_k, s_k)$,
     où $y_k := g_{k+1} - g_k$ et $s_k := x_{k+1} - x_k$.

- **Mise à jour de** M_k. On cherche à ce que M_{k+1} soit proche de M_k (stabilité), tout en vérifiant :

  – l'équation de qN : $y_k = M_{k+1} s_k$;
  – la symétrie : $M_{k+1}^\top = M_{k+1}$;
  – la définie positivité : $M_{k+1} \succ 0$.

  Cela conduit à la formule de BFGS.

$$M_{k+1} = M_k + \frac{y_k y_k^\top}{y_k^\top s_k} - \frac{M_k s_k s_k^\top M_k}{s_k^\top M_k s_k}.$$

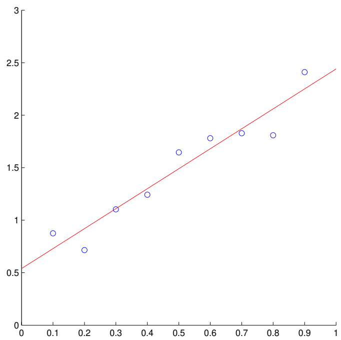
VI Problèmes de moindres-carrés

- Ce sont des problèmes de la forme

$$\min_{x \in \mathbb{R}^n} \|F(x)\|,$$

où $F : \mathbb{R}^n \to \mathbb{R}^m$. En général $m \gg n$.

- **Exemple :** la régression linéaire.



Moindres-carrés linéaire (§ 17.1)

- **Problème** : on cherche une solution de

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2, \tag{10}$$

où A est $m \times n$ et $b \in \mathbb{R}^m$.

- Équation normale : x est solution ssi

$$A^\top A x = A^\top b. \tag{11}$$

- **Existence de solution** :
  - Le problème (10) a <u>toujours</u> une solution.
  - Solution unique $\iff$ A est injective.
  - Ensemble des solutions $= x_p + N(A)$.

- **Méthodes numériques** :
  - Factorisation de Cholesky de $A^\top A$.
  - GC sur (11).
  - Factorisation QR de A.
  - Factorisation SVD de A.

Moindres-carrés non linéaire (§ 17.3)

- **Problème** : on cherche une solution de

$$\min_{x\in\mathbb{R}^n} \left(f(x) := \frac{1}{2}\|r(x)\|_2^2\right), \tag{12}$$

où $r : \mathbb{R}^n \to \mathbb{R}^m$ est non linéaire (les résidus).
Jacobienne $J \equiv J(x) \equiv r'(x)$, qui est $m \times n$.

- **Algorithme de Gauss-Newton** : RL le long de

$$d_k^{\text{GN}} \in \arg\min_{d\in\mathbb{R}^n} \frac{1}{2}\|r(x_k) + J(x_k)d\|_2^2.$$

On a $f'(x_k) \cdot d_k^{\text{GN}} \leqslant 0$ (< 0 si $\nabla f(x_k) \neq 0$).

Résultat de convergence :

Si $\{J(x_k)\}$ est *bornée* et *unif. injective*, i.e.,

$$\exists C > 0,\ \forall k \geqslant 1,\ \forall v \in \mathbb{R}^n :$$
$$C\|v\|_2 \leqslant \|J(x_k)v\|_2 \leqslant C^{-1}\|v\|_2,$$

alors l'algorithme de Gauss-Newton avec RL converge vers un point stationnaire de f (c'est-à-dire $J(x_k)^\top r(x_k) \to 0$).

- **Algorithme de Levenberg-Marquardt** (révisé) : RC avec le modèle quadratique

$$\varphi_k(s) := \frac{1}{2}\|r(x_k) + J(x_k)s\|_2^2.$$

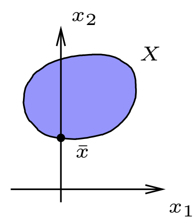
Résultat de convergence :

Si $\{J(x_k)\}$ est *bornée*,
alors l'algorithme de Levenberg-Marquardt avec RC converge vers un point stationnaire de f (c'est-à-dire $J(x_k)^\top r(x_k) \to 0$).

