NLP ANALYTICAL METHODS

PR. RAFIC YOUNES

Plan

- Introduction
- Fonction à une seule variable
- Fonction multi-variables
- PNL avec Contraintes Égalités
- Récapitulatif

Introduction

- Problème :

$$\min f(x)$$
$$\text{s.c. } x \in X$$
$$h_i(x) = 0 \text{ où } i=1,\dots,m$$
$$g_j(x) \leq 0 \text{ où } i=1,\dots,q$$

- $f: IR^n \rightarrow IR$, $h: IR^n \rightarrow IR^m$, $g: IR \rightarrow IR^q$ continues
- X ensemble fermé
- Intérieur des contraintes d'inégalité :

$$S = \{x \in X \text{ t.q. } g(x) < 0\}$$

Introduction

- 1er Problème

Fonction objectif seule :

$$\min f(x)$$

$$f : IR^n \rightarrow IR$$

Plan

- Introduction
- Fonction à une seule variable
- Fonction multi-variables
- PNL avec Contraintes Égalités
- Récapitulatif

Fonctions à une variable

$$\min f(x),\ x \in IR$$

Définitions :

- x^* est un **maximum local** de f s'il existe a > 0 tel que $f(x^*) \geq f(x)$ pour tout $x \in \,]x^*-a, x^*+a[$

- x^* est un **minimum local** de f s'il existe a > 0 tel que $f(x^*) \leq f(x)$ pour tout $x \in \,]x^*-a, x^*+a[$

- L'intervalle $]x^*-a, x^*+a[$ est appelé un **voisinage** de x^*

Fonctions à une variable

x^* est un **maximum local strict** de f s'il existe $a > 0$ tel que : $f(x^*) > f(x)$ pour tout $x \in \,]x^*-a, x^*+a[$

x^* est un **minimum local strict** de f s'il existe $a > 0$ tel que : $f(x^*) < f(x)$ pour tout $x \in \,]x^*-a, x^*+a[$

x^* est un **maximum global** de $f(x)$ si :

$$f(x^*) \geq f(x) \text{ pour tout } x \in IR$$

x^* est un **minimum global** de $f(x)$ si :

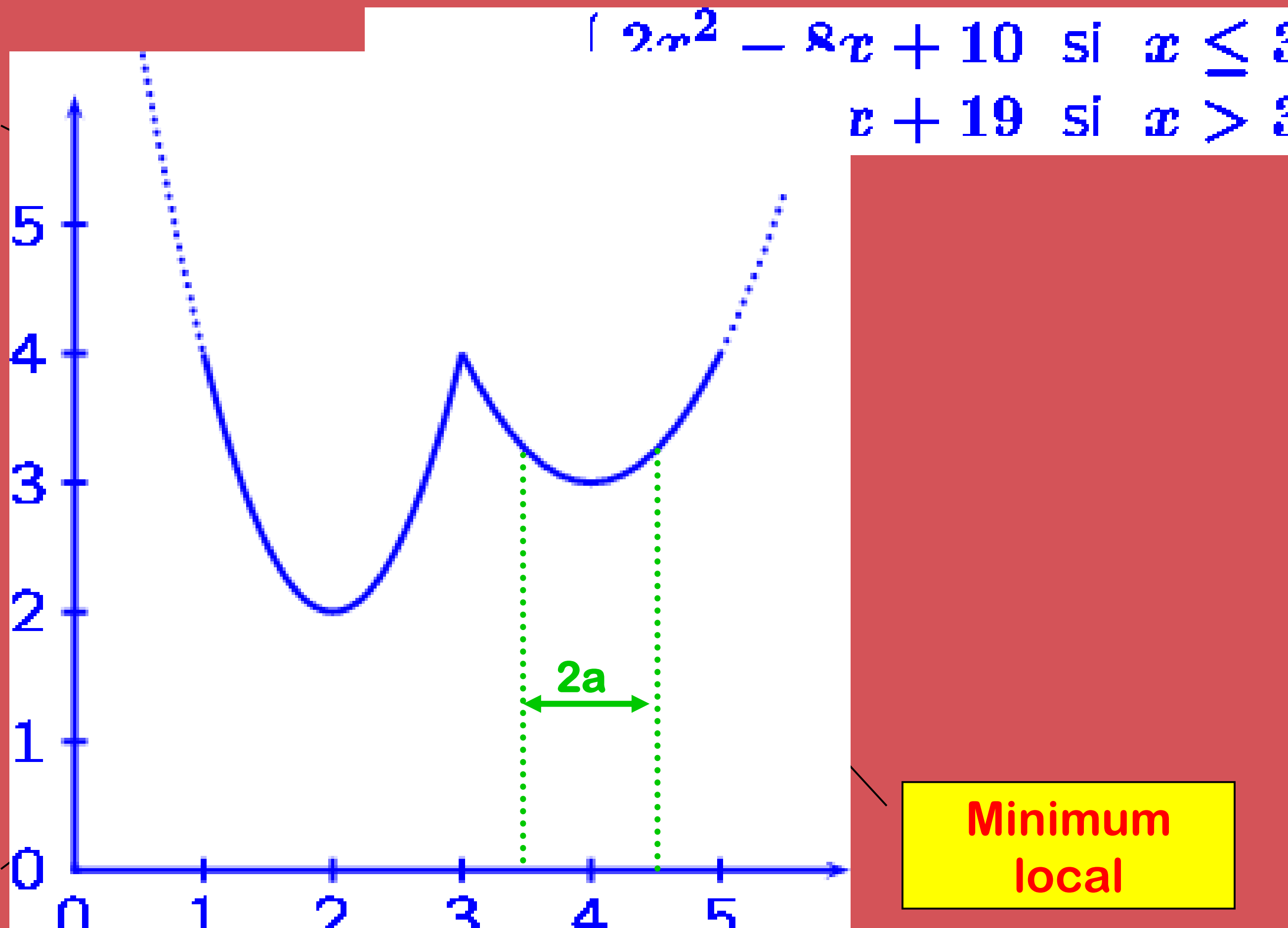
$$f(x^*) \leq f(x) \text{ pour tout } x \in IR$$

extremum : un minimum ou un maximum.

Fonctions à une variable

Maximum local

Minimum global



Minimum local

Fonctions à une variable

- Un point x où la tangente est horizontale, c'est-à-dire tel que $f'(x)=0$, est appelé un **point critique** ou **point stationnaire**.

