Optimisation Différentiable

Théorie et Algorithmes

J.Ch. GILBERT

Résumé du cours

4 septembre 2009

Informations pratiques

- Objectif du cours : l'optimisation
  - aspects théor. : convexité, CO, dualité, ...
  - aspects pratiques : algorithmes.
- Organisation :
  - 14 séances, dont 1 pour l'examen.
  - CM : 12 séances d'1h15++,
  - TD + TP : 9+5 séances d'1h45--,
  - TP : projet d'optimisation (Matlab/Scilab),
  - travail personnel.
- Supports de cours
  - syllabus : ne pas voir les sections avec $\ominus$,
  - transparents : points importants du cours,
  - exercices : en TD, dans le syllabus.
- Contrôle des connaissances
  - TP : contrôle continu
  - Séance 14 : résolution de problèmes (3h).

Plan du cours

1. Rappels, différentiabilité, convexité
2. Conditions d'optimalité I : méthode et outils
3. Conditions d'optimalité II : égalités
4. Conditions d'optimalité III : inégalités
5. Méthodes de descente : RL et RC
6. Méthodes newtoniennes : N et qN
7. Pénalisation
8. Programmation quadratique successive (SQP)
9. Dualité
10. Séance de consolidation (TD)
11. Optimisation linéaire : simplexe et PI
12. Conjugaison
13. Sous-différentiabilité

I Rappels

Existence de solution (§ 1.2)

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

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

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

Si · f est fermée sur X, · X est compact et non vide, alors (P_X) a (au moins) une solution.

En dimension finie (c'est notre cas) :

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

X compact

par

$$X \text{ fermé} \quad \text{et} \quad \lim_{\substack{x \in X \\ \|x\| \to \infty}} f(x) = +\infty.$$

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 :

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

Si · X est convexe, · f est strictement convexe sur X, alors (P_X) a au plus une solution.

Différentiabilité première (§§ C.1, C.2, C.3)

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.4, C.5)

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:

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

- 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

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

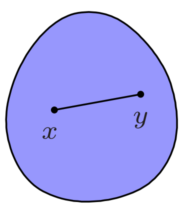
II Analyse convexe

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

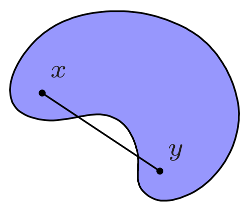
$$[x, y] := \{(1-t)x + ty : t \in [0, 1]\}.$$

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

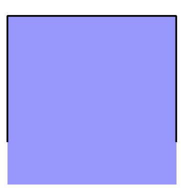
$$x, y \in C \quad \Longrightarrow \quad [x, y] \subset C.$$



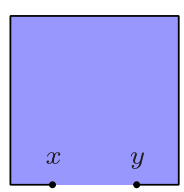
convexe



non convexe



convexe



non convexe

Polyèdre convexe (§ 2.4)

Soit $\mathbb{E}$ un espace vectoriel ($\mathbb{E} = \mathbb{R}^n$ parfois).

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

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

Description duale d'un polyèdre convexe :

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

- **Dfn** : $\hat{x}$ est un sommet de P si

$$x_1, x_2 \in P, \; \hat{x} = \frac{x_1 + x_2}{2} \quad \Longrightarrow \quad \hat{x} = x_1 = x_2.$$

L'ensemble des sommets est noté $\text{ext}\, P$.

- **Prop**.

Soit $P := \{x : Ax = b, x \geqslant 0\}$, $\hat{x} \in P$.

1) $\hat{x} \in \text{ext}\, P \quad \Longleftrightarrow \quad A_B$ est injective, où $B = \{i : x_i > 0\}$ et A_B formée des colonnes $i \in B$ de A.

2) P a au moins un sommet et au plus 2^n.

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 la projection de x sur C et est notée $P_C x$.
- **Prop** : Soit $\bar{x} \in C$. Alors[a]

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

[a]La réciproque de la dernière implication est fausse.

Séparation des convexes (§ 2.5.3)

$\mathbb{E}$ 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.
- **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.

Enveloppe convexe fermée (§ 2.5.4)

Soient $\mathbb{E}$ un e.v. avec $\langle\cdot,\cdot\rangle$ et $P \subset \mathbb{E}$.

- **Dfn**. L'enveloppe convexe de P, $\operatorname{co} P$, est le plus petit convexe contenant P.

$$P \text{ fermé} \quad \not\Rightarrow \quad \operatorname{co} P \text{ fermé}.$$

- **Dfn**. L'enveloppe convexe fermée de P, $\overline{\operatorname{co}} P$, est le plus petit convexe fermé contenant P.

- **Dfn**. Un demi-espace fermé de $\mathbb{E}$:

$$H^-(\xi,\alpha) := \{x \in \mathbb{E} : \langle \xi, x\rangle \leqslant \alpha\},$$

où $\xi \in \mathbb{E}$ est *non nul* et $\alpha \in \mathbb{R}$.

- **Prop**.

> $\overline{\operatorname{co}} P$ est l'intersection de tous les demi-espaces fermés contenant P.

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

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

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 fermé $\neq \varnothing$ 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

$$\{Ax : x \geqslant 0\} = \text{ cône, convexe, fermé, } \neq \varnothing$$

et

$$\{x : Ax \geqslant 0\}^+ = \{A^\top y : y \geqslant 0\}.$$

$(\cdot)^+$ = dual pour le produit scalaire euclidien.
[c'est une généralisation de $N(A)^\perp = R(A^\top)$]

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 $\implies$ $\sup_{i \in I} f_i$ convexe.
- f_i fermées $\implies$ $\sup_{i \in I} f_i$ 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 $X \iff$ $\forall x \in X, \forall h \in \mathbb{E}, f''(x) \cdot h^2 \geqslant 0$,
  – 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: $f(x) = x^4$.

Fonction conjuguée (§ 3.4.9)

Soient $\mathbb{E}$ un espace vectoriel, muni de $\langle \cdot, \cdot \rangle$ et $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ propre avec minorante affine.

