Chapitre 1: Outils d'optimisation

Abdelkrim El MOUATASIM
Professeur Habilité en Mathématique Appliquée - **AI**
https://sites.google.com/a/uiz.ac.ma/elmouatasim/

FPO - SMI - S6

06 février 2020

Plan

Organisation du cours

Les notations

Espaces, points, fonctions

- V, espaces vectoriels normés.
  Ex. : $V = \mathbb{R}^n$...
- x l'éléments de V.
  Ex. : $x = (x_1, ..., x_n)^t$.
- f : application de V dans $\mathbb{R}$.
  Ex. : $f(x) = \alpha_i x_i^i$.

Définition

Définition (Différentielle au sens de Fréchet)

Une application $f(x)$ de V dans $\mathbb{R}$ est différentiable en $x \in V$ si il existe une application linéaire continue de V dans W, notée $Df(x)$ telle que

$$\forall h \in V \; f(x+h) = f(x) + Df(x).h + \epsilon(h)\|h\|$$

avec $\lim_{h\to 0} \|\epsilon(h)\| = 0$.

Définitions

Définition (Différentielle au sens de Gâteaux)

Une fonction $F(x)$ de V dans $\mathbb{R}$ est une fonction différentiable au sens de Gâteaux en $x \in V$ si il existe une forme linéaire sur V, notée $DF(x)$, telle que

$$\forall h \in V \ \frac{d}{d\lambda} F(x + \lambda h)|_{\lambda=0} = DF(x).h \tag{1}$$

Théorème (Lien avec Fréchet)

Si F est différentiable au sens de Fréchet, F est différentiable au sens de Gâteaux.

Exemples

Exemple 1

Soit $V = \mathbb{R}^n$

$$f(x) = \frac{1}{2}\langle \mathbf{A}x, x\rangle + \sum_j \frac{x_j^4}{4}$$

$$Df(x).h = \frac{d}{dt} f(x+th)|_{t=0} = \langle \mathbf{A}x, h\rangle + \sum_j x_j^3 h_j$$

On pose $x^3 = (x_i^3, \ldots, x_n^3)^t$ donc

$$Df(x).h = \langle \mathbf{A}x, h\rangle + \langle x^3, h\rangle = \langle \mathbf{A}x + x^3, h\rangle$$

Exemples

Exemple 2

Soit $V = \mathbb{R}^n$ et $f(x) = \mathbf{A}x + x^3$, où $\mathbf{A}$ est une matrice et $x^3 = (x_i^3, \ldots, x_n^3)^t$. Soit $\Delta(x)$ la matrice diagonale telle que $\Delta_{i,i} = 3x_i^2$.

$$Df(x).h = \frac{d}{dt} f(x + th)|_{t=0} = \mathbf{A}h + \Delta(x)h$$

Définitions

Définition (Gradient)

Soit V un espace de dimension finie muni d'un produit scalaire $\langle \cdot, \cdot \rangle$. Il existe un vecteur, le gradient, noté $\nabla F(x)$, tel que

$$DF(x).h = \langle \nabla F(x), h \rangle$$

Si $V = \mathbb{R}^n$ et $\langle x, y \rangle = \sum_i x_i y_i$ on retrouve la définition usuelle du gradient

$$\nabla F(x) = (\frac{\partial F}{\partial x_1}, ..., \frac{\partial F}{\partial x_i}, ..., \frac{\partial F}{\partial x_n})^t$$

le gradient définit la direction de plus grande pente $\perp$ à la surface équipotentielle.