Théorème de Fermat :

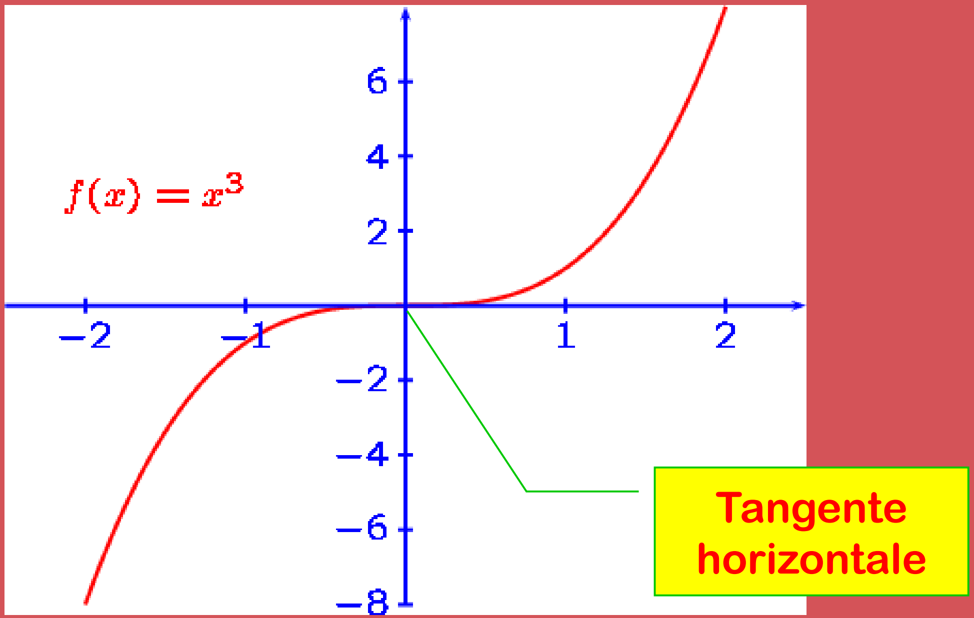
- Si une fonction continue f possède un extre mum local en x^*, et si $f'(x^*)$ existe, alors :

$$f'(x^*) = 0.$$

Fonctions à une variable

La condition $f'(x^*) = 0$ est une **condition nécessaire d'optimalité** pour une fonction différentiable.

Attention :
ce n'est pas une condition suffisante.
Rappel:
Si $P \Rightarrow Q$, alors P est suffisante et Q est nécessaire.



x^* optimal $\Rightarrow f'(x^*) = 0$

mais pas un maximum, ni un minimum…

Fonctions à une variable

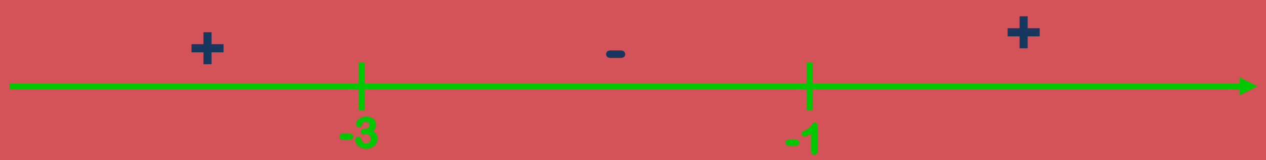
Test de premier ordre :

- Soit une fonction différentiable f, et x* un point critique. S'il existe a > 0 tel que
  - $f'(x) > 0$ si $x^*-a < x < x^*$
  - $f'(x) < 0$ si $x^* < x < x^*+a$
- Alors x* est un maximum local de f.

- Soit une fonction différentiable f, et x* un point critique. S'il existe a > 0 tel que
  - $f'(x) < 0$ si $x^*-a < x < x^*$
  - $f'(x) > 0$ si $x^* < x < x^*+a$
- Alors x* est un minimum local de f.