- **Dfn**. Conjuguée $f^* : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ de f :

$$f^*(x^*) := \sup_{x \in \mathbb{E}} \Big(\langle x^*, x \rangle - f(x) \Big).$$

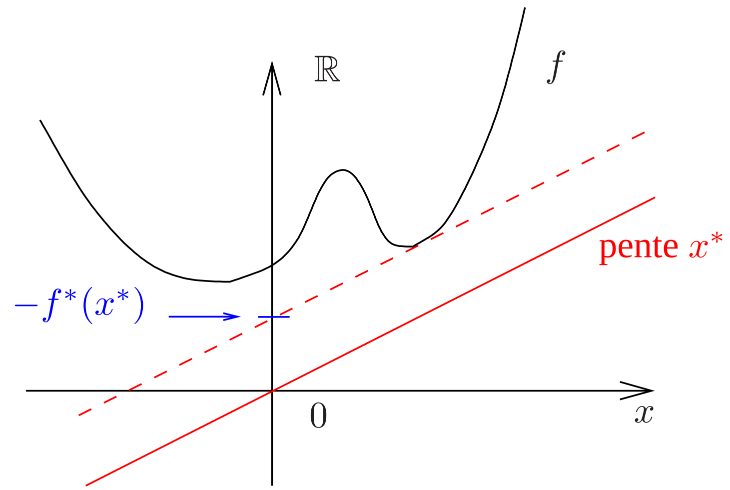
  Biconjuguée $f^{**} : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ de f :

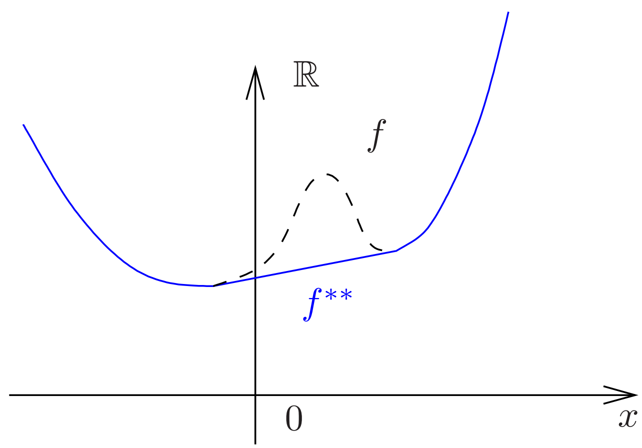
$$f^{**}(x) := \sup_{x^* \in \mathbb{E}} \Big(\langle x^*, x \rangle - f^*(x^*) \Big).$$

- **Prop**. On suppose que f est propre et a une minorante affine.

  1) $f^* \in \overline{\text{Conv}}(\mathbb{E})$ et $f^{**} \in \overline{\text{Conv}}(\mathbb{E})$.
  2) f^{**} est l'enveloppe supérieure des minorantes affines de f.
  3) $f^{**} \leqslant f$.
  4) $f^{**} = f \iff f \in \overline{\text{Conv}}(\mathbb{E})$.

Interprétation de $f^*(x^*)$ et f^{**}





Sous-différentiel des fonctions convexes (§ 3.5)

$\mathbb{E}$ un e.v. muni de $\langle \cdot, \cdot \rangle$.

Différentiabilité directionnelle

$f \in \mathrm{Conv}(\mathbb{E})$, $x \in \mathrm{dom}\, f$ 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) $f'(x;\cdot)$ est convexe;
4) $x \in (\mathrm{dom}\, f)^\circ \implies$ $f'(x;\cdot)$ Lipschitz
   $f'(x;\cdot) \in \overline{\mathrm{Conv}}(\mathbb{E})$.

Sous-différentiel

- **Dfn**. Le sous-différentiel $\partial f(x)$ de $f \in \mathrm{Conv}(\mathbb{E})$ en $x \in \mathrm{dom}\, f$ est l'ensemble des $x^* \in \mathbb{E}$ vérifiant les propriétés équivalentes :

(S_1) $\forall d \in \mathbb{E} : f'(x;d) \geqslant \langle x^*, d \rangle$,

(S_2) $\forall y \in \mathbb{E} : f(y) \geqslant f(x) + \langle x^*, y - x \rangle$,

(S_3) $f(x) + f^*(x^*) = \langle x^*, x \rangle$ [ou $\leqslant$].

- **Prop**.

1) $\hat{x} \in \arg\min f \iff 0 \in \partial f(\hat{x})$.
2) $\partial f(x)$ est un convexe fermé.
3) $x \in (\mathrm{dom}\, f)^\circ \implies$
   - $\partial f(x)$ compact non vide,
   - $f'(x;d) = \max_{x^* \in \partial f(x)} \langle x^*, d \rangle$.
5) f G-diff. en $x \iff \partial f(x) = \{\nabla f(x)\}$.

- **Calcul**.

Si $f_1, \ldots, f_m : \mathbb{E} \to \mathbb{R}$ convexes et $\alpha_i \geqslant 0$

$$\partial \left(\sum_{i=1}^m \alpha_i f_i \right)(x) = \sum_{i=1}^m \alpha_i \partial f_i(x).$$

Exemple 1D

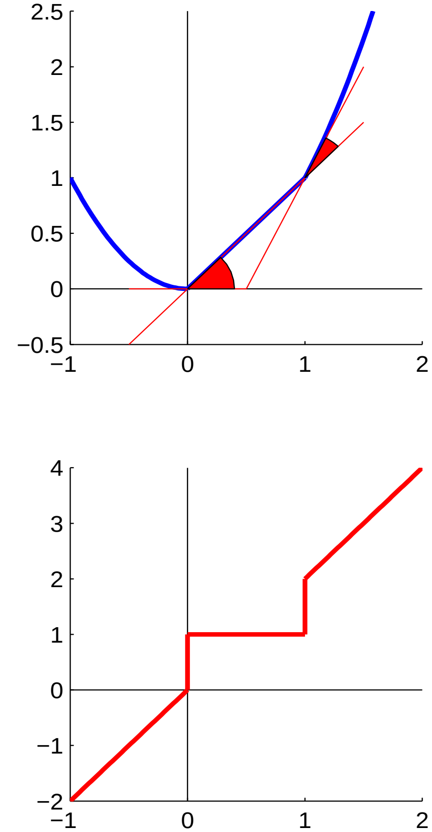


Figure 1: $f(x) = \max(x, x^2)$ et $\partial f(x)$ (en bas)

