

# Algorithmique numérique (LU3IN028)

## Cours n°2 : Introduction à l'optimisation (1/2)

Stef Graillat

Sorbonne Université



- Présentation de l'arithmétique à virgule flottante
- Introduction au logiciel MATLAB
- Introduction à la Symbolic Math Toolbox de MATLAB
- Stockage et manipulation efficace des matrices (BLAS)

- Existence et unicité des extrema
- Conditions nécessaires et suffisantes vérifiées par les extrema
- Algorithmes pour le calcul d'extrema en dimension 1

L'optimisation intervient dans tous les domaines. On peut citer par exemple

- la finance
- l'économie
- le contrôle optimal
- la météorologie
- le traitement d'image
- la gestion de la production d'électricité
- la biologie moléculaire
- etc.

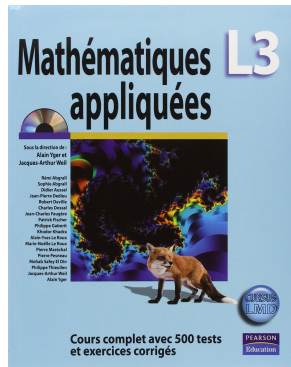
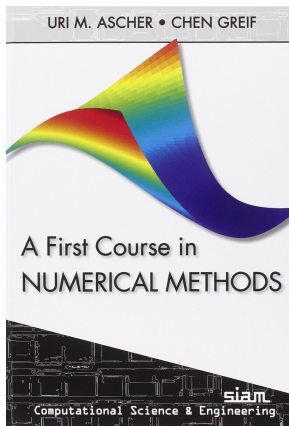
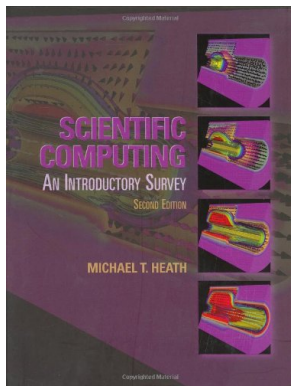
# Références principales

- **Scientific Computing, An Introductory Survey, Michael .T. Heath, McGraw-Hill, 2002** (ce cours est basé sur le livre et les transparents associés)
- A First Course in Numerical Methods, Uri M. Ascher, Chen Greif, SIAM, 2011
- Mathématiques appliquées L3, sous la direction de Jacques-Arthur Weil et Alain Yger, Pearson, 2009
- Scientific Computing with Case Studies, Dianne P. O'Leary, SIAM, 2009
- Numerical Optimization, Jorgue Nocedal et Stephen Wright, Springer, 2006
- Introduction à l'Analyse Numérique Matricielle et à l'Optimisation, Philippe G. Ciarlet, Dunod, 1982
- Analyse numérique et optimisation, Grégoire Allaire, Éditions de l'École polytechnique, 2005
- L'optimisation, J.B. Hiriart-Urruty, PUF, 1996
- Convex Optimization, S. Boyd et L. Vandenberghe, Cambridge University Press, 2004

## Références principales (suite)

- Numerical optimization : Theoretical and practical aspects, F. Bonnans, J.C. Gilbert, C. Lemaréchal et C. Sagastizábal, Springer-Verlag, 2006
- Optimisation continue, F. Bonnans, Dunod, 2006
- Optimisation et contrôle des systèmes linéaires, M. Bergounioux, Dunod, 2001.
- Convex Analysis and Nonlinear Optimization, J.M. Borwein et A.S. Lewis, Springer, 2000
- Lectures on Modern Convex Optimization - Analysis, Algorithms and Engineering Applications, A. Ben-Tal et A. Nemirovski, SIAM, 2001
- Numerical Computing with MATLAB, Cleve Moler, SIAM, 2004
- Numerical Recipes. The Art of Scientific Computing, William Press, Saul Teukolsky, William Vetterling et Brian Flannery, 3rd Edition, Cambridge University Press, 2007

# Références principales (suite)



- Généralités sur les problèmes d'optimisation
- Optimisation en dimension 1
- Optimisation en dimension  $n \geq 2$