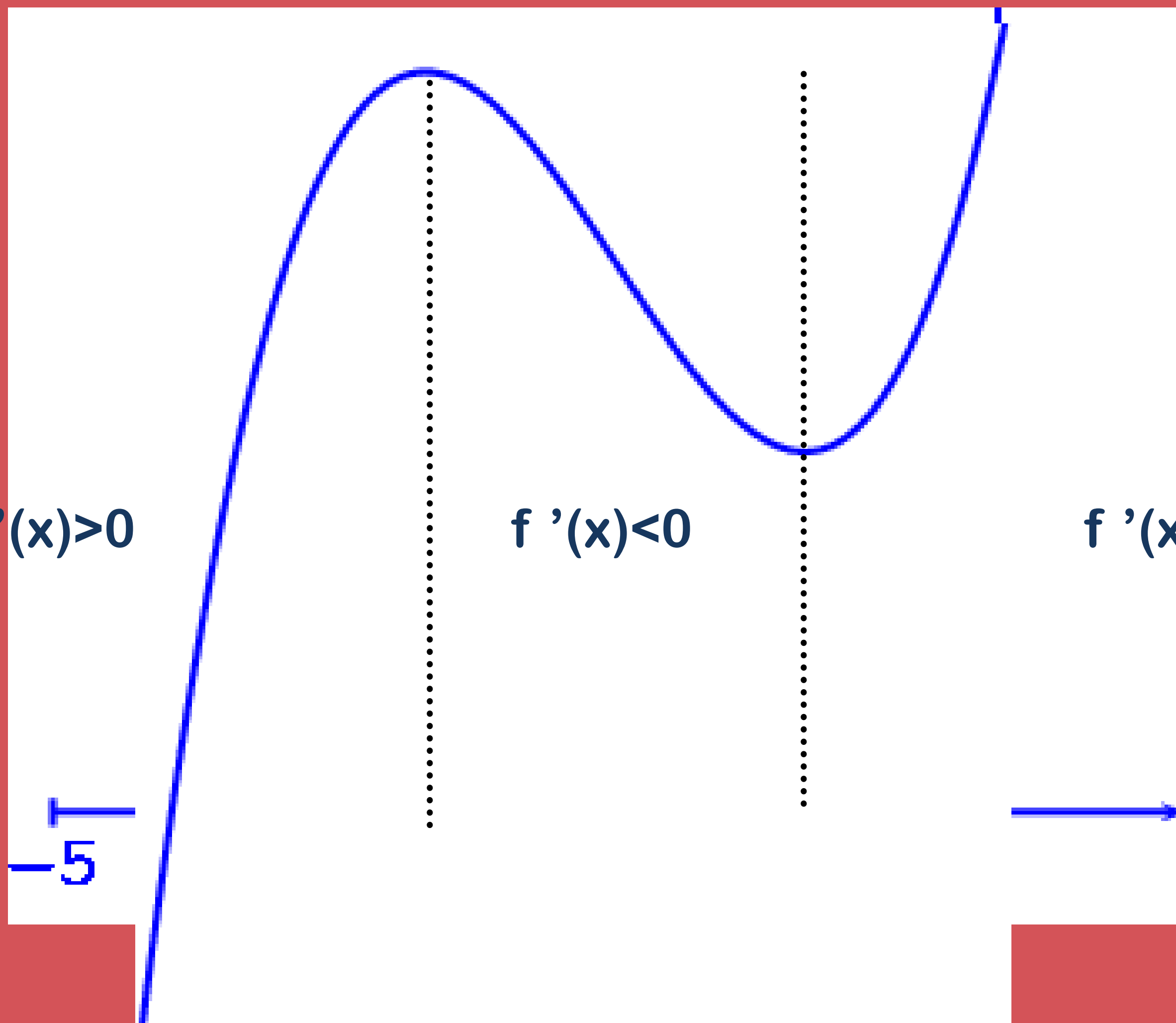
Fonctions à une variable

- Soit $f(x) = x^3 + 6x^2 + 9x + 8$
- $f'(x) = 3x^2 + 12x + 9 = 3(x+3)(x+1)$
- Points critiques : $x_1 = -3$ et $x_2 = -1$
- Signes de $f'(x)$:



- x_1 maximum local
- x_2 minimum local

Fonctions à une variable



f ’(x)>0 **f ’(x)<0** **f ’(x)>0**

Fonctions à une variable

Test de premier ordre :

- Condition suffisante d'optimalité d'un point critique.

- Inconvénient : il faut de l'information sur f(x) à d'autres points que le point critique.

Fonctions à une variable

Test de second ordre :

- Si la fonction f possède une dérivée seconde continue dans un voisinage d'un point critique x, alors :
  - $f''(x^*) < 0$ est une condition suffisante pour que x^* soit un maximum local, et
  - $f''(x^*) > 0$ est une condition suffisante pour que x^* soit un minimum local.

Fonctions à une variable

- Soit $f(x) = x^3 + 6x^2 + 9x + 8$
- $f'(x) = 3x^2 + 12x + 9 = 3(x+3)(x+1)$
- Points critiques : $x_1 = -3$ et $x_2 = -1$

$f''(x) = 6x + 12$

- $f''(-3) = -6 < 0 \Rightarrow$ x_1 maximum local
- $f''(-1) = 6 > 0 \Rightarrow$ x_2 minimum local

Plan

- Introduction
- Fonction à une seule variable
- Fonction multi-variables
- PNL avec Contraintes Égalités
- Récapitulatif

Fonctions multivariables

Rappels :

- Soit $f:\mathbb{IR}^n \to \mathbb{IR}$ deux fois continûment différentiable sur une sphère ouverte S centrée en x.
- Pour tout d tel que $x+d \in S$, il existe $0 \leq \varepsilon \leq 1$ tel que :
- $f(x+d)=f(x) + d^T\nabla f(x) + \frac{1}{2}\, d^T\nabla^2 f(x+\varepsilon d)\, d.$
- Pour tout d tel que $x+d \in S$,

  $\mathbf{f(x+d) = f(x) + d^T\nabla f(x) + \frac{1}{2}\, d^T \nabla^2 f(x)\, d + o(\lVert d \rVert^2)}$

Fonctions multivariables

Conditions nécessaires d'optimalité

- Soit x^* un minimum local de $f: IR^n \rightarrow IR$. Si f est continûment différentiable sur un ouvert S contenant x^*, alors

$$\nabla f(x^*)=0.$$

- Si, de plus, f est deux fois continûment différentiable sur S, alors

$$\nabla^2 f(x^*) \text{ est semi définie positive}$$

$$d^T \nabla^2 f(x^*)\, d \geq 0 \text{ pour tout } d \in IR^n$$

Fonctions multivariables

Preuve :

- Soit une direction arbitraire $d \in IR^n$.
- Considérons la fonction $g(\alpha)=f(x^*+\alpha d)$

$$\frac{dg}{d\alpha}(0) = d^T \nabla f(x^*)$$

$$\frac{dg}{d\alpha}(0) = \lim_{\alpha \downarrow 0} \frac{f(x^* + \alpha d) - f(x^*)}{\alpha} \geq 0$$

et donc $d^T \nabla f(x^*) \geq 0$

Fonctions multivariables

- Même raisonnement avec –d
- Considérons la fonction g(α)=f(x*-αd)

$$\frac{dg}{d\alpha}(0) = -d^T \nabla f(x^*)$$

$$\frac{dg}{d\alpha}(0) = \lim_{\alpha \downarrow 0} \frac{f(x^* - \alpha d) - f(x^*)}{\alpha} \geq 0$$

et donc $-d^T \nabla f(x^*) \geq 0$ **ou** $d^T \nabla f(x^*) \leq 0$

- **Comme $d^T \nabla f(x^*) = 0$ pour tout d, on a bien $\nabla f(x^*) = 0$**

Fonctions multivariables

- Si f est deux fois continûment différentiable, on choisit une direction d arbitraire. Pour tout $\alpha \in$ IR, le développement en série de Taylor donne:

$$\begin{aligned} f(x^* + \alpha d) - f(x^*) &= \alpha \nabla f(x^*)^T d \\ &+ \frac{\alpha^2}{2} d^T \nabla^2 f(x^*) d \\ &+ o(\alpha^2) \end{aligned}$$

Fonctions multivariables

- Comme $\nabla f(x^*) = 0$, on obtient pour α suffisamment petit

$$0 \leq \frac{f(x^*+\alpha d)-f(x^*)}{\alpha^2}$$

$$= \frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{o(\alpha^2)}{\alpha^2}$$

- **Si $\alpha \to 0$, on obtient**

$$d^T \nabla^2 f(x^*) d \geq 0$$ **cqfd**

Fonctions multivariables

Conditions suffisantes d'optimalité

- Soit f: $IR^n \rightarrow IR$ une fonction deux fois continûment différentiable sur un ouvert S. Si $x^* \in S$ satisfait les conditions

$$\nabla f(x^*)=0$$

et

$\nabla^2 f(x^*)$ est définie positive

$d^T \nabla^2 f(x^*)\, d > 0$ pour tout $d \in IR^n$, $d \neq 0$

Alors x^* est un **minimum local strict** de f

Plan

- Introduction
- Fonction à une seule variable
- Fonction multi-variables
- PNL avec Contraintes Égalités
- Récapitulatif