Exemple 2D

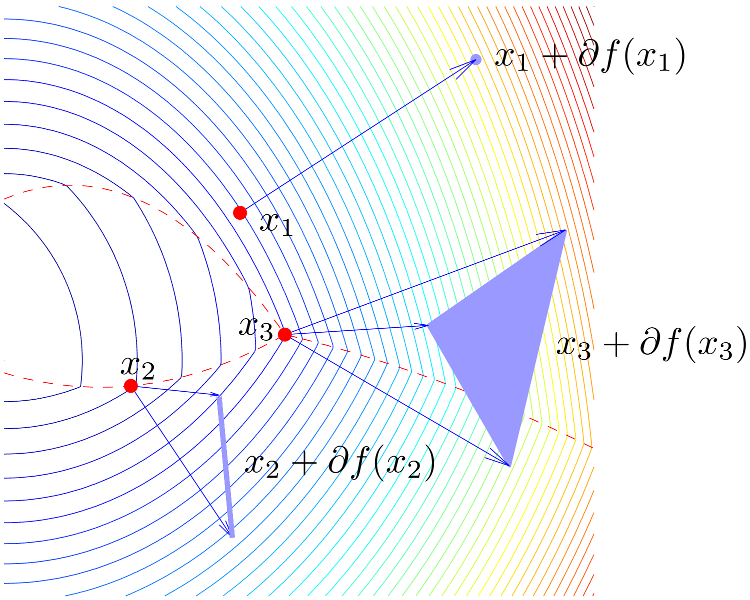


Figure 2: $f = \sup(q_1, q_2, q_3)$ et ∂f

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, prod. scal. $\langle \cdot, \cdot \rangle$).

- Ce sont des $=$ et $\leqslant$ décrivant les solutions de (P_X).
- Utilité des CO :
  - Donner des renseignements sur (P_X) (calculer la solution analytiquement ?).
  - Vérifier qu'un point est solution.
  - Définir des algorithmes de résolution.
- Il y a des CO nécessaires (CN) et des CO suffisantes (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 $\langle \cdot, \cdot \rangle$.

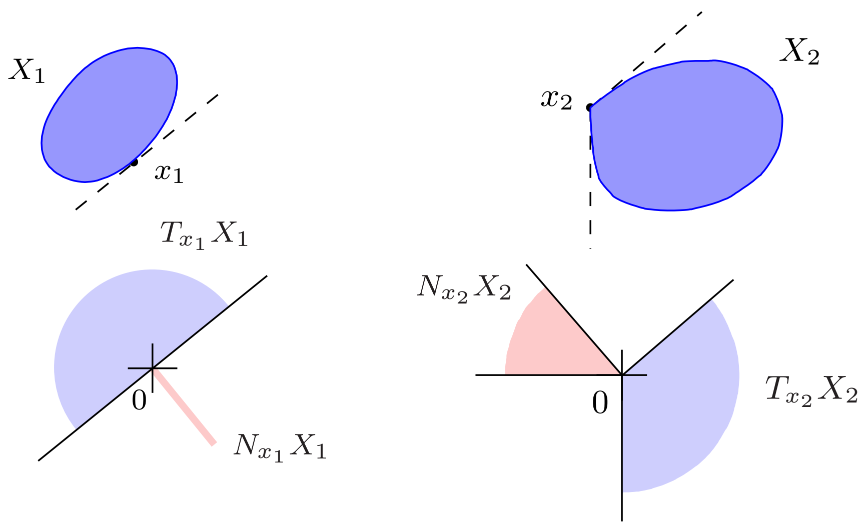
- **CN1** :
$$\nabla f(x_*) = 0.$$
  (Si f est convexe, c'est une **CS1** globale.)
- **CN2** :
$$\begin{cases} \nabla f(x_*) = 0 \\ \nabla^2 f(x_*) \succcurlyeq 0. \end{cases}$$
- **CS2** pour un minimum local strict :
$$\begin{cases} \nabla f(x_*) = 0 \\ \nabla^2 f(x_*) \succ 0. \end{cases}$$

<u>CN1 générale (§ 4.1)</u>

Le problème à résoudre :

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

Dfn : Cône tangent.



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

$$f'(x_*) \cdot d \geqslant 0, \quad \forall d \in T_{x_*}X, \tag{2}$$

où $T_{x_*}X$ est le cône tangent à X en x_*.

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

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

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

CO avec contraintes d'$=$ (§ 4.3)

Le problème à résoudre :

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

Le lagrangien du problème :

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

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

$$\begin{cases} \nabla_x \ell(x_*, \lambda_*) = 0 \\ c(x_*) = 0. \end{cases} \tag{3}$$

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

Si $\mathbb{F} = \mathbb{R}^m$, la première condition de (3) s'écrit

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

- **CN2** : si $A_* := c'(x_*)$ est surjective, on a

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

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

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

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

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

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

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

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

(Si f et c_I sont convexes et c_E est affine, ce sont des **CS1** globales.)

Qualification des contraintes (§ 4.4.2)

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

$$T_x X = T'_x X, \tag{4}$$

où

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

On a toujours : $T_x X \subset T'_x X$.

- **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* avec c'_E *surjective*, les composantes de $c_{I^0(x)}$ sont *convexes*, $\exists \hat{x} \in X$ tel que $c_{I^0(x)}(\hat{x}) < 0$.

(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 $\alpha_{I^0(x)} \geqslant 0$ $\Longrightarrow \quad \alpha_{E \cup I^0(x)} = 0$.

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 (4) avec $x = x_*$). Dès lors
$$\nabla f(x_*) \in \left(T'_{x_*}X\right)^+. \tag{5}$$

- Lemme de Farkas :
$$\{d : Ad \geqslant 0\}^+ = \{A^\top \lambda : \lambda \geqslant 0\}.$$
C'est une généralisation de $N(A)^\perp = R(A^\top)$.