# Généralités sur les problèmes d'optimisation

- Étant donné une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  et un ensemble  $S \subset \mathbb{R}^n$ , trouver  $x^* \in S$  tel que  $f(x^*) \leq f(x)$  pour tout  $x \in S$
- $x^*$  est appelé un **minimum** ou **minimiseur** de  $f$  sur  $S$
- Il suffit de ne considérer que la minimisation car maximiser  $f$  revient à minimiser  $-f$
- La **fonction objectif**  $f$  est en général différentiable et peut être linéaire ou non linéaire
- L'ensemble des **contraintes**  $S$  est défini par un système d'équations et d'inéquations qui peuvent être linéaires ou non linéaires
- Un point  $x \in S$  est dit **admissible**, **acceptable** ou **réalisable** (*feasible* en anglais)
- Si  $S = \mathbb{R}^n$ , alors le problème est dit **sans contrainte**

- Un problème d'optimisation s'écrit généralement sous la forme

$$\min f(x) \quad \text{sous} \quad g(x) = 0 \text{ et } h(x) \leq 0.$$

avec  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  et  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$

- **Programmation linéaire** :  $f$ ,  $g$  et  $h$  sont des fonctions linéaires
- **Programmation non linéaire** : au moins une des fonctions  $f$ ,  $g$  et  $h$  est non linéaire

- Minimiser le poids d'une structure en prenant en compte des contraintes sur les forces
- Minimiser la surface d'un cylindre avec une contrainte sur le volume :

$$\min_{x_1, x_2} f(x_1, x_2) = 2\pi x_1(x_1 + x_2)$$

$$\text{sous } g(x_1, x_2) = \pi x_1^2 x_2 - V = 0$$

où  $x_1$  et  $x_2$  sont respectivement le rayon et la hauteur du cylindre et  $V$  le volume requis

# Optimisation locale vs globale

- $x^* \in S$  est un **minimum global** si  $f(x^*) \leq f(x)$  pour tout  $x \in S$ .
- $x^* \in S$  est un **minimum local** si  $f(x^*) \leq f(x)$  pour  $x$  dans un voisinage de  $x^*$



- Trouver (ou même vérifier) un minimum global est en général très difficile
- La plupart des méthodes d'optimisation sont conçues pour trouver un minimum local (qui peut être ou non global)
- Si un minimum global est recherché, on peut essayer d'appliquer une méthode d'optimisation avec des points initiaux différents
- Pour certains problèmes tels que la programmation linéaire, la recherche d'un minimum global est atteignable

# Existence d'un minimum

- Si  $f$  est continue sur un fermé borné  $S \subset \mathbb{R}^n$  alors  $f$  admet un minimum global sur  $S$
- Si  $S$  n'est pas fermé ou non borné, alors  $f$  peut n'avoir ni minimum local, ni minimum global sur  $S$
- Une fonction continue  $f$  sur un ensemble non borné  $S$  est dite **coercive** si

$$\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty,$$

*i.e.*  $f(x)$  devient grand quand  $\|x\|$  devient grand

- Si  $f$  est coercive sur un ensemble fermé non borné  $S \subset \mathbb{R}^n$  alors  $f$  admet un minimum global sur  $S$

- Une **ligne de niveau** pour une fonction  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  est l'ensemble des points de  $S$  pour lesquels  $f$  a une valeur donnée constante
- Étant donné  $\gamma \in \mathbb{R}$ , un **ensemble de sous-niveau** est défini par

$$L_\gamma = \{x \in S : f(x) \leq \gamma\}$$

- Si une fonction continue  $f$  sur  $S \subset \mathbb{R}^n$  a un ensemble de sous-niveau fermé et borné alors  $f$  a un minimum global sur  $S$
- Si  $f$  est non bornée alors  $f$  est coercive sur  $S$  si et seulement si tous ses ensembles de sous-niveau sont bornés

# Unicité d'un minimum

- L'ensemble  $S$  est **convexe** s'il contient tous les segments compris entre deux de ses points

$$\forall x, y \in S, \forall \alpha \in [0;1], \quad \alpha x + (1 - \alpha)y \in S$$

- Une fonction  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  est **convexe** sur l'ensemble convexe  $S$  si

$$\forall x, y \in S, \forall \alpha \in [0;1], \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