PNL avec Contraintes Egalités

- Problème :

$$\min f(x)$$
$$\text{s.c. } x \in X$$
$$h_i(x) = 0 \text{ où } i=1,\ldots,m$$
$$g_j(x) \leq 0 \text{ où } i=1,\ldots,q$$

- $f: IR^n \rightarrow IR$, $h: IR^n \rightarrow IR^m$, $g: IR \rightarrow IR^q$ continues
- X ensemble fermé
- Intérieur des contraintes d'inégalité :

$$S = \{x \in X \text{ t.q. } g(x) < 0\}$$

PNL avec Contraintes Egalités

- Problème :

$$\min f(x)$$

$$h_i(x) = 0 \text{ où } i=1,\ldots,m$$

- $f: IR^n \rightarrow IR$, $h: IR^n \rightarrow IR^m$,

Introduction

Idée : transformer le problème avec contraintes en un problème sans contrainte.

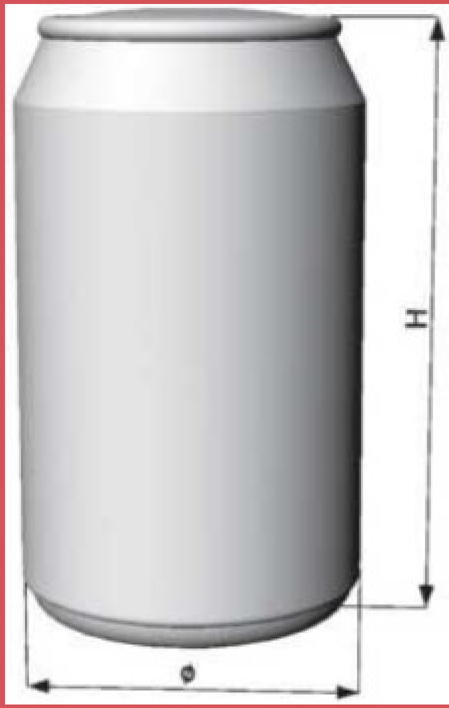
- Plusieurs approches :
  - Par Élimination.
  - Fonction de Pénalité.
  - Théorie des multiplicateurs de Lagrange.
  - Lagrangien Pénalisé.
  - … …

PNL avec Contraintes Egalités

- **Méthode d'Élimination**
- Fonction de Pénalité
- Théorie de Lagrange
- Lagrangien Pénalisé

Par Élimination

- Les contenants (canettes) ont une forme cylindrique de hauteur h et de rayon r.
- Afin de réduire, on veut minimiser la surface d'aluminium nécessaire à la construction des contenants.
- Cependant, ils doivent s'assurer qu'un contenant ait un volume de 128π cm^3.
- Trouver les dimensions qui réalise nt l'objectif et satisfait la contrainte ?



Par Élimination

- Objectif : Min $S(r,h) = 2\pi r^2 + 2\pi r h$
- Contrainte : volume = $\pi r^2 h = 128\pi$

$$h = \frac{128\pi}{\pi r^2} \Rightarrow h = \frac{128}{r^2}$$

- Substituer ces expressions dans l'objectif:

$$S(r,h) = 2\pi r^2 + 2\pi r h$$

$$S(r) = 2\pi r^2 + 2\pi r\left(\frac{128}{r^2}\right)$$

$$S(r) = 2\pi r^2 + \frac{256\pi}{r}$$

Par Élimination

- Un point optimale est obtenu lorsque :

$$S'(r) = 0$$

$$S'(r) = \left(2\pi r^2 + \frac{256\pi}{r}\right)' \Rightarrow S'(r) = 4\pi r - \frac{256\pi}{r^2}$$ $$\Rightarrow r = 4$$

- Il faut s'assurer qu'on retrouve bien un minimum en ce point : $$S''(r) = \left(4\pi r - \frac{256\pi}{r^2}\right)' \Rightarrow S''(r) = 4\pi + \frac{512\pi}{r^3}$$

- Pour r > 0, la dérivée seconde est toujours positive. La surface est minimisée lorsque :

r = 4 et h = 8.

Par Élimination

- On prouve uniquement le cas linéaire :

(P1) $\min f(x)$ s.c. $Ax=b$

$A \in IR^{mxn}$ de rang plein, $b \in IR^m$

- On suppose sans perte de généralité que les m premières colonnes de A sont linéairement indépendants.

$A = (B\ R)$ avec $B \in IR^{mxm}$ inversible, $R \in IR^{mx(n-m)}$

$x = (x_B\ x_R)^T$ avec $x_B \in IR^m$ $x_R \in IR^{mx(n-m)}$

(P2) $\min f(x_B,x_R)$ s.c. $Bx_B+Rx_R=b$

- Pour vérifier les contraintes, il faut que

$$x_B = B^{-1}(b-Rx_R)$$

Par Élimination

(P3) min $F(x_R)=f(B^{-1}(b-Rx_R), x_R)$ s.c. $x_R \in IR^{mx(n-m)}$

- Si (x_B^*,x_R^*) est solution optimale de (P1), x_R^* est solution optimale de (P3), problème sans contrainte.

$$0 = \nabla F(x_R^*) = -R^T B^{-T} \nabla_B f(x^*) + \nabla_R f(x^*)$$

  où
  - $\nabla_B f$ est le gradient de f par rapport à x_B
  - $\nabla_R f$ est le gradient de f par rapport à $x_{R.}$

Par Élimination

- $0 = \nabla F(x_R^*) = -R^T B^{-T} \nabla_B f(x^*) + \nabla_R f(x^*)$
- Posons $\lambda^* = -B^{-T} \nabla_B f(x^*)$
- Donc :

$$B^T\lambda^* + \nabla_B f(x^*) = 0$$
$$R^T\lambda^* + \nabla_R f(x^*) = 0$$

- c'est-à-dire

$$A^T\lambda^* + \nabla f(x^*) = 0$$

- C'est la condition nécessaire du premier ordre pour le cas linéaire.
- La condition du second ordre est admise sans preuve.

