Pénalisation

1) Introduction

A Vue d'ensemble

Pour la resolution des problèmes d'optimisation avec contraintes

- On a vu une famille d'algorithmes duaux dans la séance sur la dualité : ils se concentrent sur la génération d'une suite de multiplicateurs $(\lambda_k) \to \lambda_*$; une solution primale x_* peut parfois être obtenue en resolvant le problème

$$\inf_{x \in X} \varphi(x, \bar{y}) = \lambda_*$$

- On présente dans cette séance une famille d'algorithmes primaux

  ⊗ ils se concentrent sur la génération d'une suite primale $(x_k) \to x_*$

  ⊗ un multiplicateur optimal λ_* peut être obtenu comme sous-produit des algorithmes

→ Proposition 1.11 (Faire passer un terme de l'objectif en contrainte)

$$\underset{x \in E}{\inf} \; f(x) + g(x) \quad (P) \qquad = \qquad \underset{\substack{(x,\gamma) \in E \times \mathbb{R} \\ g(x) \le \gamma}}{\inf} \; f(x) + \gamma \quad (P')$$

x_* est sol de (P) avec $\gamma_* = g(x_*)$

$\Longleftrightarrow$ (x_*, γ_*) est solution de (P')

→ Pénalisation (faire passer les contraintes dans l'objectif)

$$\underset{x \in X}{\inf} \; f(x)$$

[b] Principes généraux

– On cherche à résoudre (algorithmiquement ou théoriquement) le problème avec contrainte

$$(P_X) \begin{cases} \inf f(x) \\ x \in X \end{cases} \qquad \begin{array}{l} f : X \to \mathbb{R} \\ X \subset \text{espace vectoriel } \mathbb{E} \end{array}$$

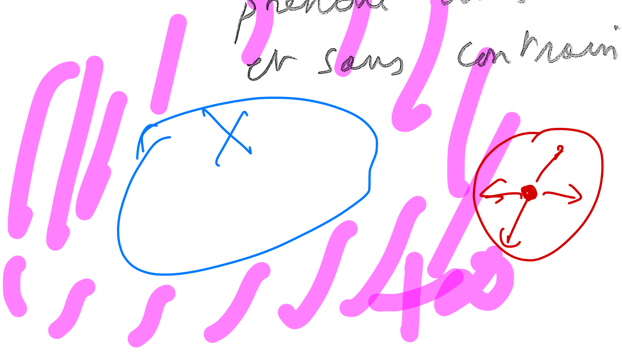
– Dans les méthodes de pénalisation, on remplace (P_X) par un ou plusieurs problèmes sans contrainte, de la forme

$$\boxed{(P_r) \quad \inf_{x \in \mathbb{E}} \theta_r(x)}$$

où (souvent) $\theta_r : \mathbb{E} \to \mathbb{R}$ (parfois $\overline{\mathbb{R}}$) prend en $x \in \mathbb{E}$ la valeur

$$\boxed{\theta_r(x) = f(x) + \underset{\substack{\text{facteur de pénalisation}\\ (\text{souvent } > 0)}}{r}\; \underbrace{p(x)}_{\substack{p : \mathbb{E} \to \overline{\mathbb{R}} \\ \text{fonction pénalisante}}}}$$

– On pourrait prendre $p = I_X$ (indicatrice de X) et $r = 1$, mais ce n'est pas utile en pratique ; c'est utile en théorie pour pouvoir prendre dans le même cadre les problèmes avec et sans contrainte.



$$I_X(x) = \begin{cases} 0 & \text{si } x \in X \\ +\infty & \text{sinon} \end{cases}$$

– À quoi ça sert ?

- On peut alors utiliser des algorithmes d'optimisation sans contrainte pour résoudre le problème avec contrainte (P_X).

  C'est la "méthode du pauvre"
  (on ne connaît pas grand chose et on en fait le moins possible → de moins en moins le cas, grâce au développement de bons solveurs d'optimisation)

- C'est aussi un outil pour construire des algorithmes plus sophistiqués (points intérieurs, par ex.) (globalisation de SQP)

- Cela permet d'interpréter certains algorithmes, de mieux les comprendre ; cela définit un cadre conceptuel.

- Sert aussi à la théorie
  - pour montrer l'existence de solution de (P_X) à partir de celles des (P_r)
  - pour donner des propriétés de (P_X) à partir de celles des (P_r)
  - pour établir des conditions d'optimalité de (P_X) à partir de celles de (P_r)

— Pénalisation exacte et inexacte

(concepts vagues mais très utiles !)