VII $\boxed{\textbf{Dualité (§ 13)}}$

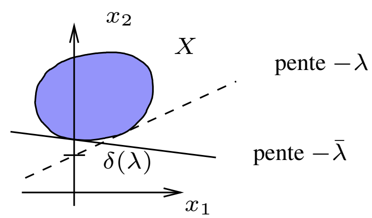
- **Un premier problème** :

$$(P)\quad \begin{cases} \inf\ x_2 \\ x \in X \\ x_1 = 0. \end{cases}$$



- **Un second problème** :

$$(D)\quad \sup_{\lambda \in \mathbb{R}} \ \delta(\lambda)$$



- (P) et (D) sont duaux l'un de l'autre.
- **Intérêts de la dualité** :
  - obtenir des propriétés sur un problème à partir des propriétés d'un pbl dual (e.g., une borne sur la valeur optimale);
  - construire des pbls duaux équivalents au pbl primal, mais plus faciles à résoudre;
  - algorithmique : recherche de point-selle, du multiplicateur optimal.

Dualité min-max (§ 13.1)

Soient X un ensemble, $f : X \to \overline{\mathbb{R}}$ et

$$(P) \qquad \inf_{x \in X} f(x).$$

- **Récriture du problème primal.**
  On suppose que

$$f(x) = \sup_{y \in Y} \varphi(x, y),$$

  où $\varphi : X \times Y \to \overline{\mathbb{R}}$. Donc

$$(P) \qquad \inf_{x \in X} \sup_{y \in Y} \varphi(x, y) = \mathrm{val}(P).$$

- **Le problème dual** : on inverse l'inf et le sup

$$(D) \qquad \sup_{y \in Y} \inf_{x \in X} \varphi(x, y) = \mathrm{val}(D).$$

  On peut aussi l'écrire $\inf_{y \in Y} \ \delta(y)$, où

$$\delta(y) := -\inf_{x \in X} \ \varphi(x, y). \qquad (13)$$

  - $\delta \equiv$ fonction duale,
  - $(13) \ \equiv$ problème interne en $y \in Y$.

Liens entre (P) et (D)

- **Dualité faible** :

$$\mathrm{val}(D) \leqslant \mathrm{val}(P).$$

Saut de dualité $:= \mathrm{val}(P) - \mathrm{val}(D) \geqslant 0.$

- **Dfn** : On dit que $(\bar{x}, \bar{y})$ est un point-selle de φ sur $X \times Y$, si $\forall x \in X$ et $\forall y \in Y$

$$\varphi(\bar{x}, y) \leqslant \varphi(\bar{x}, \bar{y}) \leqslant \varphi(x, \bar{y}).$$

- **Théor** : $(\bar{x}, \bar{y})$ est un point-selle de φ sur $X \times Y$ SSI
  1) $\bar{x}$ est solution de (P),
  2) $\bar{y}$ est solution de (D),
  3) il n'y a pas de saut de dualité.

- **Coroll** : Si φ a un point-selle et $\bar{y} \in \mathrm{Sol}(D)$:

$$\varnothing \neq \mathrm{Sol}(P) \subset \arg\min_{x \in X} \varphi(x, \bar{y}).$$

Dualisation de contraintes fonctionnelles (§ 13.5)

On cherche à écrire un problème dual du problème d'optimisation avec contraintes :

$$(P_{X,EI}) \begin{cases} \min\ f(x) \\ x \in X \\ c_E(x) = 0 \\ c_I(x) \leqslant 0. \end{cases}$$

où $X \subset \mathbb{E}$, sans qu'il y ait de saut de dualité.