Définitions

Définition (Gradient)

Soit V un espace de dimension finie muni d'un produit scalaire $\langle \cdot, \cdot \rangle$. Il existe un vecteur, le gradient, noté $\nabla F(x)$, tel que

$$DF(x).h = \langle \nabla F(x), h \rangle$$

Exemples

$$Df(x).h = \langle \mathbf{A}x + x^3, h \rangle$$

$$\nabla F(x) = \mathbf{A}x + x^3$$

Définitions

Définition (Matrice jacobienne)

Si $V = W = \mathbb{R}^n$ la différentielle $Df(x)$ est une application linéaire représentée donc par une matrice $\mathbf{J}f(x)$, appelée matrice jacobienne, *de coefficients*

$$\mathbf{J}f(x)_{i,j} = \frac{\partial f_i(x)}{\partial x_j}$$

Exemples

Soit $V = \mathbb{R}^n$ et $f(x) = \mathbf{A}x + x^3$, où $\mathbf{A}$ est une matrice.

$$Df(x).h = \mathbf{A}h + \Delta(x)h = (\mathbf{A} + \Delta(x))h$$

$$\mathbf{J}f = \mathbf{A} + \Delta(x)$$

Définitions

Soit $V = \mathbb{R}^n$ et $F(x)$ une fonction différentiable en tout point. Si $f(x) = F'(x) = \nabla F(x)$ est une application différentiable de $\mathbb{R}^n$ dans $\mathbb{R}^n$, la matrice jacobienne $\mathbf{J}f(x)$ est symétrique .

Définition (Hessien)

Le Hessien de la fonction $F(x)$, supposée deux fois différentiable, est la matrice $\mathbf{H}F(x)$ telle que

$$\frac{d^2}{d\lambda^2} F(x+\lambda h)_{\lambda=0} = \langle \mathbf{H}F(x)h, h\rangle$$

Ce qui implique

$$\mathbf{H}F(x)_{i,j} = \frac{\partial^2 F(x)}{\partial x_i x_j}$$

Définitions

Soit $V = \mathbb{R}^n$ et $F(x)$ une fonction différentiable en tout point. Si $f(x) = F'(x) = \nabla F(x)$ est une application différentiable de $\mathbb{R}^n$ dans $\mathbb{R}^n$, la matrice jacobienne $\mathbf{J}f(x)$ est symétrique .

Exemples

Soit $V = \mathbb{R}^n$

$$F(x) = \frac{1}{2}\langle \mathbf{A}x, x\rangle + \sum_j \frac{x_j^4}{4}$$

$$f(x) = \nabla F(x) = \mathbf{A}x + x^3$$

$$\mathbf{J}f = \mathbf{A} + \Delta(x)$$

$$\mathbf{H}F = \mathbf{A} + \Delta(x)$$

Definition

Une matrice réelle D de dimension $n \times n$ est semi-définie positive si $x^T Dx \geqslant 0$ pour tout $x \in \mathbb{R}^n$.

Definition

Une matrice réelle D de dimension $n \times n$ est définie positive si $x^T Dx > 0$ pour tout $x \in \mathbb{R}^n, \quad x \neq 0$.

Caractérisation des minimums

On note $B_r(x)$ la boule de centre x et de rayon r.

Theorem

Si $x \in V$ est un minimum local de $F(x)$ et (si $V = \mathbb{R}^n$) le Hessien $\mathbf{H}F(x)$ est semi-défini positif.
Réciproquement :
Si $DF(x) = 0$ et si il existe une boule $B_r(x)$ telle que le Hessien est semi-défini positif alors x est un minimum local. Si

$$\forall h \in V \;\; F(x+h) = F(x) + \langle \nabla\, F(x), h \rangle + \frac{1}{2}\langle \mathbf{H}F(x)h, h\rangle + \epsilon(h)\|h\|^2$$

où $\lim_{h \to 0} \|\epsilon(h)\| = 0$.