- On dit que la pénalisation est exacte si

$$\boxed{x_* \text{ "solution" de } (P_X) \text{ "}\Longleftrightarrow\text{" } x_* \text{ "solution" de } (P_r)}$$

  - il faut toujours dire ce que l'on entend par "solution" (point stationnaire, min local/global)
  - l'équivalence est rarement vraie ; on est déjà content si l'on peut montrer "$\Rightarrow$" pour un certain r (l'implication réciproque est rarement vraie, sauf si l'on suppose $x_* \in X$)
  - il faut préciser les valeurs de r pour lesquelles la propriété est vraie.
  - avantage : il suffit de résoudre un (P_r) pour résoudre (P_X) !
  - inconvénient : (P_r) est souvent un problème non différentiable.

- On dit que la pénalisation est inexacte dans tous les autres cas.

  - il faut alors résoudre plusieurs (P_r) (une infinité ?) pour résoudre (P_X)

Proposition (monotonie en pénalisation)

Si $\forall r$ considérés, $\theta_r(x) = f(x) + r\,p(x)$ a un minimiseur $\bar{x}_r$

alors lorsque $r \nearrow$, on a

1) $p(\bar{x}_r) \searrow$

2) $f(\bar{x}_r) \nearrow$ (si $r \geq 0$)

Dém 1) $r_1 < r_2$, $x_i = \bar{x}_{r_i}$

$\theta_{r_1}(x_1) \leq \theta_{r_1}(x_2)$

$$\cancel{f(x_1)} + r_1\,p(x_1) \leq \cancel{f(x_2)} + r_1\,p(x_2) \quad \leq r_1\,p(x_1)$$

$\theta_{r_2}(x_2) \leq \theta_{r_2}(x_1)$

$$+\quad \cancel{f(x_2)} + r_2\,p(x_2) \leq \cancel{f(x_1)} + r_2\,p(x_1)$$

$$\underbrace{(r_2 - r_1)}_{>0}\,p(x_2) \leq \underbrace{(r_2 - r_1)}_{>0}\,p(x_1)$$

$$\Rightarrow \quad p(x_2) \leq p(x_1)$$

2) $f(x_1) \leq f(x_2)$

$\square$

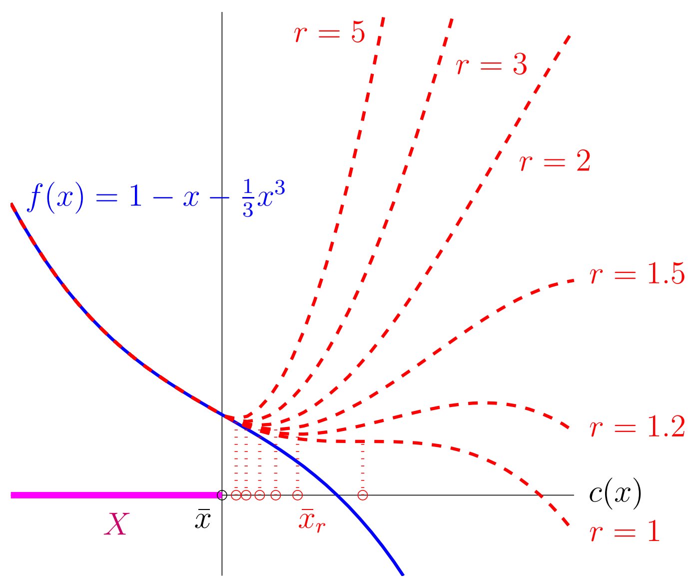
2) Pénalisation extérieure

A Exemples, définition, convergence

• Exemples $(P_X) \quad \inf_{x\in X} f(x), \quad X \subset E$

définition de X		Θ_r
⊗ $c(x) = 0$	$\longrightarrow$	$f(x) + \frac{r}{2}\|c(x)\|_2^2$
⊗ $c(x) \leq 0$	$\longrightarrow$	$f(x) + \frac{r}{2}\|c(x)^+\|_2^2$

$c(x)^+ = \max(0, c(x))$



Pénalisation quadratique

$\begin{cases} \min f(x) \\ x \leq 0 \end{cases}$

$\min f(x) + r[\max(0,x)]^2$

• Définition On va établir un résultat général montrant ce qui se passe pour $\bar{x}_r \in \arg\min \Theta_r$ lorsque $r \nearrow \infty$, sous les hypothèses suivantes

$$\boxed{\Theta_r(x) = f(x) + r\,p(x)}$$

$$(H)\begin{cases} (i) & p \text{ est s.c.i sur } E \\ (ii) & p(x) \geq 0, \quad \forall x \in E \\ (iii) & p(x) = 0 \iff x \in X \end{cases}$$

la pénalisation de f par p n'intervient qu'à l'extérieur de X (d'où l'appellation)