- Le lemme de Farkas permet d'exprimer (5) autrement : $\exists \lambda_* \in \mathbb{R}^m$ tel que l'on ait (KKT).

Signification des multiplicateurs optimaux (§ 4.7.1)

- Problème perturbé : pour $p \in \mathbb{R}^m$, on définit
$$(P_{EI}^p) \quad \begin{cases} \min\ f(x) \\ c_E(x) + p_E = 0 \\ c_I(x) + p_I \leqslant 0. \end{cases}$$

- **Dfn**. La fonction valeur associée à (P_{EI}^p) 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_{EI}^p).
$$(P_{EI}) \text{ convexe} \implies v \text{ convexe}.$$

- Cas différentiable régulier.

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

- **Dfn**. (x_*, λ_*) est une solution primale-duale globale si

$$\begin{cases} x_* \in \arg\min \ell(\cdot, \lambda_*) \\ c_E(x_*) = 0, \quad c_I(x_*) \leqslant 0 \\ (\lambda_*)_I \geqslant 0, \quad (\lambda_*)_I^\top c_I(x_*) = 0. \end{cases}$$

- Cas convexe non différentiable.

Si · x_* est solution de (P_{EI}),
· v est convexe,
· $\partial v(0) \neq \varnothing$,
alors $\partial v(0) = \{\lambda_* : (x_*, \lambda_*) \text{ solution PD globale}\}$.

Remarque : 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.

CN d'optimalité (cas convexe non diff.).

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

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

Si (x_*, λ_*) solution PD globale,
alors x_* solution de (P_{EI}).

IV Méthodes à directions de descente

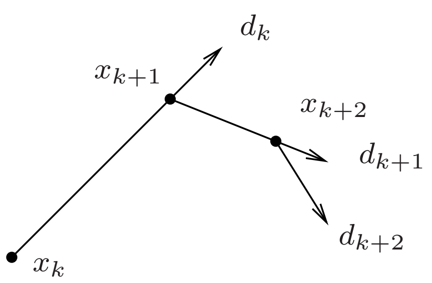
Schéma des algorithmes (§ 6.1)

- **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.

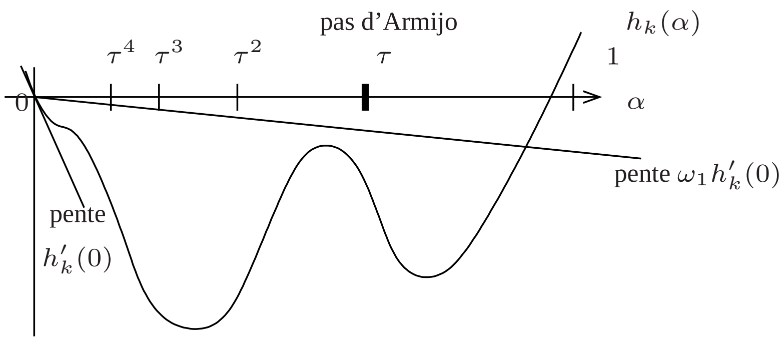
Soient d_k une direction de descente et

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

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

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

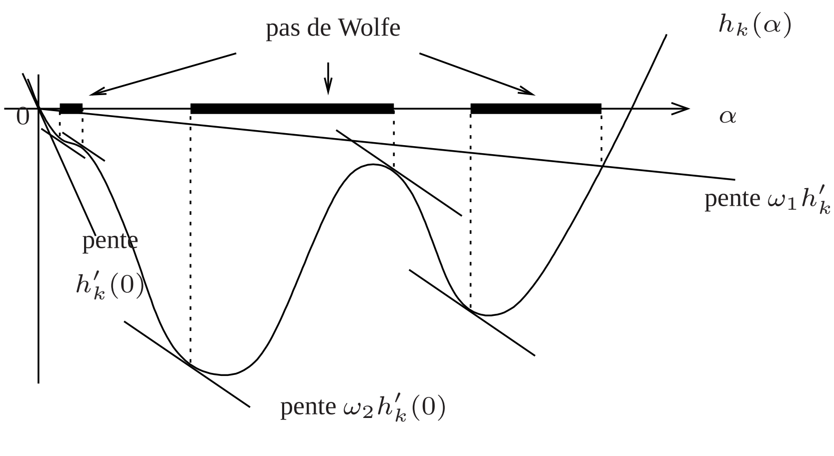
où i_k est le plus petit dans $\{0, 1, 2, \ldots\}$.



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 $\omega_2 = 0.99$.

V Méthodes à régions de confiance

Principe de l'algorithme (§ 8.1.1)

- **Le problème** :

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

- Modèle quadratique de f autour d'un itéré x_k :
$f(x_k + s) \simeq f(x_k) + \psi(s)$, où

$$\psi_k(s) := \langle g_k, s \rangle + \frac{1}{2} \langle M_k s, s \rangle.$$

- Région de confiance : région dans laquelle ce modèle est considéré comme bon. Le plus souvent

$$B(0, \Delta_k) := \{s \in \mathbb{E} : \|s\| \leqslant \Delta_k\}.$$

$\Delta_k > 0$ est le rayon de confiance du modèle.

- **Schéma d'un algorithme à RC** :
il génère une suite $\{x_k\}$ par

1. *Déplacement* :

$$s_k \simeq \arg\min_{\|s\| \leqslant \Delta_k} \psi_k(s);$$

2. *Appréciation du déplacement* :
si la concordance

$$\rho_k := \frac{f(x_k + s_k) - f(x_k)}{\psi(s_k)}$$

n'est pas bonne ($\rho_k \leqslant \omega_1$), diminuer Δ_k et retour en 1.

3. *Nouvel itéré* :

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

4. *Nouveau modèle* :
   – nouveau rayon de confiance

$$\Delta_{k+1} \in \begin{cases} [\tau_2 \Delta_k, \Delta_k] & \text{si } \rho_k \leqslant \omega_2 \\ [\Delta_k, \tau_3 \Delta_k] & \text{sinon.} \end{cases}$$

   – calculer $g_{k+1} := \nabla f(x_{k+1})$, M_{k+1}.