Dérivation numérique

- Méthode "naïve" :

$$\boxed{f'(x) \approx \frac{f(x+h)-f(x)}{h}}$$

- en théorie, la formule est vraie pour h ≈ 0
- en pratique, attention au choix de h !
    - h trop grand : calcul trop approximatif
    - h trop petit : problèmes d'arrondis

Méthode des différences centrales

- Taylor :
  - $f(x+y) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(x) + \dots$
- On connaît f sur un ensemble de points $\{x_i, y_i\}$
  - $h = x_{i+1} - x_i$
  - $f(x+h) \ \rightarrow \ y_{i+1} = y_i + hf'(x_i) + \frac{h^2}{2!}f''(x_i) + \frac{h^3}{3!}f'''(x_i) + \dots$
  - $f(x-h) \ \rightarrow \ y_{i-1} = y_i - hf'(x_i) + \frac{h^2}{2!}f''(x_i) - \frac{h^3}{3!}f'''(x_i) + \dots$

- $f(x+h)$-$f(x-h) \Rightarrow$ y_{i+1} - $y_{i-1} = 2\mathrm{h}f^{'}(x_i) + \frac{2h^3}{3!}f^{''}(x_i) + \ldots$
- en négligeant les termes en h^3 : $\boxed{f^{'}(x_i) \approx \frac{y_{i+1}-y_{i-1}}{2h}}$
- meilleure approximation que la méthode "naïve" (h^3/h^2)

Calcule des dérivées d'ordre supérieur :

- $f''(x_i)$?

$$y_{i+1} = y_i + hf'(x_i) + \frac{h^2}{2!}f''(x_i) + \frac{h^3}{3!}f'''(x_i) + \ldots$$

$$y_{i-1} = y_i - hf'(x_i) + \frac{h^2}{2!}f''(x_i) - \frac{h^3}{3!}f'''(x_i) + \ldots$$

$$y_{i+1} + y_{i-1} = 2y_i + \frac{2h^2}{2!}f''(x_i) + \ldots$$

- *en négligeant les termes en* h^4 :

$$\boxed{f''(x_i) \approx \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2}}$$

- *et pour les autres dérivées ?*

Organisation du cours

Soient a et b deux points dans $\mathbb{R}^n$.
Le segment de droite joignant ces deux points est l'ensemble des points

$$\{x \in \mathbb{R}^n \,|\, \exists\ \lambda \in [0,1] \text{ tel que } x = a + \lambda(b-a) = \lambda b + (1-\lambda)a\}.$$

Une fonction est *convexe* si tout point sur le segment de droite joignant $f(a)$ et $f(b)$ se trouve au-dessus du graphe de f :

$$f(\lambda b + (1-\lambda)a) \leqslant \lambda f(b) + (1-\lambda)f(a), \tag{2}$$

pour toute paire de points (a,b).
Si cette inégalité est stricte, la fonction f est dite *strictement convexe*.

La Figure 1 illustre une fonction strictement convexe.

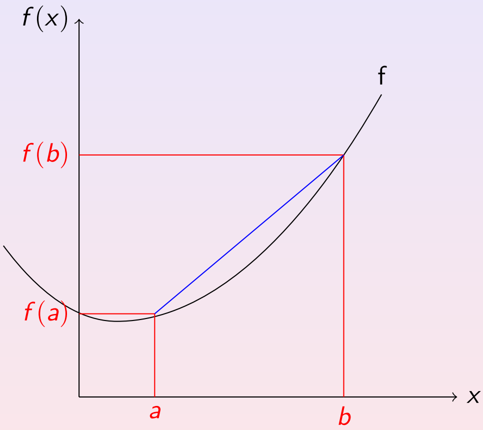


FIGURE : Fonction convexe

En remplaçant le sens de l'inégalité dans (2), la fonction f est dite *concave* :

$$f(\lambda b + (1-\lambda)a) \geqslant \lambda f(b) + (1-\lambda) f(a),$$

pour toute paire (a, b).
Si cette dernière inégalité est stricte, la fonction f est *strictement concave*.

La Figure 2 illustre une fonction strictement concave.