<u>théo</u> (Convergence de la pénalisation quadratique)

<u>Si</u>
- X <u>fermé</u>, <u>non vide</u> de $\mathbb{E}$
- $p : \mathbb{E} \to \mathbb{R}$ vérifie (H)
- f est s.c.i $\qquad \theta_r(x) = f(x) + r\,p(x)$
- $\exists\, r_0 \geq 0 :\ \theta_{r_0}$ est coercive

<u>alors</u>
1) $\forall r \geq r_0$, (P_r) a une solution, disons $\bar{x}_r$
2) $(\bar{x}_r)_{r \nearrow \infty}$ est bornée
3) tout point d'adhérence de $(\bar{x}_r)$ est solution de (P_X)

<u>Leçon</u> : pour résoudre (P_X), on génère une <u>suite</u> $(\bar{x}_r)$ de minimiseurs de θ_r avec $r \nearrow \infty$

<u>Dém</u> 1) Si $r \geq r_0$, $x \in \mathbb{E}$

$$\theta_r(x) = \theta_{r_0}(x) + \underbrace{(r - r_0)}_{\geq 0}\, \underbrace{p(x)}_{\geq 0\ (H)}$$

$$\geq \underbrace{\theta_{r_0}(x)}_{\text{coercive}}$$

$\Rightarrow$ θ_r coercive si $r \geq r_0$

$\Rightarrow$ $\exists$ solution $\bar{x}_r \in \underset{x \in \mathbb{E}}{\arg\min}\ \theta_r(x)$

2) $\theta_{r_0}(\bar{x}_r) \leq \theta_r(\bar{x}_r) \leq \theta_r(x) \leq f(x)$ ci-dessus de r

$\forall r \geq r_0 \quad \forall x \in E \quad \forall x \in X$

on fixe x

θ_{r_0} coercive $\Rightarrow (\bar{x}_r)$ bornée

3) Soit $\bar{x}_r \longrightarrow \bar{x}$ (sous-suite)

$\overset{?}{\Rightarrow} \bar{x} \in \text{Sol}\,(P_X)$

$\bar{x} \in X$?

$\underbrace{\theta_{r_0}(\bar{x}_r)}_{\geq \min \theta_{r_0}} + (r-r_0)\,p(\bar{x}_r) = \theta_r(\bar{x}_r) \leq f(x)$

$\forall x \in X$

$\Rightarrow 0 \leq p(\bar{x}_r) \leq \dfrac{f(x) - \min \theta_{r_0}}{r - r_0}$ ($x \in X$ fixé)

$\longrightarrow 0$ lorsque $r \to \infty$

$0 \leq p(\bar{x}) \leq \liminf_{r\to\infty} p(\bar{x}_r) = 0$

$\Rightarrow p(\bar{x}) = 0 \Rightarrow \bar{x} \in X$

$f(\bar{x}) \leq f(x), \quad \forall x \in X$?

$f(\bar{x}_r) \leq \theta_r(\bar{x}_r) \leq f(x), \quad \forall x \in X$

$\Rightarrow f(\bar{x}) \leq \liminf_{r\to\infty} f(\bar{x}_r) \leq f(x)$

$\square$

B Schéma algorithmique

On passe de $(x_k, r_k) \to (x_{k+1}, r_{k+1})$ par

1) si x_k est satisfaisant $\to$ arrêt
2) prendre $r_{k+1} > r_k$ (heuristique)
3) x_{k+1} solution (approchée) de $\min_{x \in E} \theta_{r_{k+1}}(x)$

Remarques

1) "x_k satisfaisant" revient souvent à vérifier que les contraintes sont presque satisfaites.

2) Il ne suffit pas de prendre r très grand une fois pour toutes !! Étrange :
   - c'est quoi r grand
   - si r trop grand, on ne "voit" plus f
   - si r est grand, θ_r est mal conditionnée (voir ci-dessous) et prendre x_k comme point de départ à l'étape 3 permet en partie de faire face au mauvais conditionnement

3) si r grand, θ_r est mal conditionnée ; voyons cela pour (P_E) et la pénalisation quadratique :