PNL avec Contraintes Egalités

- Méthode d'Élimination
- **Fonction de Pénalité**
- Théorie de Lagrange
- Lagrangien Pénalisé

Par pénalité

- Soit le problème P

$$\min f(x)$$
$$\text{s.c. } h(x) = 0.$$

- Pour k=1,2,3..., on définit

$$F^k(x) = f(x) + \frac{k}{2}\|h(x)\|^2$$

où

- **x^* est minimum local de P.**
- **$\alpha \in$ IR, $\alpha > 0$**

Par pénalité

- Le terme $\frac{k}{2}\|h(x)\|^2$ **est le terme de pénalité**

- La suite $\nabla F^k(x^k)$ converge donc vers x^*

$$\begin{aligned} 0 &= \nabla F^k(x^k) \\ &= \nabla f(x^k) + k\nabla h(x^k)h(x^k) \end{aligned}$$

- x^k est à l'intérieur de S pour k suffisamment grand.
- x^k est solution du problème sans contraintes pour k suffisamment grand.

PNL avec Contraintes Egalités

- Méthode d'Élimination
- Fonction de Pénalité
- **Théorie de Lagrange**
- Lagrangien Pénalisé

Lagrange

- Joseph Louis Lagrange
- Né à Turin en 1736
- Mort à Paris 1813

Définition :

- Soit $L: IR^{n+m} \rightarrow IR$

$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$$

Fonction lagrangienne

- Condition du premier ordre

$$\nabla_x L(x^*,\lambda^*) = 0$$

- Contraintes

$$\nabla_\lambda L(x^*,\lambda^*) = 0$$

- Condition du second ordre

$$y^T \nabla_{xx}^2 L(x^*,\lambda^*) y \geq 0 \quad \forall y \in V(x^*)$$

Fonction lagrangienne

Exemple :

$$\min \tfrac{1}{2}\,(x_1^2+x_2^2+x_3^2)$$

$$\text{s.c. } x_1+x_2+x_3=3$$

$L(x_1,x_2,x_3,\lambda)= \tfrac{1}{2}(x_1^2+x_2^2+x_3^2)+\lambda(x_1+x_2+x_3-3)$

$\partial L/\partial x_i = x_i+\lambda \qquad i=1,2,3$

$\partial L/\partial \lambda= x_1+x_2+x_3-3$

Conditions suffisantes

Conditions suffisantes du second ordre

- Soient $f: IR^n \rightarrow IR$ et $h: IR^n \rightarrow IR^m$ deux foix continûment différentiables.
- Soient $x^* \in IR^n$ et $\lambda^* \in IR^m$ tels que
  - $\nabla_x L(x^*, \lambda^*) = 0$
  - $\nabla_\lambda L(x^*, \lambda^*) = 0$
  - $y^T \nabla^2_{xx} L(x^*, \lambda^*) y > 0 \ \forall y$ tel que $\nabla h(x^*)^T y = 0$
- Alors x^* est une minimum local strict de f s.c. $h(x)=0$.

PNL avec Contraintes Egalités

- Méthode d'Élimination
- Fonction de Pénalité
- Théorie de Lagrange
- **Lagrangien Pénalisé**

Lagrangien pénalisé

- Soit le problème

$$\min f(x)$$
$$\text{s.c. } h(x) = 0,\ x \in X$$

avec $f: IR^n \to IR$, $h: IR^n \to IR^m$ et $X \subseteq IR^n$.

- En général, on aura $X = IR^n$.
- On considérera x^* et λ^* qui vérifient les conditions suffisantes d'optimalité du second ordre.

Lagrangien pénalisé

Définition

- Soit $c \geq 0$. La fonction $L_c : IR^n x IR^m \to IR$

$$L_c(x,\lambda) = f(x) + \lambda^T h(x) + (c/2) \|h(x)\|^2$$

est appelée le Lagrangien pénalisé de f.

Lagrangien pénalisé

Idée

- Deux mécanismes, basés sur la minimisation sans contraintes de $L_c(.,\lambda)$, donnent des points proches de x*.

1. Rendre λ le plus proche possible de λ^*.

   Car x* est un minimum local strict de $L_c(.,\lambda^*)$.

2. Rendre c très grand.

   Le coût des points non admissibles sera alors très élevé, et le minimum de $L_c(.,\lambda)$ sera quasi-admissible. De plus, si x est quasi-admissible, on a $L_c(x,\lambda) \approx f(x)$.

Exemple

$$\min f(x) = \tfrac{1}{2}\ (x_1^2 + x_2^2)$$
$$\text{s.c. } x_1 = 1$$

- Solution optimale : $x^* = (1\ 0)^T$
- Multiplicateur : $\lambda^* = -1$
- En effet

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = x^*_1 + x^*_2 + \lambda^* = 0$$