Itérés générés par RC dans le contre-exemple de Powell

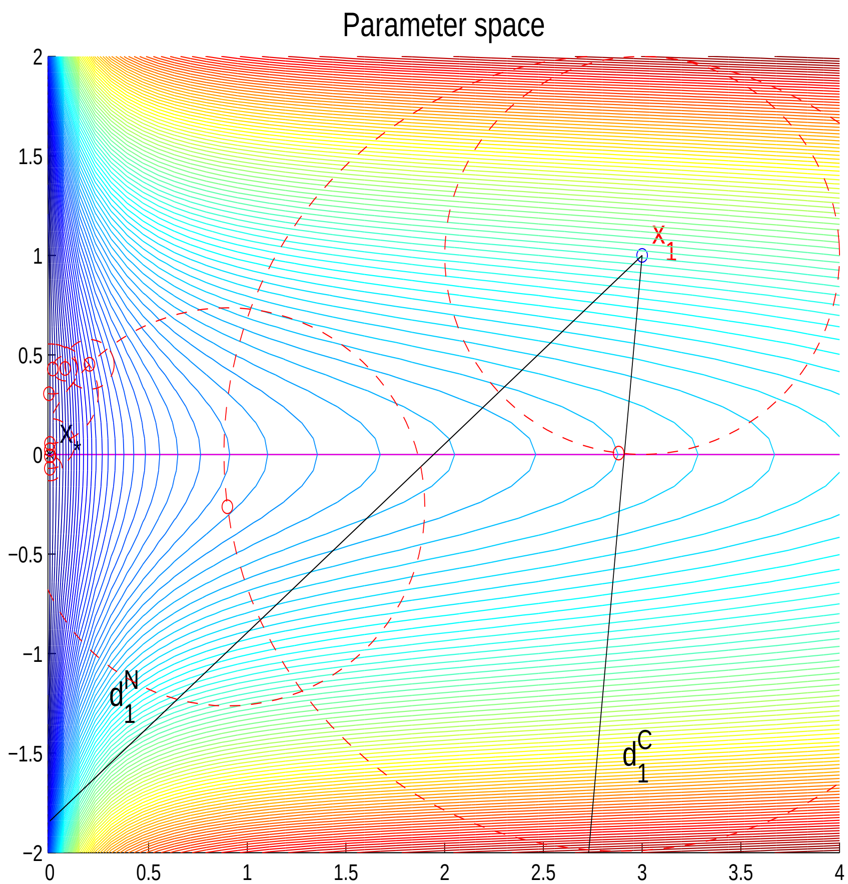


Figure 3: RC dans le contre-exemple de Powell

Résolution du problème quadratique (§§ 8.3.1, 8.4)

- **Le problème** :

$$\begin{cases} \min \langle g, s\rangle + \frac{1}{2}\langle Ms, s\rangle \\ \|s\| \leqslant \Delta. \end{cases}$$

- CNS d'optimalité : $\exists \hat{\lambda} \in \mathbb{R}$ tel que

$$\begin{cases} (M + \hat{\lambda} I)\hat{s} = -g \\ \|\hat{s}\| \leqslant \Delta, \quad \hat{\lambda} \geqslant 0 \\ \hat{\lambda}(\Delta - \|\hat{s}\|) = 0 \\ (M + \hat{\lambda} I) \succcurlyeq 0. \end{cases}$$

- *Résolution approchée* :
  - algorithme dogleg de Powell
  - algorithme du GC tronqué.

  Résolution fine :
  - algorithme de Moré-Sorensen.

VI 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 $r \in [0,1[$ tels que $\forall\, k \geqslant k_0$:
$$\frac{\|x_{k+1}-x_*\|}{\|x_k-x_*\|} \leqslant r.$$

- Convergence superlinéaire :
$$\frac{\|x_{k+1}-x_*\|}{\|x_k-x_*\|} \to 0.$$

- Convergence quadratique : il existe une constante $C>0$ telle que $\forall\, k \geqslant 1$:
$$\frac{\|x_{k+1}-x_*\|}{\|x_k-x_*\|^2} \leqslant C.$$

$\chi_k =$ nombre de chiffres significatifs corrects.

	superlinéaire		quadratique	
k	x_k	χ_k	x_k	χ_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		

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

$$\text{Superlinéaire} \Longrightarrow \chi_{k+1} - \chi_k \to \infty.$$

$$\text{Quadratique} \Longrightarrow \liminf_{k\to\infty} \frac{\chi_{k+1}}{\chi_k} \geqslant 2.$$

Algorithme de Newton pour systèmes non linéaires (§ 9.1.1

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

$$F(x) = 0.$$

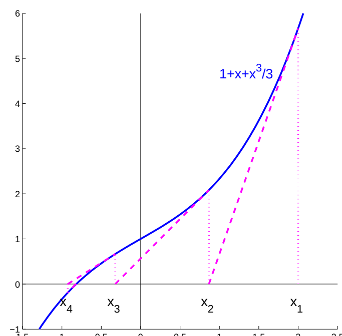
- **Algorithme de Newton.** De x_k à x_{k+1} :
  - *Résoudre* en d_k l'équation de Newton :

$$F'(x_k)\, d_k = -F(x_k). \tag{6}$$

  - *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
V, l'algorithme de Newton (6) est bien défi
et génère une suite $\{x_k\} \subset V$ qui converg
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 $\omega_1 \in]0, \frac{1}{2}[$)

  $$f(x_+) \leqslant f(x) + \alpha\, \omega_1\, f'(x) \cdot d^{\mathrm{N}}.$$

- 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 . . . !