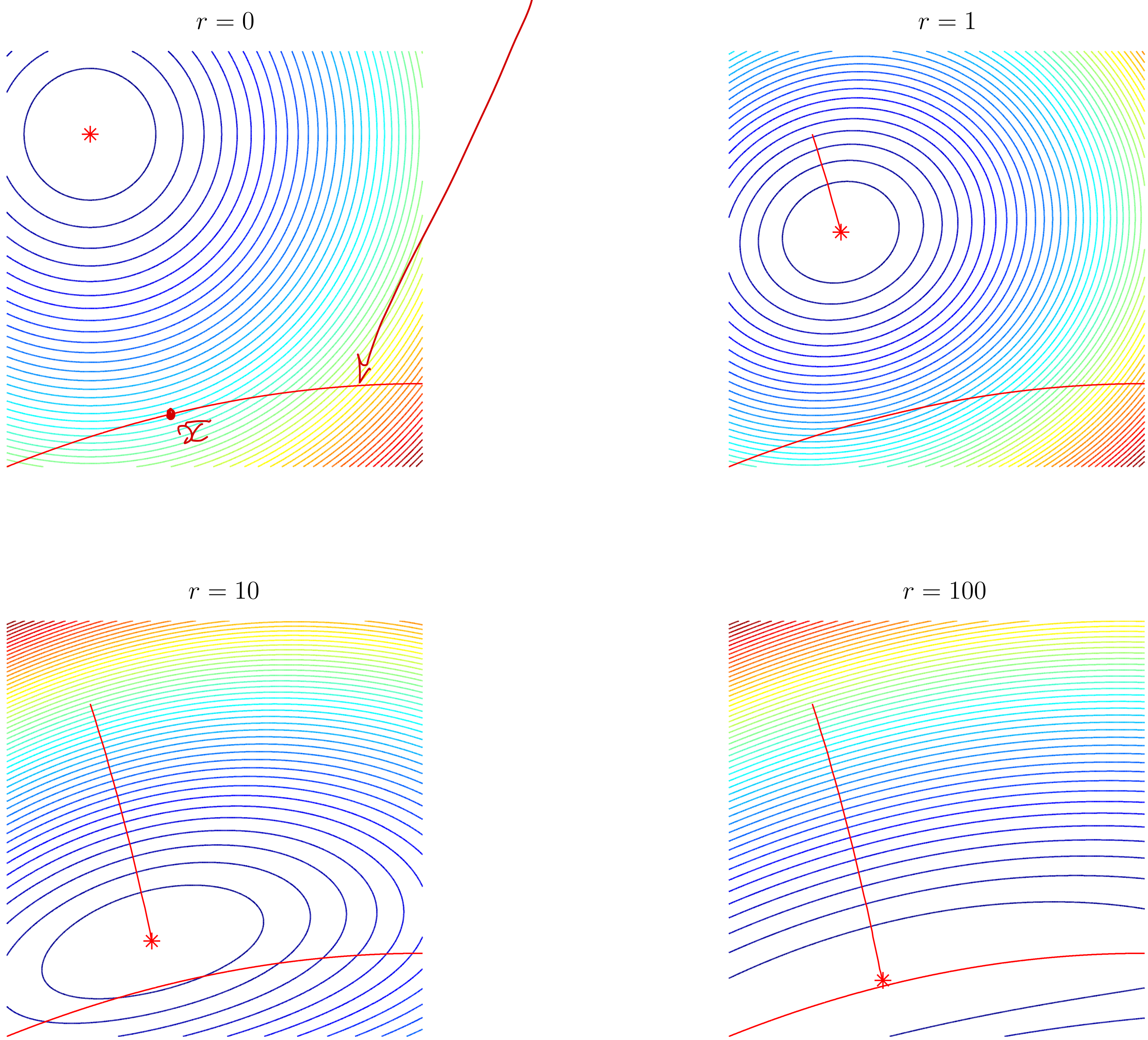
$$\begin{cases} \min f(x) \\ c(x) = 0 \end{cases}$$

- $\theta_r(x) = f(x) + \frac{r}{2}\|c(x)\|_2^2$
- $D\theta_r(x) = Df(x) + r\, c'(x)^T c(x)$
- $D^2\theta_r(x) = D^2 f(x) + \sum_i \underbrace{r\, c_i(x)}_{\substack{\text{multiplicateur optimal} \\ \text{(voir ci-après)}}} D^2 c_i(x) + \underbrace{r\, c'(x)^T c'(x)}_{\text{explose avec } r}$

$\Rightarrow K_2(D^2\theta_r(x_r)) \nearrow \infty$

figure

$$\begin{cases} \min f(x) \\ c(x) = 0 \end{cases} \rightsquigarrow \min_{x \in \mathbb{R}^2} f(x) + r \|c(x)\|_2^2$$



Chemin des minimiseurs

Prop (récupération d'un multiplicateur optimal)

$$(P_E)\begin{cases}\inf f(x)\\ c(x)=0\end{cases}\qquad \theta_r(x)=f(x)+\frac{r}{2}\|c(x)\|_2^2$$