- On a

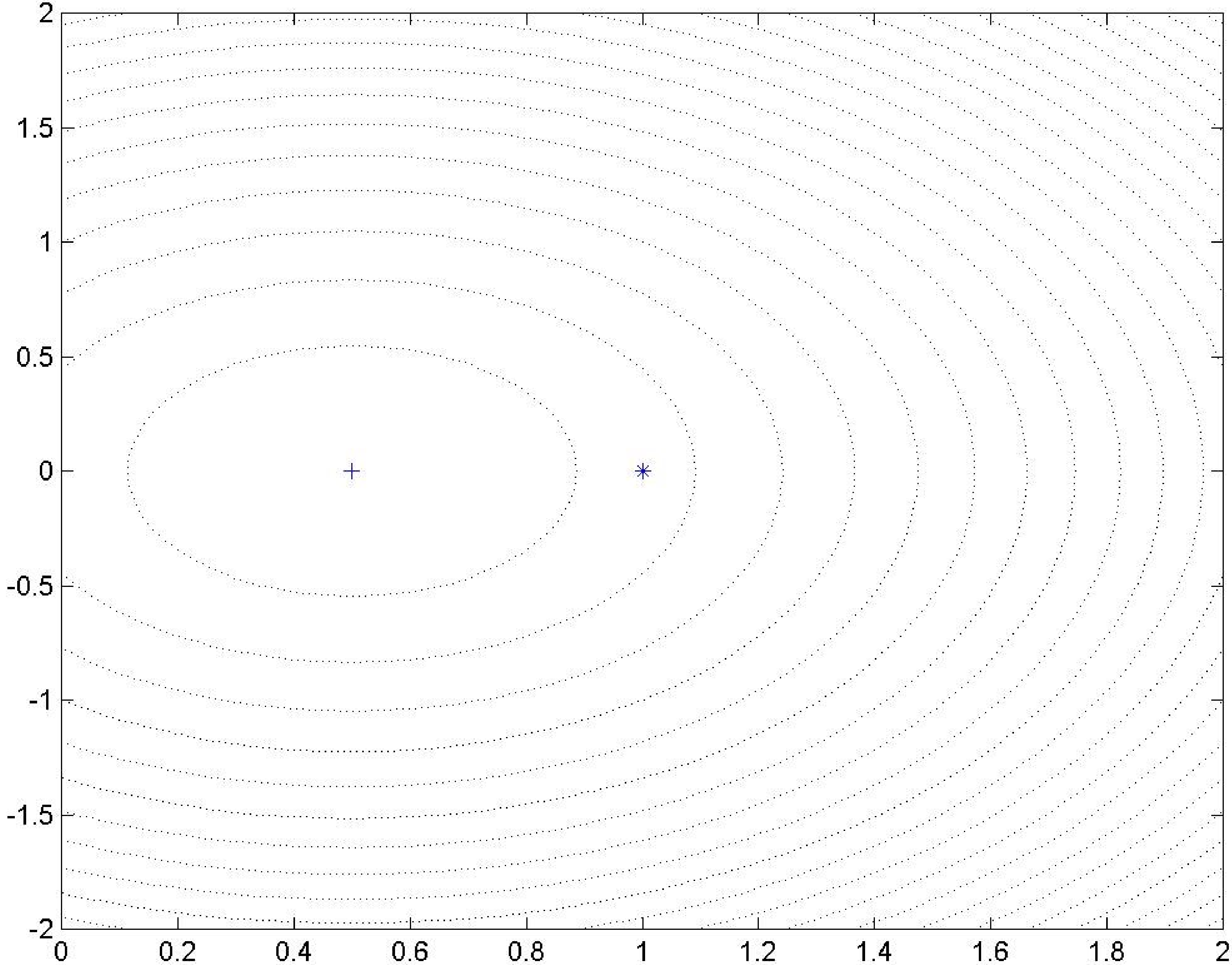
$$L_c(x,\lambda) = \tfrac{1}{2}(x_1^2 + x_2^2) + \lambda(x_1 - 1) + (c/2)(x_1 - 1)^2$$

Exemple

$$L_c(x,\lambda) = ½(x_1^2 + x_2^2) + \lambda(x_1-1)+(c/2)(x_1-1)^2$$

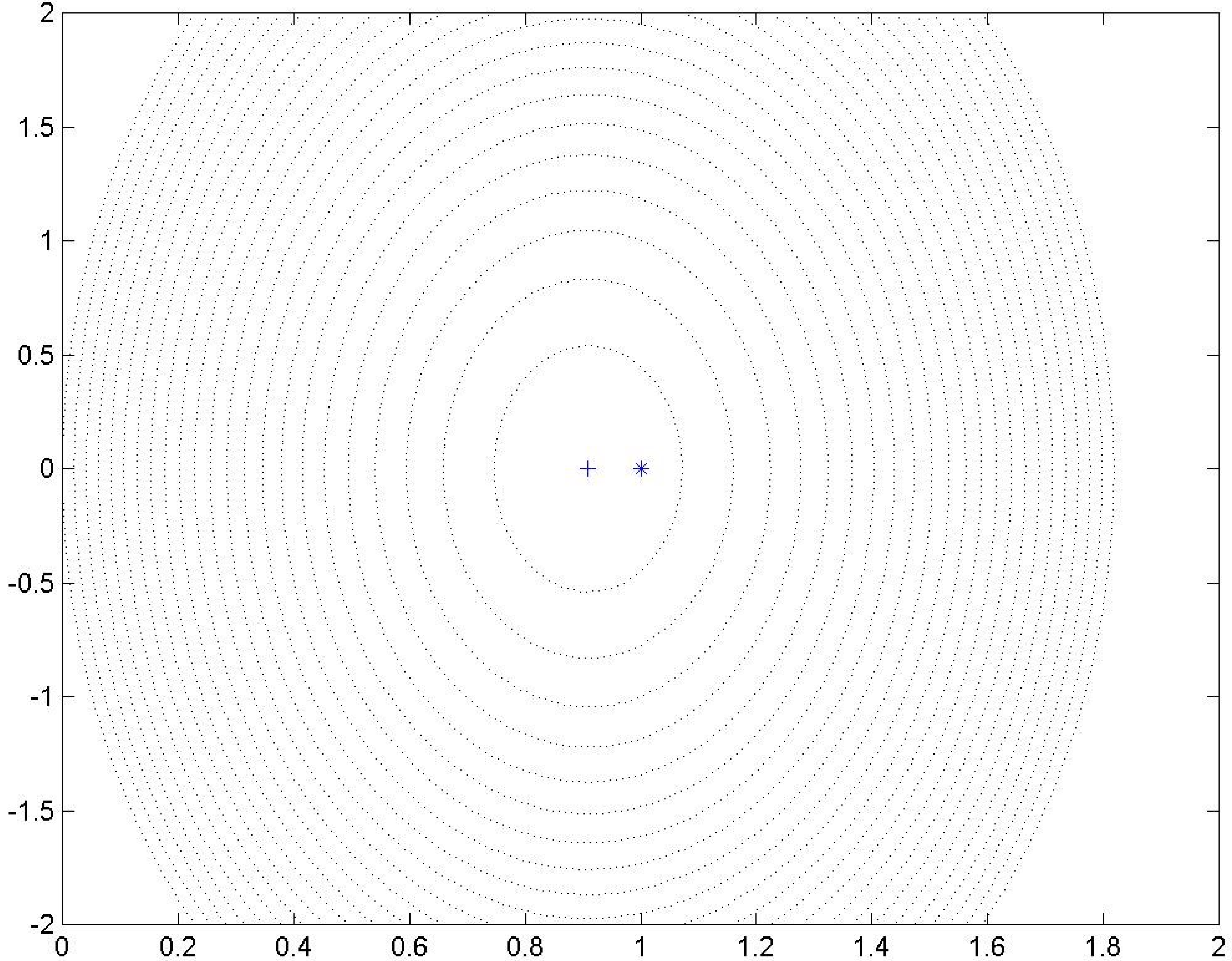
- $\partial L_c(x,\lambda) / \partial x_1 = x_1+\lambda+c(x_1-1)$
- $\partial L_c(x,\lambda) / \partial x_2 = x_2$
- $\partial L_c(x,\lambda) / \partial x_1 = 0 \Leftrightarrow x_1 = (c-\lambda)/(c+1)$
- $\partial L_c(x,\lambda) / \partial x_2 = 0 \Leftrightarrow x_2 = 0$
- Si $\lambda=\lambda^*=-1$, on a $x_1 = x^*_1 = 1$ pour tout $c > 0$
- Si $c\to\infty$, $x_1 \to x^*_1 = 1$ pour tout λ