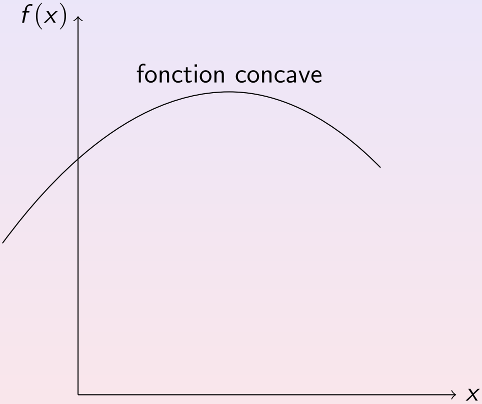


FIGURE : Fonction concave

Proposition

Soit $X \subset \mathbb{R}^n$ un ensemble convexe. Soit $f : X \longrightarrow \mathbb{R}^1$ une fonction convexe. Si $\bar{x}$ est un minimum local de f sur X, alors $\bar{x}$ est un minimum global de f sur X.

Démonstration.

La preuve se fait par contradiction en supposant qu'il existe un point $\tilde{x} \in X$ tel que $f(\tilde{x}) < f(\bar{x})$. Puisque f est convexe,
$f(\theta\tilde{x} + (1-\theta)\bar{x}) \leqslant \theta f(\tilde{x}) + (1-\theta)f(\bar{x}) < \theta f(\bar{x}) + (1-\theta)f(\bar{x}) = f(\bar{x})$
pour tout $\theta \in (0, 1]$. Or pour $\theta > 0$ suffisamment petit,

$$x(\theta) = \theta\tilde{x} + (1-\theta)\bar{x} \in B_\epsilon(\bar{x}) \cap X.$$

Ainsi $f(x(\theta)) < f(\bar{x})$ où $x(\theta) \in B_\epsilon(\bar{x}) \cap X$, contredisant que $\bar{x}$ est un minimum local de f sur X. □

Proposition (Inégalité du gradient)

Soit $X \subset \mathbb{R}^n$ un ensemble convexe. Soit $f : X \longrightarrow \mathbb{R}^1$ une fonction de classe C^1/X. Alors f est convexe sur X si et seulement si

$$f(x) \geqslant f(y) + \nabla f(y)^T (x - y)$$

pour tout paire de points $x, \quad y \in X$.

Corollaire

Soit $X \subset \mathbb{R}^n$ un ensemble convexe. Soit $f : X \longrightarrow \mathbb{R}^1$ une fonction convexe de classe C^1/X. Alors si $\nabla f(x^*)^T (x - x^*) \geqslant 0$ pour tout $x \in X$, alors x^* est un minimum global de f sur X.

Démonstration.

Le résultat découle directement de l'inégalité de gradient. En effet, puisque f est convexe sur X, alors

$$f(x) \geqslant f(x^*) + \nabla f(x^*)^T(x - x^*)$$

pour tout $x \in X$. Puisque par hypothèse $\nabla f(x^*)^T(x - x^*) \geqslant 0$, alors

$$f(x) - f(x^*) \geqslant \nabla f(x^*)^T(x - x^*) \geqslant 0$$

d'où $f(x) \geqslant f(x^*)$ pour tout $x \in X$. Donc x^* est un minimum global de f sur X. □

Conséquence

Si $\nabla f(x^*) = 0$, alors x^* est un minimum global de f sur X.

Supposons que f est une fonction deux fois dérivable d'une seule variable. Alors,

1. f est convexe si et seulement si
$$\forall\ x,\ \frac{d^2 f(x)}{dx^2} \geqslant 0;$$
2. f est strictement convexe si
$$\forall\ x,\ \frac{d^2 f(x)}{dx^2} > 0;$$
3. f est concave si et seulement si
$$\forall\ x,\ \frac{d^2 f(x)}{dx^2} \leqslant 0;$$
4. f est strictement concave si
$$\forall\ x,\ \frac{d^2 f(x)}{dx^2} < 0.$$

Supposons que f est une fonction deux dérivable de deux variables. Dans ce cas, nous utilisons la Table 1 comme conditions nécessaires.