Si
- $\nabla_x\theta(\bar{x}_r)=0$
- $\bar{x}_r\to\bar{x}$ lorsque $r\to\infty$
- $c'(\bar{x})$ surjective

alors
1) $c(\bar{x})=0$
2) $\exists\,\bar{\lambda}\in\mathbb{R}^m$: $\nabla_x\ell(\bar{x},\bar{\lambda})=0$
3) $r\,c(\bar{x}_r)\to\bar{\lambda}$

Dém

1) $\nabla_x\theta_r(\bar{x}_r)=0=\underbrace{\frac{1}{r}}_{\to 0}\nabla f(\bar{x}_r)+\cancel{r}\sum_i c_i(\bar{x}_r)\nabla c_i(\bar{x}_r)$

$\frac{1}{r}$ + $\lim_{r\to\infty}$

$\Rightarrow \sum_i c_i(\bar{x})\underbrace{\nabla c_i(\bar{x})}_{LI}=0$

$\Rightarrow c(\bar{x})=0$

2) $\nabla f(\bar{x}_r)+\sum_i\underbrace{(r\,c_i(\bar{x}_r))}_{}\nabla c_i(\bar{x}_r)=0$

$\downarrow$ $\qquad\downarrow$ néc $\qquad\downarrow$

$\nabla f(\bar{x})+\sum_i$ q-ch. $\underbrace{\nabla c_i(\bar{x})}_{LI}=0$

$\;\;\;\;\;\;\; \overset{!!}{:}\;\bar{\lambda}$

$\Rightarrow \nabla_x\ell(\bar{x},\bar{\lambda})=0$

$\square$

3) Pénalisation intérieure

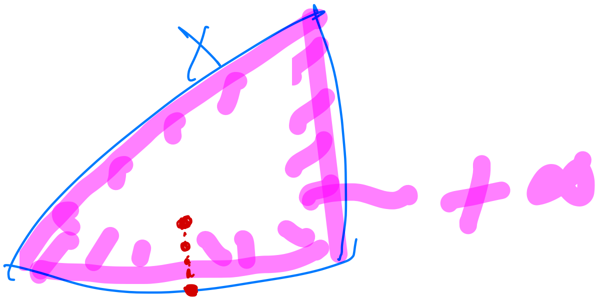
Exemple (rien de plus...)

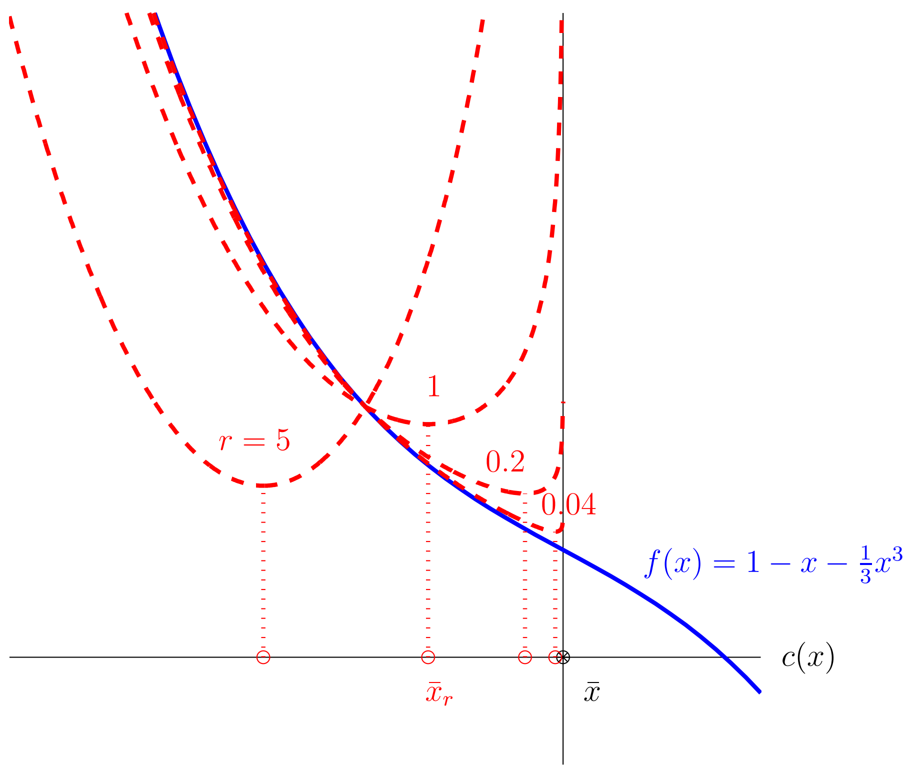
$$\begin{cases} \min f(x) \\ c(x) \leq 0 \end{cases} \rightsquigarrow \min_{c(x)<0} f(x) - \mu \sum_i \log(-c_i(x))$$