Globalisation de l'algorithme de Newton par régions de confiance (§ 9.3.2)

- **Principes** :
  1. On ne s'intéresse plus qu'aux points stationnaires de
  $$f(x) = \frac{1}{2}\|F(x)\|_2^2.$$
  Remarque : la RL n'est pas toujours capable d'en trouver.
  2. On prend comme modèle quadratique de f en x_k :
  $$\varphi_k(s) := \frac{1}{2}\|F(x_k) + F'(x_k)s\|_2^2. \qquad (7)$$
  Avantage : minimiseur s_k est défini même si $F'(x_k)$ n'est pas inversible (c'est l'origine de l'affaiblissement des hypothèses).
  3. On minimise φ_k sur une région de confiance :
  $$\begin{cases} \min_s \ \varphi_k(s) \\ \|s\|_2 \leqslant \Delta_k. \end{cases} \qquad (8)$$

- **Résultat de convergence** :

  Si $\{F'(x_k)\}$ est *bornée*,
  alors l'algorithme de Newton avec RC converge vers un point stationnaire de f : $\nabla f(x_k) \to 0$.

  Remarque : on n'a plus besoin d'hypothèse sur $F'(x)^{-1}$!

VII

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 :

$$\nabla^2 f(x_k)\, d_k = -\nabla f(x_k). \qquad (9)$$

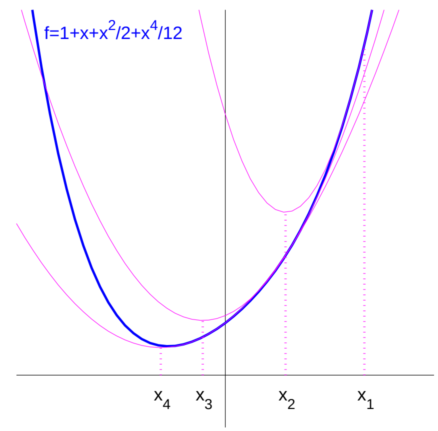
  – *Nouvel itéré* :

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

- **Le problème quadratique tangent.**

  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.**

⊕⊕ 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_*.

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

⊖ Pas de distinction entre min, max, point stationnaire.

⊖ Les directions ne sont pas nécessairement de descente.

⊖ Il faut calculer les dérivées secondes de f.

Algorithmes de quasi-Newton (§ 10)

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} dfn. pos.

  Cela conduit à la formule de BFGS.

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

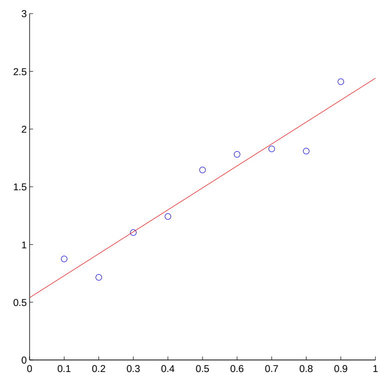
VIII **Problèmes de moindres-carrés**

- 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 linéaires (§ 16.1)

- **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 :

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

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

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

Moindres-carrés non linéaires (§ 16.2)

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

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

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 $J(x_k)^\top r(x_k) \to 0$).

IX $\boxed{\textbf{Pénalisation (§ 12)}}$

- **À quoi ça sert ?**

  En optimisation avec contraintes :

  – pour la théorie: obtenir des propriétés à partir de problèmes approchés sans contrainte,

  – pour l'algorithmique: résoudre un problème avec contraintes "sans trop en faire".

- **Exemple**. On veut résoudre

$$(P) \quad \left\{ \begin{array}{l} \min\ f(x) \\ c(x) \leqslant 0. \end{array} \right.$$

On *approche* ce problème par ($r > 0$)

$$(P_r) \quad \min\ f(x) + \frac{r}{2}\|c(x)^+\|_2^2,$$

que l'on résout par un algorithme de descente, pour une <u>suite</u> de $r \to \infty$.

- **Pénalisation ℓ_2 en 1D**

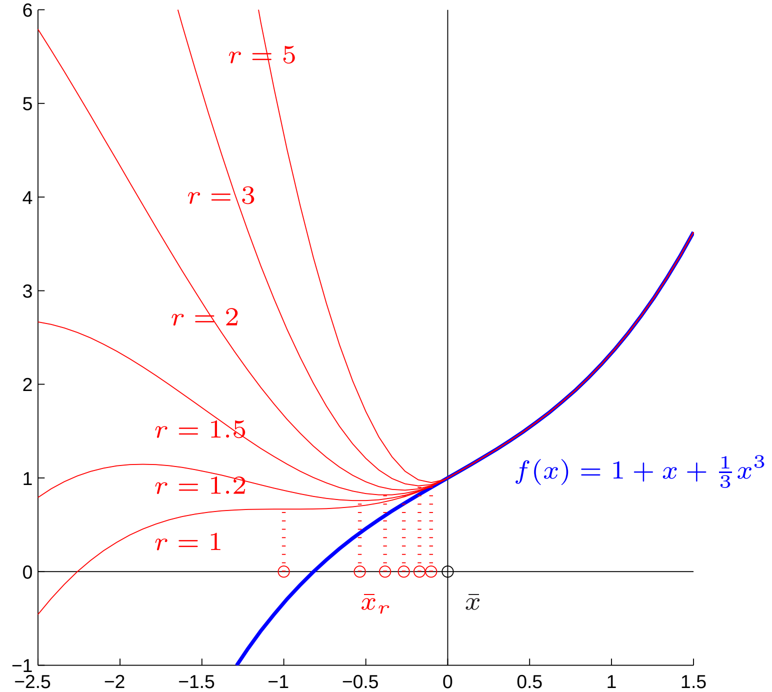


Figure 4: Pénalisation quadratique

- **Pénalisation extérieure (§ 12.2).**
Plus généralement, si $X \subset \mathbb{E}$, on approche

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

par une <u>suite</u> de problèmes sans contrainte

$$(P_r) \quad \min \ f(x) + rp(x),$$

où la fonction p vérifie

$$(H_p) \quad \begin{cases} p \text{ est continue sur } \mathbb{E} \\ p(x) \geqslant 0, \ \forall x \in \mathbb{E} \\ p(x) = 0 \iff x \in X. \end{cases}$$

<u>Résultat</u> <u>d'approximation</u>.