	Convexe	Strict. convexe	Concave	Strict. concave
$\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2}$	$\geqslant 0$	> 0	$\leqslant 0$	< 0
$\frac{\partial^2 f(x_1,x_2)}{\partial x_2^2}$	$\geqslant 0$	> 0	$\leqslant 0$	< 0
$\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2} \cdot \frac{\partial^2 f(x_2,x_2)}{\partial x_2^2} - \left(\frac{\partial^2 f(x_1,x_2)}{\partial x_1 x_2}\right)^2$	$\geqslant 0$	> 0	$\geqslant 0$	> 0

TABLE : Test de convexité et de concavité pour une fonction à deux variables deux fois dérivable

Ces tests se généralisent à des fonctions de plus de deux variables en utilisant le Hessien (la matrice des dérivées secondes). Deux résultats supplémentaires sont utiles pour tester la convexité :

- une fonction qui s'exprime comme la somme de fonctions convexes (concaves) est convexe (concave) ;
- l'opposé d'une fonction concave est convexe et vice-versa.

Pour les fonctions f de classe C^2, il existe un critère pour vérifier la convexité qui est basé sur le Hessien $\nabla^2 f$.

Proposition

Soit $X \subset \mathbb{R}^n$ un ensemble convexe dont l'intérieur est non vide. Soit $f : X \longrightarrow \mathbb{R}^1$ une fonction de classe C^2/X. Alors f est convexe sur X si et seulement si son Hessien $\nabla^2 f$ est une matrice semi définie positive pour tout $x \in X$.

Théorème

Une matrice symétrique définie positive possède des valeurs propres strictement positives.

Démonstration.

1. Supposons A symétrique définie positive. Soit λ_i une valeur propre associée au vecteur propre x_i. On a donc $Ax_i = \lambda_i x_i$, et $x_i^T Ax_i = \lambda_i x_i^T x_i = \lambda_i \|x_i\|^2 > 0$.
2. Réciproquement, si toutes les valeurs propres sont positives, on a pour tout vecteur propre x_i $x_i^T Ax_i > 0$. Ces derniers formant une base orthonormée, on écrit pour tout $x \neq 0$

$$x^T Ax = \left(\sum_{i=1}^{n} \alpha_i x_i\right)^T A \left(\sum_{i=1}^{n} \alpha_i x_i\right) = \sum_{i=1}^{n} \alpha_i^2 \lambda_i > 0.$$

$\square$

Organisation du cours

Un ensemble est convexe si, pour toute paire de points de l'ensemble, le segment de droite joignant ces points est contenu dans l'ensemble. Il est possible de démontrer qu'un ensemble est convexe grâce aux propriétés suivantes :

- si f est convexe, alors $\{x \mid f(x) \leqslant b\}$ est convexe ;
- si f est concave, alors $\{x \mid f(x) \geqslant b\}$ est convexe ;
- l'intersection d'ensembles convexes est convexe.

Example (Wyndor Glass)

Soit le problème d'optimisation suivant :

$$\max_{x} z = 3x_1 + 5x_2$$

sous les contraintes

$$\begin{aligned} x_1 \qquad &\leqslant 4 \quad \text{(usine 1)} \\ 2x_2 &\leqslant 12 \quad \text{(usine 3)} \\ 3x_1 + 2x_2 &\leqslant 18 \quad \text{(usine 3)} \end{aligned}$$

$$x_1 \geqslant 0, \; x_2 \geqslant 0 \text{ (non-négativité)}$$

Example (Wyndor Glass)

Modifions l'exemple Wyndor Glass en remplaçant certaines contraintes par une contrainte non linéaire.

$$\begin{aligned} \max z &= 3x_1 + 5x_2 \\ \text{s.c. } x_1 &\leqslant 4 \\ 9x_1^2 + 5x_2^2 &\leqslant 216 \\ x_1 \geqslant 0,\ x_2 &\geqslant 0. \end{aligned}$$

Le problème ainsi modifié peut se représenter comme sur la Figure 3.