- Une fonction  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  est **strictement convexe** sur l'ensemble convexe  $S$  si

$$\forall x, y \in S, \forall \alpha \in ]0;1[, \quad f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$$

- Tout minimum local d'une fonction convexe  $f$  sur un ensemble convexe  $S \subset \mathbb{R}^n$  est un minimum global de  $f$  sur  $S$
- Tout minimum local d'une fonction strictement convexe  $f$  sur un ensemble convexe  $S \subset \mathbb{R}^n$  est le minimum global (unique) de  $f$  sur  $S$

# Condition d'optimalité du premier ordre

- Pour des fonctions d'une variable, on trouve les extrema en calculant les zéros de la dérivée
- Pour des fonctions à  $n$  variables, on cherche les points critiques, *i.e.* les solutions du système

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

où  $\nabla f(x)$  est le gradient de  $f$  dont la  $i$ -ième composante est  $\partial f(x)/\partial x_i$

- Si  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  est de classe  $C^1$  alors un minimum local  $x^*$  appartenant à l'intérieur de  $S$  est un point critique de  $f$
- Attention : tous les points critiques ne sont pas forcément des minimas (ils peuvent être des maxima ou des points selle)

# Condition d'optimalité du second ordre

- Pour une fonction  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  de classe  $C^2$ , on distingue les points critiques en considérant la matrice Hessienne  $H_f(x)$  définie par

$$(H_f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

qui est symétrique. On note aussi  $H_f(x) = \nabla^2 f(x)$

- À un point critique  $x^*$ , si  $H_f(x^*)$  est
  - définie-positive alors  $x^*$  est un minimum de  $f$
  - définie-négative alors  $x^*$  est un maximum de  $f$
  - indéfinie alors  $x^*$  est un point selle
  - singulière alors on ne peut rien dire

- Si le problème est contraint, alors seuls les points réalisables sont intéressants
- Pour le problème

$$\min f(x) \text{ sous } g(x) = 0$$

où  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  et  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , une condition nécessaire pour qu'un point réalisable  $x^*$  soit solution est que

$$-\nabla f(x^*) = J_g(x^*)^T \lambda$$

où  $J_g$  est la matrice jacobienne de  $g$  et  $\lambda$  est le vecteur des **multiplicateurs de Lagrange**

## Condition sous contraintes (suite)

- Le Lagrangien  $\mathcal{L} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$  est défini par

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T g(x)$$

- Son gradient est donné par

$$\nabla \mathcal{L}(x, \lambda) = \begin{pmatrix} \nabla f(x) + J_g^T(x)\lambda \\ g(x) \end{pmatrix}$$

- La Hessienne est donnée par

$$H_{\mathcal{L}}(x, \lambda) = \begin{pmatrix} B(x, \lambda) & J_g^T(x) \\ J_g(x) & 0 \end{pmatrix}$$

où

$$B(x, \lambda) = H_f(x) + \sum_{i=1}^m \lambda_i H_{g_i}(x)$$

- Une condition nécessaire est donc un point critique du Lagrangien

$$\nabla \mathcal{L}(x, \lambda) = \begin{pmatrix} \nabla f(x) + J_g^T(x)\lambda \\ g(x) \end{pmatrix} = 0$$

- La Hessienne du Lagrangien est symétrique mais pas en général définie. Par conséquent un point critique de  $\mathcal{L}$  est un point selle plus qu'un minimum ou un maximum
- Un point critique  $(x^*, \lambda^*)$  de  $\mathcal{L}$  est un minimum contraint de  $f$  si  $B(x^*, \lambda^*)$  est défini positif sur le noyau de  $J_g(x^*)$
- Si les colonnes de  $Z$  forment une base du noyau de  $J_g(x^*)$  alors il suffit de tester le caractère défini positif du Hessien projeté  $Z^T B Z$

- Développement de Taylor de  $f$  au voisinage de  $x^*$

$$f(\widehat{x}) = f(x^* + h) = f(x^*) + f'(x^*)h + \frac{1}{2}f''(x^*)h^2 + \mathcal{O}(h^3)$$

Puisque  $f'(x^*) = 0$ , si  $|f(\widehat{x}) - f(x^*)| \leq \varepsilon$  alors  $|\widehat{x} - x^*|$  est de l'ordre de  $\sqrt{2\varepsilon/|f''(x^*)|}$