Dualité lagrangienne
(problèmes convexes)

On prend pour φ, le lagrangien (ici $y \equiv \lambda$)

$$\varphi(x,\lambda) \equiv \ell(x,\lambda) = f(x) + \lambda^\top c(x),$$

$X \subset \mathbb{E}$ et $Y := \mathbb{R}^{m_E} \times \mathbb{R}^{m_I}_+$.

- **Problème primal** :

$$(P_{X,EI}) \equiv \inf_{x\in X} \sup_{\lambda\in Y} \ \ell(x,\lambda).$$

- **Problème dual** :

$$\sup_{\lambda\in Y} \inf_{x\in X} \ \ell(x,\lambda).$$

- **Résultat de dualisation** :

Si
- $X = \mathbb{E}$,
- $(P_{X,EI})$ est « convexe » (i.e., f et c_I convexes et c_E affine),
- $(\bar{x},\bar{\lambda})$ vérifie (KKT),

alors $(\bar{x},\bar{\lambda})$ est point-selle de ℓ sur $\mathbb{E} \times Y$.

- **Relaxation lagrangienne (Uzawa)**

On passe de λ_k à λ_{k+1} par :

1. $x_k \in \arg\min_{x\in\mathbb{E}} \ \ell(x,\lambda_k)$,
2. arrêt si (x_k,λ_k) est satisfaisant,
3. $\lambda_{k+1} = P_Y\,[\lambda_k + \alpha_k c(x_k)]$.

Explications de la formule de mise à jour de λ_k (algorithme du gradient avec projection) :

- P_Y est le projecteur orthogonal sur Y (permet de maintenir les λ_k dans Y),
- $\alpha_k > 0$ est déterminé de manière à faire croître δ,
- $-c(x_k)$ est un sous-gradient de δ.

Résultat de convergence :

Si
- f fortement convexe, c_E affine et c_I convexe,
- $(\bar{x},\bar{\lambda})$ vérifie (KKT),
- $\alpha_k > 0$ « petit »,

alors $x_k \to \bar{x}$.

Dualité lagrangienne augmentée
(problèmes non convexes)

On prend pour φ, le lagrangien augmenté

$$\ell_r(x,\lambda) = f(x) + \sum_{i\in E}\left[\lambda_i c_i(x) + \frac{r}{2}c_i(x)^2\right] + \sum_{i\in I}\left[\lambda_i \max\left(\frac{-\lambda_i}{r}, c_i(x)\right) + \frac{r}{2}\left(\max\left(\frac{-\lambda_i}{r}, c_i(x)\right)\right)^2\right].$$

- **Problème primal** :

$$\inf_{x\in X}\ \sup_{\lambda\in\mathbb{R}^m}\ \ell_r(x,\lambda).$$

- **Problème dual** :

$$\sup_{\lambda\in\mathbb{R}^m}\ \inf_{x\in X}\ \ell_r(x,\lambda).$$

- **Résultat de dualisation** :

Si • $X = \mathbb{E}$,
• $(\bar{x},\bar{\lambda})$ vérifie les CS2,

alors il existe un *voisinage* V de $\bar{x}$ et un seuil $r_0 > 0$ tels que, pour tout $r \geqslant r_0$, $(\bar{x},\bar{\lambda})$ est point-selle de ℓ_r sur $V \times \mathbb{R}^m$.

- **Relaxation lagrangienne augmentée (méthode des multiplicateurs)**

On passe de (λ_k, r_k) à (λ_{k+1}, r_{k+1}) par :

1. $x_k \in \arg\min_{x\in\mathbb{E}} \ell_{r_k}(x,\lambda_k)$,
2. arrêt si (x_k,λ_k) est satisfaisant,
3. $\lambda_{k+1} = P_Y\left[\lambda_k + r_k c(x_k)\right]$ (pas besoin de RL !),
4. adapter $r_k \curvearrowright r_{k+1}$ (heuristique).