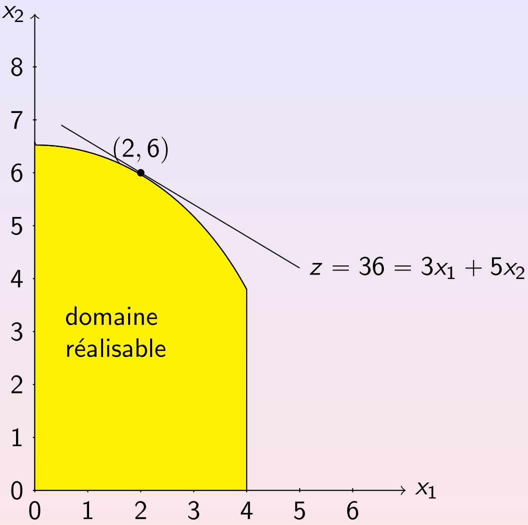
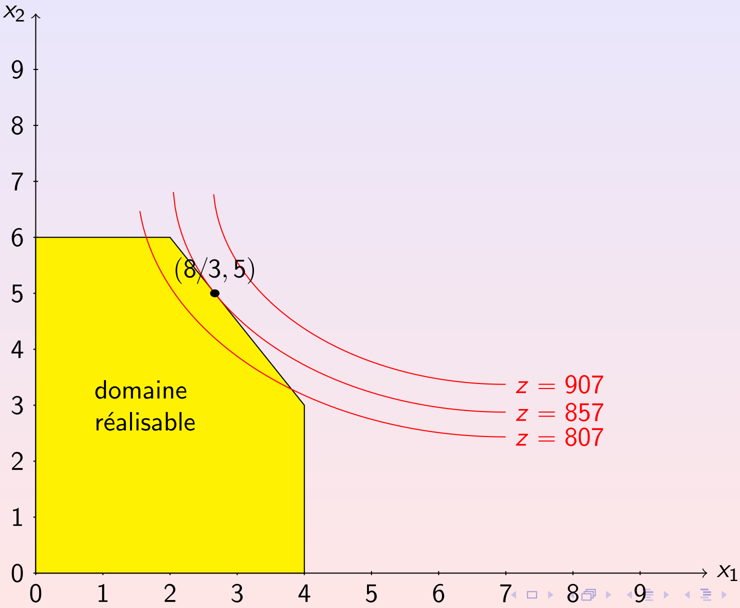


FIGURE : exemple Wyndor Glass

Remarquons que l'objectif est concave (car linéaire) et que le domaine réalisable est convexe : c'est un modèle de *programmation convexe*.
Dans le cas présent, la solution optimale est sur la frontière du domaine réalisable, mais ne correspond pas à un coin (l'intersection de deux contraintes).
Modifions encore l'exemple Wyndor Glass, mais cette fois en remplaçant l'objectif lineaire par une fonction non lineaire, comme illustré sur la Figure 4 :

$$
\begin{aligned}
\max z \; &= 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2 \\
& x_1 \leqslant 4 \\
& 2x_2 \leqslant 12 \\
& 3x_1 + 2x_2 \leqslant 18
\end{aligned}
$$


x_2
9
8
7
6
5
4
3
2
1
0
(8/3, 5)
domaine
réalisable
$z = 907$
$z = 857$
$z = 807$
0
1
2
3
4
5
6
7
8
9
x_1

Nous pouvons à nouveau remarquer que l'objectif est concave et que le domaine réalisable est convexe : c'est aussi un modèle de programmation convexe. Ici, la solution optimale $(8/3, 5)$ est sur la frontière du domaine réalisable, mais ne correspond pas à un point extrême du domaine réalisable.