- En conséquence, basée seulement sur l'évaluation de fonction, un minimum ne peut être calculé qu'avec une précision divisée par deux

# Optimisation en dimension 1

- Pour minimiser une fonction d'une seule variable, on a besoin d'encadrer la solution par une méthode analogue à la dichotomie pour le calcul des racines d'une fonction
- Une fonction à valeur réelle  $f$  est **unimodale** sur l'intervalle  $[a, b]$  s'il existe un unique point  $x^* \in [a, b]$  tel que  $f(x^*)$  soit minimal sur  $[a, b]$ ,  $f$  soit strictement décroissant pour  $x \leq x^*$  et strictement croissante pour  $x^* \leq x$
- L'unimodalité permet de restreindre la recherche à un sous intervalle (analogue à la dichotomie)

# Méthode de la section dorée (golden section search)

Appelée aussi **méthode du nombre d'or**

- On suppose  $f$  unimodale sur  $[a, b]$  et soient  $x_1, x_2$  deux points de  $[a, b]$  avec  $x_1 < x_2$
- En évaluant et comparant  $f(x_1)$  et  $f(x_2)$ , on peut retirer un des intervalles  $]x_2, b]$  ou  $[a, x_1[$ , le minimum appartenant à l'un des sous-intervalles restants
- On peut itérer le procédé en ne calculant à chaque fois qu'une seule évaluation de la fonction
- On veut réduire l'intervalle de recherche d'un même facteur à chaque itération et de plus, on veut conserver les mêmes relations entre les points du nouvel intervalle qu'avec celui de l'ancien

- Pour se faire, on choisit les positions relatives des deux points par  $\tau$  et  $1 - \tau$  avec  $\tau^2 = 1 - \tau$ . Donc  $\tau = (\sqrt{5} - 1)/2 \approx 0.618$  et  $1 - \tau \approx 0.382$
- Quel que soit le sous-intervalle retenu, sa longueur sera de  $\tau$  fois celui de l'intervalle précédent et les nouveaux points seront en position  $\tau$  et  $1 - \tau$  relativement au nouvel intervalle
- À chaque itération, on n'effectue qu'une seule évaluation de la fonction
- La taux de convergence est linéaire

# Méthode de la section dorée (suite)

$$\tau = (\sqrt{5} - 1)/2$$

$$x_1 = a + (1 - \tau)(b - a); \quad f_1 = f(x_1)$$

$$x_2 = a + \tau(b - a); \quad f_2 = f(x_2)$$

**tant que**  $((b - a) > \text{tol})$  **faire**

**si**  $(f_1 > f_2)$  **alors**

$$a = x_1$$

$$x_1 = x_2$$

$$f_1 = f_2$$

$$x_2 = a + \tau(b - a)$$

$$f_2 = f(x_2)$$

**sinon**

$$b = x_2$$

$$x_2 = x_1$$

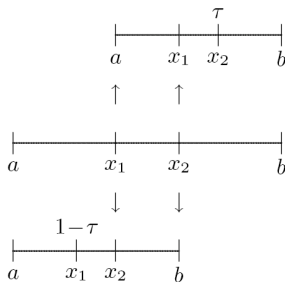
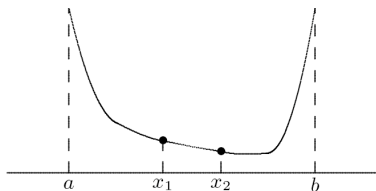
$$f_2 = f_1$$