c=1
λ=0



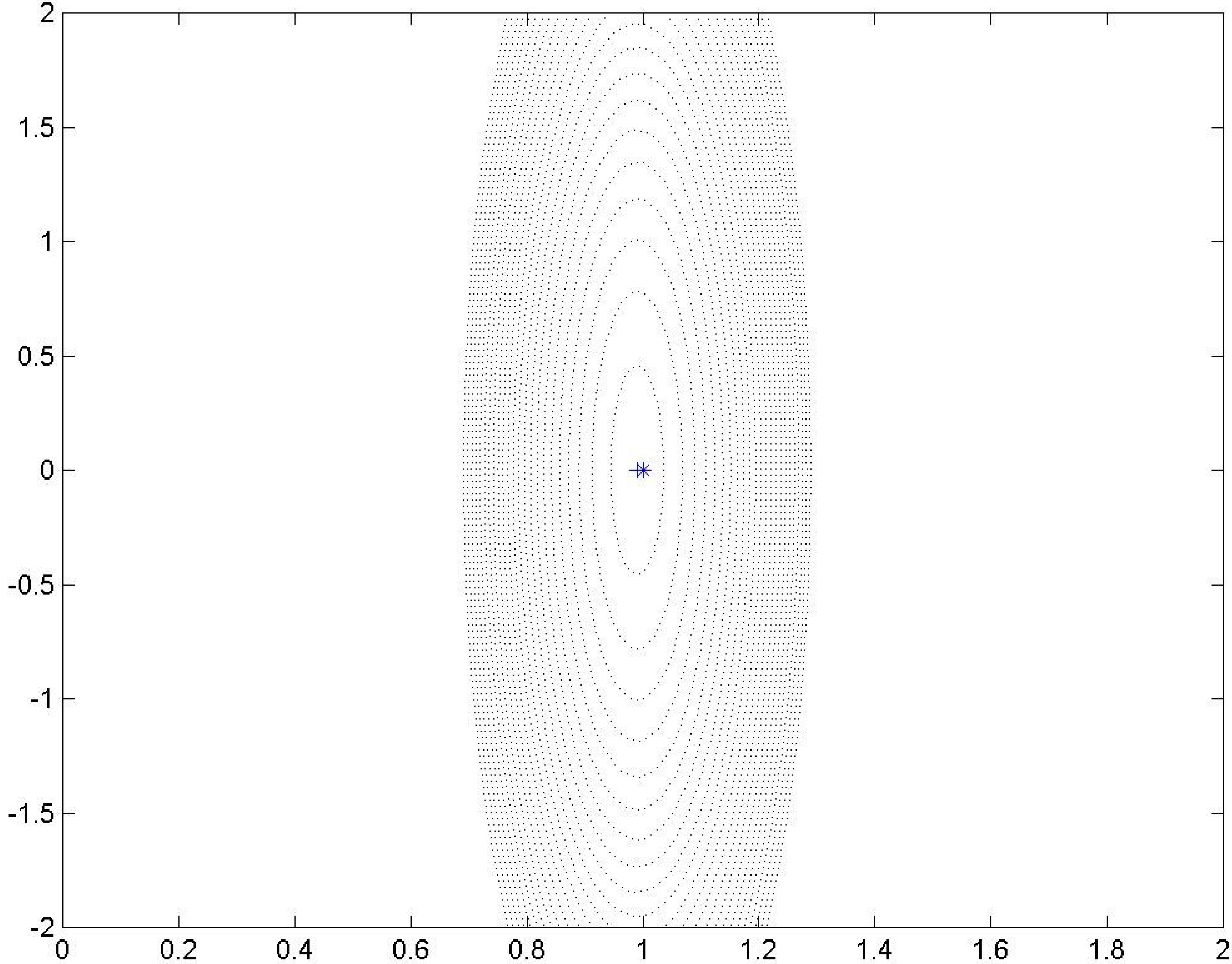
c=10

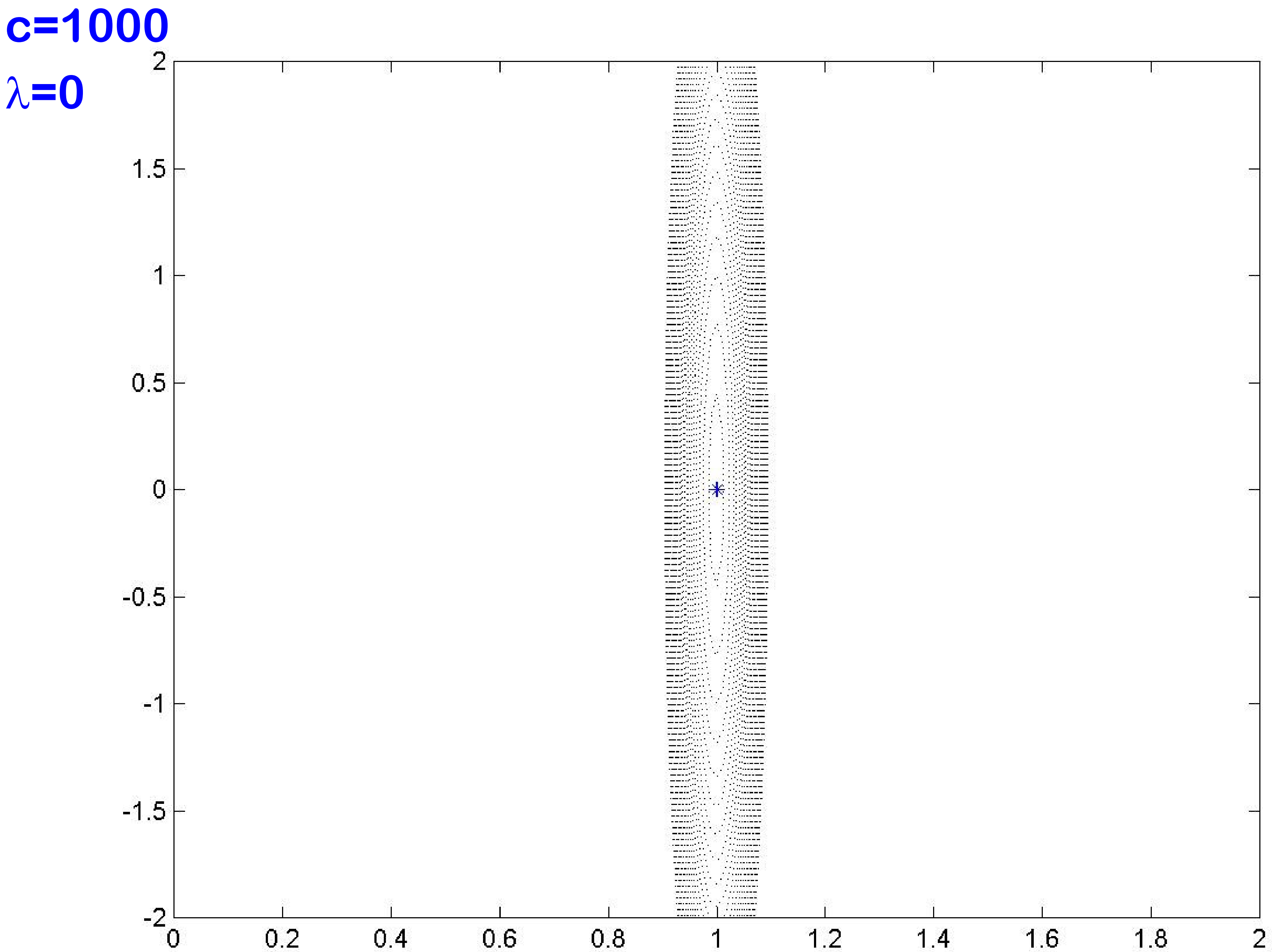
λ=0



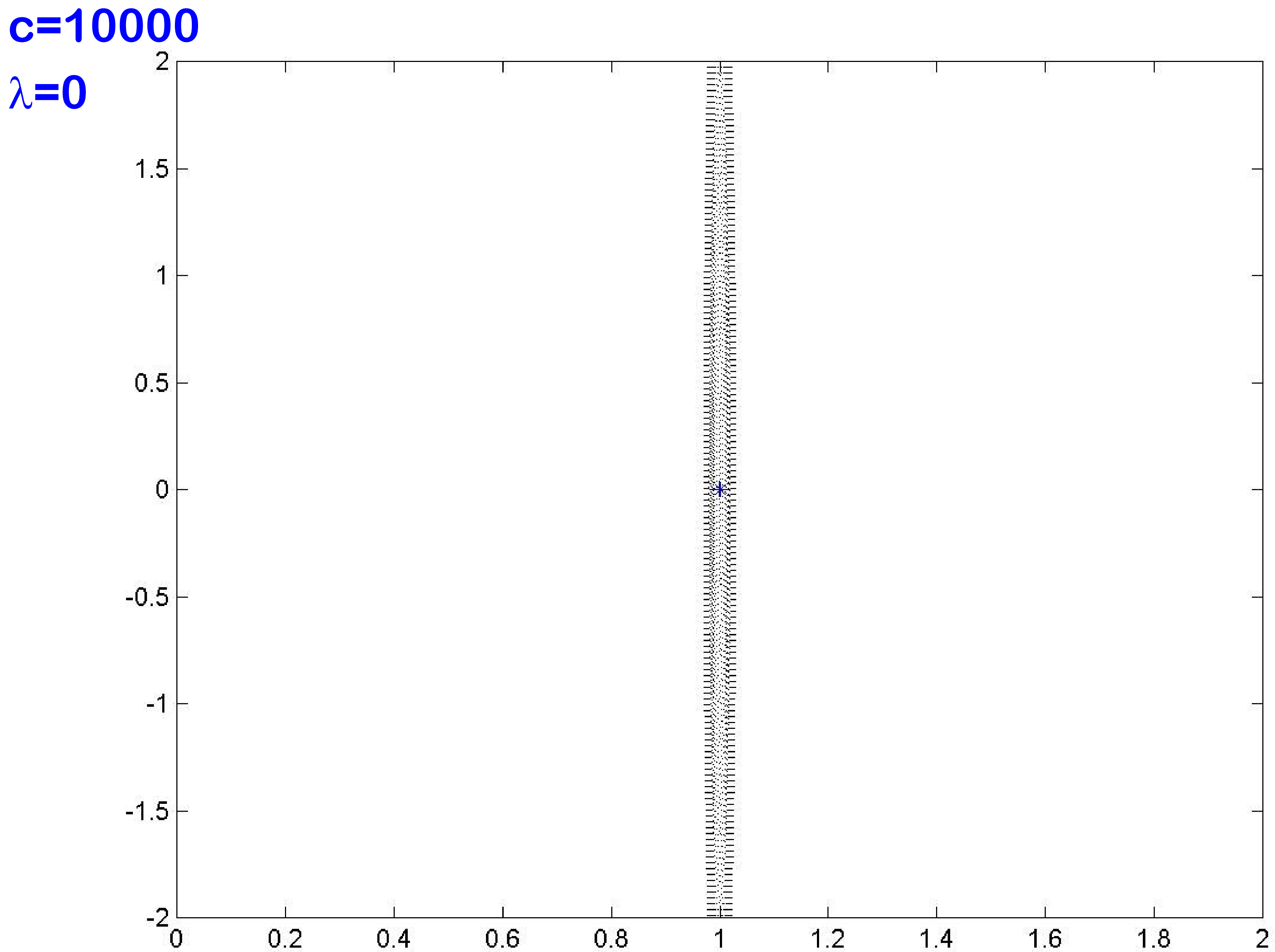
c=100

λ=0





c=1000
λ=0
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2


c=10000
λ=0
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

RECAPITULATIF

Fonction des pénalités :

$$\Phi(x,k) = f(x) + \frac{k}{2} \cdot \sum_{i=1}^{m} [h_i(x)]^2$$

Lagrangien Augmenté :

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i \cdot h_i(x)$$

Lagrangien Pénalisé :

$$L_k(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i \cdot h_i(x) + \frac{k}{2} \cdot \sum_{i=1}^{p} [h_i(x)]^2$$