$\mu \downarrow 0$

on reste à l'intérieur de l'ensemble admissible

barrière logarithmique ($\to +\infty$ si x s'approche du bord de l'ensemble admissible)





Pénalisation logarithmique

4) Lagrangien augmenté

[A] Objectifs

On cherche une fonction de pénalisation telle que

- on ne doit plus prendre $r \nearrow \infty$ (donc on évite le mauvais conditionnement)
- pour un r fini, θ_r a une solution en une solution de (P_X), c-à-d θ_r est une fonction de pénalisation exacte

[B] Exactitude du Lagrangien

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

prop

si
- (P_{EI}) est convexe (f, c_I convexes, c_E affine)
- $\bar{x}$ solution de (P_{EI})
- f et c sont dérivables en $\bar{x}$
- $\exists\, \bar{\lambda}$ tel que $(\bar{x}, \bar{\lambda})$ vérifie KKT

alors $x \longmapsto \ell(x, \bar{\lambda})$ a un min global en $\bar{x}$

Dem (déjà vu dans le cours sur la dualité)

$$\ell(\bar{x}, \bar{\lambda}) \leq \ell(x, \bar{\lambda}), \quad \forall x \in E$$

Deux inconvénients

① $\ell(\cdot, \bar{\lambda})$ est pas convexe si (P_{EI}) ne l'est pas

exemple $\begin{cases} \min & -x^2 \\ x = 0 \end{cases}$ $\longrightarrow (\bar{x}, \bar{\lambda}) = (0,0)$

$\ell(x, \bar{\lambda}) = -x^2$ n'est pas convexe !

② il faudrait connaître $\bar{\lambda}$

→ dans cette approche, il faut résoudre une suite de problème

$$\inf_{x \in E} \ell(x, \lambda)$$

avec mise à jour de $\lambda \to \bar{\lambda}$

→ c'est la relaxation lagrangienne (voir cours sur la dualité)

C Le lagrangien augmenté de (P_E)