<u>Si</u> · f continue et $f(x) \to +\infty$ lorsque $\|x\| \to \infty$,
· X est fermé et non vide,
· $p : \mathbb{E} \to \mathbb{R}$ vérifie (H_p),

<u>alors</u> 1) $\forall r > 0$, (P_r) a au moins 1 solution $\bar{x}_r$,
2) $\{\bar{x}_r\}_{r\uparrow\infty}$ est bornée,
3) tout point d'adhérence de la suite $\{\bar{x}_r\}_{r\uparrow\infty}$ est solution de (P_X).

- **$\oplus$ et $\ominus$ de l'approche par pénalisation**

$\oplus$ Facile à mettre en œuvre (avec algo. sans contrainte).

$\ominus$ <u>Suite</u> de problèmes non linéaires (bon r inconnu, premier r très grand ne convient pas).

$\ominus$ Le mauvais conditionnement augmente avec r (i.e., les courbes de niveau s'allongent).

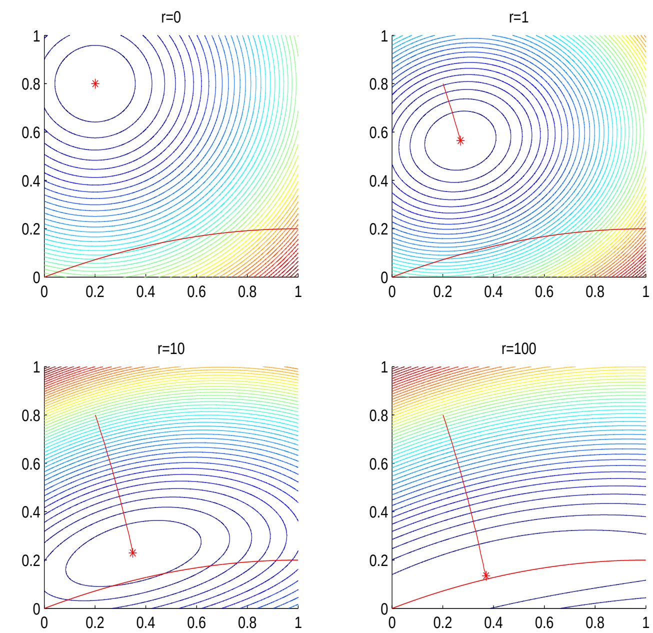
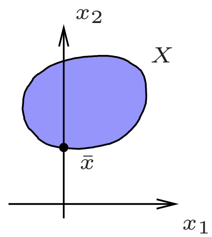


Figure 5: Chemin des minimiseurs

X $\boxed{\textbf{Dualité } (\S\,13)}$

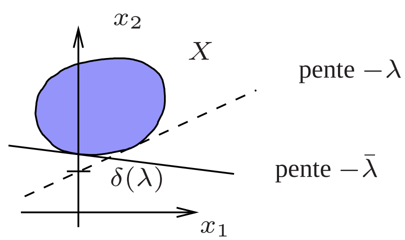
- **Un premier problème :**

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



- **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 ($\S$ 13.1)

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

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

- **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

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

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

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

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** :

$$\sup_{y\in Y} \inf_{x\in X} \varphi(x,y) \leqslant \inf_{x\in X} \sup_{y\in Y} \varphi(x,y).$$

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

- **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 $(\bar{x},\bar{y})$, alors les solutions de (P) sont solutions du pbl interne en $y = \bar{y}$.

Dualisation de contraintes fonctionnelles (§ 13.4)

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 (§ 13.4.1)

(problèmes convexes)

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

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

- **Problème primal** : $Y := \mathbb{R}^{m_E} \times \mathbb{R}^{m_I}_+$ et

$$(P_{X,EI}) \equiv \inf_{x \in X} \sup_{\lambda : \lambda_I \geqslant 0} \ell(x,\lambda).$$

- **Problème dual** :

$$\sup_{\lambda : \lambda_I \geqslant 0} \inf_{x \in X} \ell(x,\lambda).$$

- **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 (§ 13.4.2)
(problèmes non convexes)

On prend pour φ, le lagrangien augmenté

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

- **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** :

<u>Si</u> · $X = \mathbb{E}$,
· $(\bar{x},\bar{\lambda})$ vérifie les CS2,

<u>alors</u> 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. $\lambda_{k+1} = P_Y\,[\lambda_k + r_k c(x_k)]$
(pas besoin de RL !),
4. adapter $r_k \curvearrowright r_{k+1}$ (heuristique).

XI $\boxed{\textbf{Optimisation linéaire : simplexe}}$

- On considère le problème d'optimisation linéaire sur $\mathbb{R}^n$ (forme standard)

$$(P_L) \quad \left\{ \begin{array}{l} \min\ c^\top x \\ Ax = b \\ x \geqslant 0, \end{array} \right.$$

  où
  - $c \in \mathbb{R}^n$,
  - A est $m \times n$ surjective ($m \leqslant n$),
  - $b \in \mathbb{R}^m$.

- On note

$$X := \{x \in \mathbb{R}^n : Ax = b,\ x \geqslant 0\}$$

  l'ensemble admissible.

- Pour $x \in X$, on note

$$\begin{array}{rcl} I^+(x) & := & \{i : x_i > 0\} \\ I^0(x) & := & \{i : x_i = 0\}. \end{array}$$

Structure de l'ensemble admissible (§ 17.2.1)

- On appelle sommet de X, un point $x \in X$ tel que

$$\begin{array}{l} x_1, x_2 \in X, \\ x = (1-t)x_1 + tx_2, \quad t \in]0,1[\\ \Longrightarrow \quad x = x_1 = x_2. \end{array}$$

- **Théor** (reconnaître un sommet) :

$$\begin{array}{l} x \in X \text{ est un sommet de } X \\ \Longleftrightarrow \quad \text{les colonnes } \{A^j : x_j > 0\} \text{ sont LI.} \end{array}$$

- Un sommet $x \in X$ est dit non dégénéré, si $|I^+(x)| = m$ (donc $A_{I^+(x)}$ est inversible)

- **Théor** (existence de sommet) :

$$\begin{array}{l} X \neq \varnothing \\ \Longrightarrow \quad X \text{ a (au moins) un sommet.} \end{array}$$

Existence de solution (§ 17.2.2)

- **Théor** (existence de solution) :

  (P_L) a une solution

  $\iff$ (P_L) est réalisable et borné.