Considérons le même exemple, mais avec une fonction objectif différente, tel que représenté sur la Figure 5 :

$$\begin{aligned}
\max z \ &= 54x_1 - 9x_1^2 + 78x_2 - 13x_2^2 \\
&x_1 \leqslant 4 \\
&2x_2 \leqslant 12 \\
&3x_1 + 2x_2 \leqslant 18
\end{aligned}$$

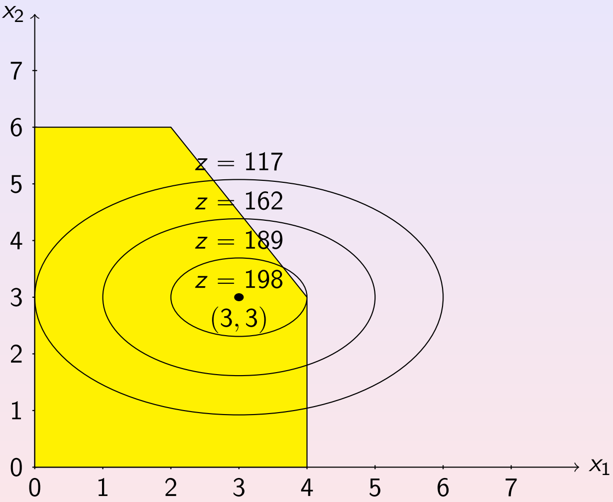
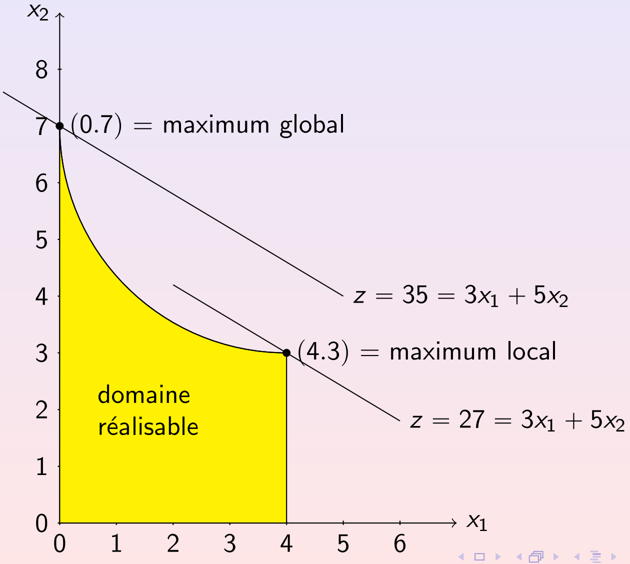


FIGURE : exemple Wyndor Glass - cas 3

Il s'agit à nouveau d'un modèle de programmation convexe, avec comme solution optimale $(3, 3)$, laquelle se trouve à l'intérieur du domaine réalisable. La fonction objectif est la somme de deux fonctions d'une seule variable. Si nous annulons les dérivées de chacune de ces fonctions, nous obtenons la solution unique $(3, 3)$, qui se trouve à l'intérieur du domaine ; c'est donc nécessairement la solution optimale.
Revenons sur la fonction objectif initiale, mais introduisons une contrainte non-linéaire, qui définit un domaine réalisable non-convexe, comme illustré sur la Figure 6 :

$$\begin{aligned} \max z &= 3x_1 + 5x_2 \\ \text{s.c. } & x_1 \leqslant 4 \\ & x_2 \leqslant 7 \\ & 8x_1 - x_1^2 + 14x_2 - x_2^2 \leqslant 49 \\ & x_1 \geqslant 0,\ x_2 \geqslant 0. \end{aligned}$$

Le problème ainsi modifié peut se représenter comme sur la Figure 6.



Dans ce modèle de *programmation non convexe*, nous remarquons la présence de deux maxima locaux :

- x est un *maximum local* si $f(x) \geqslant f(a)$ pour tout a réalisable suffisamment près de a ;
- $(4,3)$ et $(0,7)$ sont des maxima locaux, mais $x^* = (0,7)$ est la *maximum global* : $f(x) \geqslant f(a)$ pour tout a réalisable.

Les méthodes classiques de programmation non-linéaire permettent d'identifier un maximum local, mais pas nécessairement le maximum global, alors qu'en programmation convexe, tout maximum local est global.