L'idée est la suivante $\qquad (P_E)\begin{cases}\min f(x)\\ c(x)=0\end{cases}$

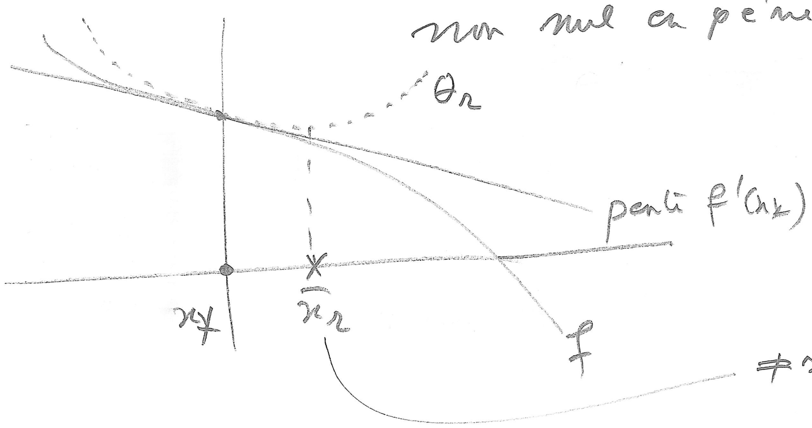
- La pénalisation quadratique

$$\Theta_r(x) = f(x) + \frac{r}{2}\|c(x)\|_2^2$$

ne peut pas être exacte (en général) car en une solution x_* de (P_E), on a

$$\nabla\Theta_r(x_*) = \nabla f(x_*) + r\, c'(x_*)^T \underbrace{c(x_*)}_{=0}$$

$$= \underbrace{\nabla f(x_*)}_{\text{non nul en général}}.$$



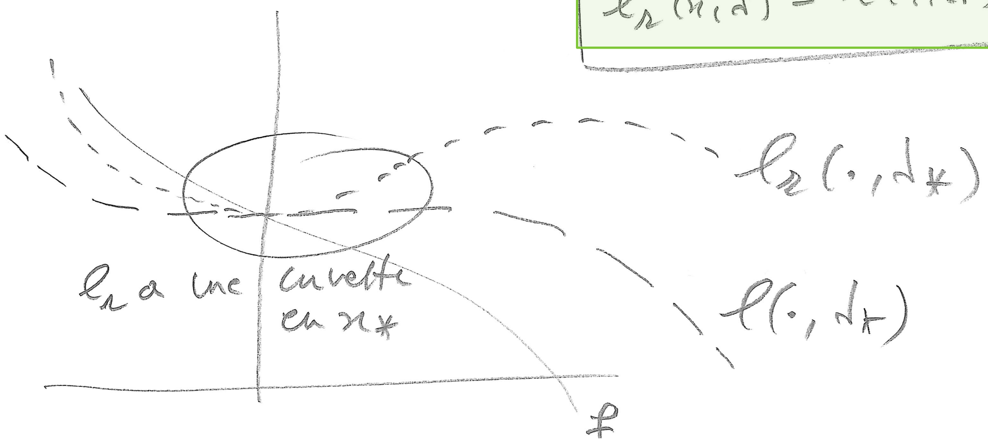
Contrainte $x = 0$

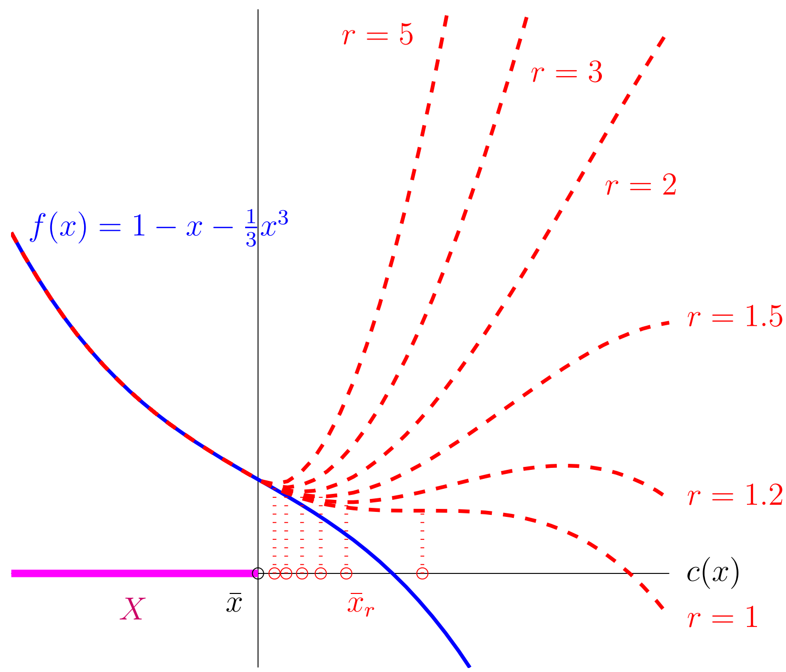
$\neq x_*$ car $\Theta'(x_*) = f'(x_*) \neq 0$

- Le lagrangien augmenté "redresse" d'abord f pour que la "fonction redressée" ait une pente nulle en x_*, puis on pénalise. La fonction redressée est le lagrangien, qui a en effet une pente nulle en x_*

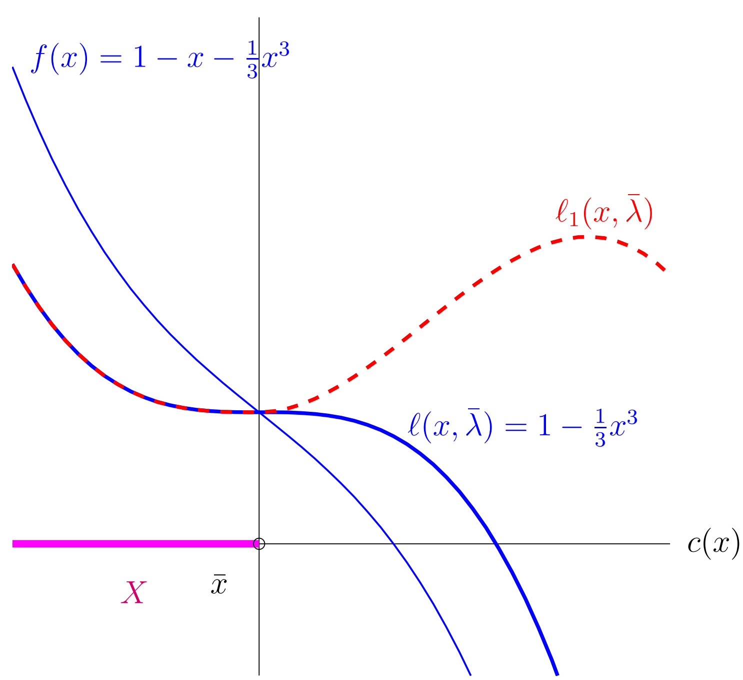
$$\ell_r : \mathbb{E}\times\mathbb{F} \to \mathbb{R}$$

$$\boxed{\ell_r(x,\lambda) = \ell(x,\lambda) + \frac{r}{2}\|c(x)\|^2}$$





Pénalisation quadratique



Pénalisation quadratique du lagrangien

Lemme de FINSLER (1937)