$$x_1 = a + (1 - \tau)(b - a)$$

$$f_1 = f(x_1)$$

**fin**

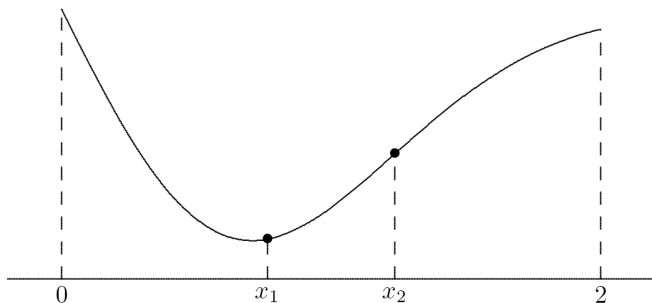
**fin**



# Méthode de la section dorée : exemple

Utilisation de la méthode de la section dorée pour minimiser

$$f(x) = 0.5 - xe^{-x^2}$$

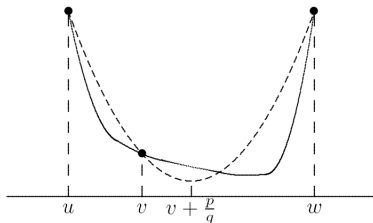


## Méthode de la section dorée : exemple (suite)

$x_1$	$f_1$	$x_2$	$f_2$
0.764	0.074	1.236	0.232
0.472	0.122	0.764	0.074
0.764	0.074	0.944	0.113
0.656	0.074	0.764	0.074
0.584	0.085	0.652	0.074
0.652	0.075	0.695	0.071
0.695	0.071	0.721	0.071
0.679	0.072	0.695	0.071
0.695	0.071	0.705	0.071
0.705	0.071	0.711	0.071

# Méthode d'interpolation quadratique

- On approxime la fonction par une fonction quadratique (polynôme de degré 2) en 3 valeurs
- On minimise la fonction quadratique pour obtenir une approximation du minimum

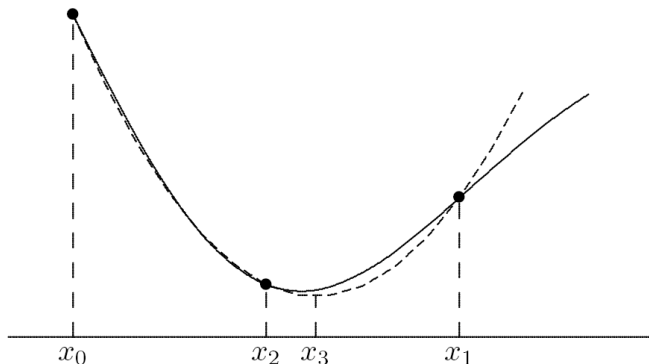


- on remplace un des trois points par le nouveau point et on itère le procédé jusqu'à convergence
- Le taux de convergence est superlinéaire

# Méthode d'interpolation quadratique

Utilisation de la méthode d'interpolation quadratique pour minimiser

$$f(x) = 0.5 - xe^{-x^2}$$



# Méthode d'interpolation quadratique

$x_k$	$f(x_k)$
0.000	0.500
0.600	0.081
1.200	0.216
0.754	0.073
0.721	0.071
0.692	0.071
0.707	0.071

- Une autre approximation quadratique est la formule de Taylor à l'ordre 2

$$f(x+h) \approx f(x) + f'(x)h + \frac{1}{2}f''(x)h^2$$

- Le minimum de cette fonction quadratique de  $h$  est donné par  
 $h = -f'(x)/f''(x)$

- L'itération

$$x_{k+1} = x_k - f'(x_k)/f''(x_k)$$

est l'itération de Newton pour la résolution du système non linéaire  
 $f'(x) = 0$

- La méthode de Newton pour trouver un minimum a donc une convergence quadratique à condition de commencer l'itération suffisamment près de la solution

# Méthode de Newton : exemple

- Utiliser la méthode de Newton pour minimiser  $f(x) = 0.5 - xe^{-x^2}$
- Les dérivées premières et secondes sont données par

$$f'(x) = (2x^2 - 1)e^{-x^2} \quad \text{et} \quad f''(x) = 2x(3 - 2x^2)e^{-x^2}$$

- L'itération de Newton est donc donnée par

$$x_{k+1} = x_k - (2x_k^2 - 1)/(2x_k(3 - 2x_k^2))$$

- En commençant à  $x_0 = 1$ , on obtient

$x_k$	$f(x_k)$
1.000	0.132
0.500	0.111
0.700	0.071
0.707	0.071

On a présenté

- des généralités sur les problèmes d'optimisation (existence, unicité, CNS, etc)
- des algorithmes pour résoudre des problèmes d'optimisation en dimension 1

La semaine prochaine, on présentera

- des algorithmes pour résoudre des problèmes d'optimisation en dimension  $n \geq 2$