- **Théor** (existence de solution-sommet) :

  (P_L) a une solution

  $\Longrightarrow$ (P_L) a une solution-sommet.

Conditions d'optimalité (§ 17.2.2)

- **Théor** :

  (P_L) a une solution

  $$\iff \exists y \in \mathbb{R}^m,\ \exists s \in \mathbb{R}^n\ : \left\{ \begin{array}{l} A^\top y + s = c, \quad s \geqslant 0, \\ Ax = b, \quad x \geqslant 0, \\ x^\top s = 0. \end{array} \right.$$

Algorithme du simplexe (§ 17.4)

(description géométrique)

- **Hypothèse** : A est $m \times n$ surjective.

- **Phase I** : trouver $\hat{x} \in X$, un sommet de X.

  On prend $B \supset I^+(\hat{x})$ avec $|B| = m$ et $N = B^c \subset I^0(\hat{x})$.

- **Phase II** : on itère de sommet en sommet. Voici une itération.
  - Coût réduit : $r := c_N - A_N^\top A_B^{-\top} c_B$.
  - Optimalité :
    si $r \geqslant 0$, $\hat{x}$ est solution (arrêt);
    sinon $\exists j$ t.q. $r_j < 0$.
  - Direction de déplacement d :
    $d_N = e_N^j$ et $d_B = -A_B^{-1} A_N e_N^j$.
  - Si $d_B \geqslant 0$, (P_L) est non borné (arrêt).
  - Nouveau sommet $\hat{x}^+$:
    prendre le plus grand $\alpha \geqslant 0$ tel que

    $$\hat{x}^+ := \hat{x} + \alpha d \in X$$

    ($\alpha > 0$ si $\hat{x}$ est non dégénéré).

XII Optimisation linéaire : points intérieurs

La une du New York Times

(19 novembre 1984)

Voici comment l'algorithme de Karmarkar (le premier algorithme de points intérieurs) était "révélé" au grand public :

> *"The discovery, which is to be formally published next month, is already circulating rapidly through the mathematics world. It has also set off a deluge of inquiries from brokerage houses, oil companies and airlines, industries with millions of dollars at stake in problems known as linear programming."*

Notations

Le problème et son dual

$$(P)\ \begin{cases} \inf\ c^\top x \\ Ax = b \\ x \geqslant 0, \end{cases} \qquad (D)\ \begin{cases} \sup\ b^\top y \\ A^\top y + s = c \\ s \geqslant 0. \end{cases}$$

Ensembles admissibles

$$\mathcal{F}_P := \{x \in \mathbb{R}^n : Ax = b,\ x \geqslant 0\}$$

$$\mathcal{F}_D := \{(y,s) \in \mathbb{R}^{m+n} : A^\top y + s = c,\ s \geqslant 0\}.$$

Intérieurs relatifs

$$\mathcal{F}_P^o := \{x \in \mathbb{R}^n : Ax = b,\ x > 0\}$$

$$\mathcal{F}_D^o := \{(y,s) \in \mathbb{R}^{m+n} : A^\top y + s = c,\ s > 0\}.$$

Conditions d'optimalité

$$\begin{cases} Ax = b, \quad x \geqslant 0 \\ A^\top y + s = c, \quad s \geqslant 0 \\ x^\top s = 0. \end{cases}$$

Le chemin central primal-dual (§ 18.1)

- Conditions d'optimalité perturbées par $\mu > 0$:

$$(\mathrm{KKT}_\mu) \quad \begin{cases} Ax = b \quad (x > 0) \\ A^\top y + s = c \quad (s > 0) \\ Xs = \mu\, e. \end{cases}$$

- **Existence et unicité** :

$$\boxed{\begin{array}{l}\underline{\text{Si}}\ \ \mathcal{F}_P^o \neq \varnothing,\ \mathcal{F}_D^o \neq \varnothing \text{ et } \mu > 0, \\ \underline{\text{alors}}\ \ (\mathrm{KKT}_\mu) \text{ a une solution unique.}\end{array}}$$

  Si $\mu = 0$, il y a existence mais pas néc. unicité !

- **Dfn** : le chemin central est l'ensemble des solutions de (KKT_μ) :

$$\{(x_\mu, y_\mu, s_\mu) : \mu > 0\}.$$

Algorithme PD de suivi de chemin

(éléments constitutifs, § 18.2)

- On suppose A surjective.
- Soit $z = (x, y, s)$ l'itéré courant admissible.
- Choix de

$$\mu := \frac{x^\top s}{n}.$$

- Facteur de réduction $\sigma \in \,]0, 1[$ de μ.
- Direction de Newton $dz = (dx, dy, ds)$

$$\begin{pmatrix} 0 & A^\top & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \sigma\mu e - Xs \end{pmatrix}.$$

  où $X = \mathrm{diag}(x_i)$, $S = \mathrm{diag}(s_i)$, $e = (1 \ \cdots \ 1)^\top$.

$$(x, s) > 0 \implies \text{SL inversible.}$$

- Nouvel itéré $z^+ = z + \alpha dz$.
- Contrôle du pas α pour que ($\gamma \simeq 10^{-3}$)

$$z^+ \in V_{-\infty}(\gamma) := \{z \in \mathcal{F}^o : Xs \geqslant \gamma\mu e\}.$$

Algorithme PD de suivi de chemin

(algorithme des grands déplacements, § 18.3.3)

- **Paramètres** :
  $0 < \sigma_{\min} < \sigma_{\max} < 1$ et $\gamma \in]0, 1[$.
- **Donnée** :
  $z = (x, y, s) \in V_{-\infty}(\gamma)$ primal-dual admissible (i.e., $Ax = b$ et $A^\top y + s = c$).
- **Une itération** :
  - Choix de $\sigma \in]0, 1[$.
  - Calcul de la direction de Newton dz.
  - Choisir $\alpha \in [0, 1]$ le plus grand possible pour que $z + \alpha dz \in V_{-\infty}(\gamma)$.
  - Nouvel itéré $z^+ := z + \alpha dz$.