Si
- M et $A \in S^n$ (matrices symétriques d'ordre n)
- $A \succcurlyeq 0$ (semi-définie positive)
- $M \succ 0$ sur $N(A)$, c-à-d

$$v^T M v > 0, \quad \forall v \in N(A) \setminus \{0\}$$

alors $\exists \underline{r}$, $\forall r \geq \underline{r}$, $M + rA \succ 0$

Dém (par l'absurde)

$\Rightarrow$ $\exists\, r_k \nearrow \infty$: $M + r_k A \not\succ 0$

$\exists\, v_k \neq 0$: $v_k^T (M + r_k A)\, v_k \leq 0$

- on peut supposer que $\|v_k\| = 1$
- Soit $v_k \to v$ (sous-suite)
- $Av = 0$ $\quad \left(\begin{array}{l} v^T A v = 0 \\ A \succcurlyeq 0 \end{array} \right) \Rightarrow Av = 0$

$$\underbrace{\frac{v_k^T M v_k}{r_k}}_{\to\, 0} + \cancel{r_k}\, \underbrace{v_k^T A v_k}_{\to\, v^T A v} \leq 0 \quad \Rightarrow \quad v^T A v = 0$$

- $v^T M v \leq 0$

$$v_k^T M v_k + \underbrace{r_k\, v_k^T A v_k}_{\geq 0} \leq 0$$

$$\Rightarrow v_k^T M v_k \leq 0 \Rightarrow v^T M v \leq 0$$

- $v \neq 0$

$\Rightarrow$ Contradiction

$\square$

Théo (exactitude du lagrangien augmenté)

Si
- f et c sont 2 fois dérivable en $\bar{x}$
- $\bar{x}$ minimum local de (P_E)
- $\exists \bar{\lambda}$ tel que $\nabla_x \ell(\bar{x}, \bar{\lambda}) = 0$ et CS2

alors $\exists \underline{r}$, $\forall r \geq \underline{r}$, $\bar{x}$ est un minimum local strict de $\ell_r(\cdot, \bar{\lambda})$.

Dém

$$\ell_r(x, \lambda) = f(x) + \lambda^T c(x) + \frac{r}{2} \|c(x)\|_2^2$$

$$\nabla_x \ell_r(x, \lambda) = \nabla f(x) + c'(x)^T \lambda + r\, c'(x)^T c(x)$$

$$\nabla_x \ell_r(\bar{x}, \bar{\lambda}) = 0$$

$$\nabla^2_{xx} \ell_r(\bar{x}, \bar{\lambda}) = \underbrace{\nabla^2_{xx} \ell(\bar{x}, \bar{\lambda})}_{M} + r\, \underbrace{c'(\bar{x})^T c'(\bar{x})}_{A}$$

$M > 0$ sur $N(A)$ par la CS2

Finsler $\Rightarrow$ pour r assez grand, $\nabla^2_{xx} \ell_r(\bar{x}, \bar{\lambda}) > 0$

$\square$

Remarques

1) On ne doit pas faire tendre $r \nearrow \infty$

2) On doit toujours trouver une suite de $\lambda_k \to \bar{\lambda}$

$\longrightarrow$ c'est la relaxation lagrangienne augmentée

D Méthode des multiplicateurs pour (P_E)

On passe de $(\lambda_k, r_k) \to (\lambda_{k+1}, r_{k+1})$ par

1) on minimise $\ell_{r_k}(\cdot, \lambda_k) \longrightarrow x_k$
2) si (x_k, λ_k) est satisfaisant $\longrightarrow$ arrêt
3) $\lambda_{k+1} = \lambda_k + r_k\, c(x_k)$
4) $r_{k+1} \nearrow$ si nécessaire (heuristique)

D'où vient la formule de λ_{k+1} ?

- c'est « heuristique » (elle se justifie très bien pour les problèmes convexes)

$$0 = \nabla_x \ell_{r_k}(x_k, \lambda_k) \quad \text{par l'étape 1}$$

$$= Df(x_k) + c'(x_k)^T \underbrace{(\lambda_k + r_k\, c(x_k))}_{= \lambda_{k+1}}$$

λ_{k+1} car on souhaite que $\nabla_x \ell(\cdot, \lambda_{k+1}) = 0$

- si (P_{EI}) convexe, on montre que l'algorithme est une méthode proximale sur la fonction duale