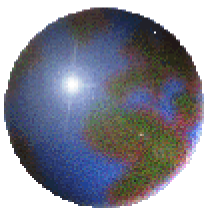
Optimisation

Chapitre 2

Algorithmes d’optimisation numérique



Master IASD

Tarik AMTOUT

tarik.amtout@gmail.com

Formulation d'un problème d'optimisation

La forme standard d'un problème d'optimisation continue est la suivante.

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{sous} \quad \begin{cases} c_E(x) = 0 \\ c_I(x) \le 0 \end{cases} \qquad (1.1)$$

Le vecteur $x \in \mathbb{R}^n$ représente les n variables ou paramètres du problème.

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

La fonction à minimiser $f : \mathbb{R}^n \to \mathbb{R}$ est appelée coût ou objectif ou encore critère.

La fonction $c_E : \mathbb{R}^n \to \mathbb{R}^p$ représente un vecteur de p contraintes d'égalité.

$$c_E(x) = \begin{pmatrix} c_{E1}(x) \\ \vdots \\ c_{Ep}(x) \end{pmatrix} \quad \text{avec} \quad c_{Ej} : \mathbb{R}^n \to \mathbb{R}$$

La fonction $c_I : \mathbb{R}^n \to \mathbb{R}^q$ représente un vecteur de q contraintes d'inégalité.

$$c_I(x) = \begin{pmatrix} c_{I1}(x) \\ \vdots \\ c_{Iq}(x) \end{pmatrix} \quad \text{avec} \quad c_{Ij} : \mathbb{R}^n \to \mathbb{R}$$

Les contraintes c_E et c_I sont regroupées dans le vecteur c de dimension $m = p + q$.

$$c(x) = \begin{pmatrix} c_E(x) \\ c_I(x) \end{pmatrix} \in \mathbb{R}^m$$

Tout problème d'optimisation peut se mettre sous la forme standard (1.1) en utilisant les transformations suivantes.

Maximisation/minimisation : $\max_x f(x) \Leftrightarrow \min_x -f(x)$

Contrainte supérieur / inférieur : $c(x) \geq 0 \Leftrightarrow -c(x) \leq 0$

Hypothèses :

- Continuité : Fonctions continues de variables réelles
  $\rightarrow$ Optimisation continue
  $\neq$ Optimisation combinatoire, Programmation en nombres entiers

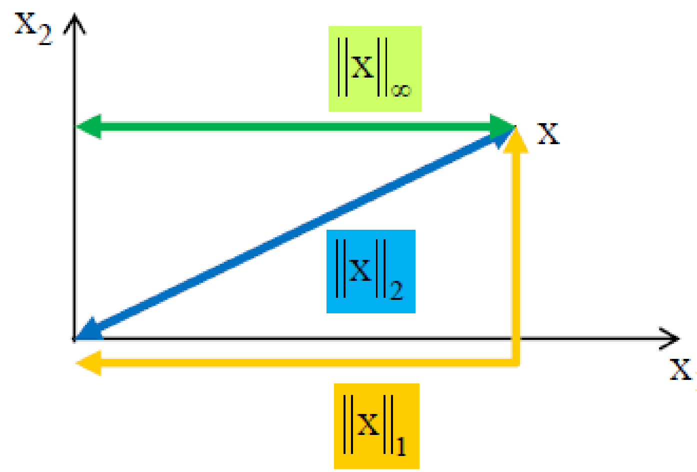
- Différentiabilité : Fonctions différentiables
  $\rightarrow$ **Méthodes à base de gradient**
  $\neq$ Méthodes sans dérivées

- Déterminisme : Les données du problème sont parfaitement connues
  $\neq$ Optimisation stochastique

Norme

Norme vectorielle sur R^n

- Fonction $\|\,.\,\| : R^n \to R$ vérifiant $\begin{cases} \|x\| \geq 0 \\ \|x\| = 0 \Leftrightarrow x = 0 \\ \|x+y\| \leq \|x\| + \|y\| \\ \|\alpha x\| = |\alpha|\,\|x\| \end{cases}$
- Norme p : $\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$
- **Norme ∞** : $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$
- **Norme 2** = norme euclidienne



Norme matricielle

- Norme induite sur $R^{m \times n}$ par la norme vectorielle $\|\,.\,\|$
- Fonction $\|\,.\,\|_{m \times n} : R^{m \times n} \to R$ définie par $\|A\|_{m \times n} = \max_{x \in R^n, x \neq 0} \dfrac{\|Ax\|}{\|x\|}$

Suite

Suite dans R^n

- Suite : $\{ x_k , k=0,1,2,\ldots\} = \{x_0 , x_1 , x_2 ,\ldots, x_n, \ldots\}$

- Limite : $\lim\limits_{k\to\infty} x_k = x^* \Leftrightarrow \lim\limits_{k\to\infty} \|x_k - x^*\|$

Vitesse de convergence

- Convergence linéaire : $\|x_{k+1} - x^*\| \le c \ \|x_k - x^*\|$ avec $0 \le c < 1$
  $\rightarrow$ lent à partir d'un certain rang k_0

- Convergence superlinéaire : $\|x_{k+1} - x^*\| \le c_k \|x_k - x^*\|$ avec $\lim\limits_{k\to\infty} c_k = 0$
  $\rightarrow$ bonne à partir d'un certain rang k_0

- Convergence d'ordre p : $\|x_{k+1} - x^*\| \le c \ \|x_k - x^*\|^p$ avec $0 \le c < 1$
  à partir d'un certain rang k_0

- **Convergence quadratique** si p=2
  $\rightarrow$ rapide

Différentiabilité

Différentiabilité ordre 1

f fonction continue de R^n dans R

Dérivée partielle

Dérivée partielle de f en x par rapport à x_i :

si la limite existe

$$f_{x_i}(x) = \frac{\partial f(x)}{\partial x_i} = \lim_{s \to 0} \frac{f(x_1,..,x_i+s,..,x_n) - f(x_1,..,x_i,..,x_n)}{s}$$

Gradient

Gradient

Gradient de f en x : $g(x) = \nabla f(x)$

$g(x) : R^n \to R^n$

si toutes les dérivées partielles existent

$$\mathbf{g(x)} = \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Dérivée directionnelle

Dérivée directionnelle de f en x dans la direction $d \in R^n$:

si la limite existe

(dérivée directionnelle = produit scalaire avec le gradient)

$$f_d(x) = \lim_{s \to 0} \frac{f(x + sd) - f(x)}{s}$$

$$\Rightarrow \boxed{f_d(x) = g(x)^T d}$$

Fonction différentiable

f différentiable en x $\Leftrightarrow$ f admet une dérivée directionnelle pour tout $d \in R^n$

Hessien

Différentiabilité ordre 2

f fonction deux fois différentiable de R^n dans R

Hessien

Hessien de f en x : $H(x) = \nabla^2 f(x)$

$H(x) : R^n \rightarrow R^{n \times n}$

$$H(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j=1,\dots,n} = \begin{pmatrix} \dfrac{\partial^2 f(x)}{\partial^2 x_1} & \dots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dots & \dots & \dots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \dots & \dfrac{\partial^2 f(x)}{\partial^2 x_n} \end{pmatrix}$$

Jacobien

Matrice gradient

c fonction continue de R^n dans R^m

Gradient de c en x : $\nabla c(x) : R^n \to R^{n\times m}$

$$\boxed{\nabla c(x) = \left(\nabla c_1(x), \ldots, \nabla c_m(x)\right) = \left(\frac{\partial c_j(x)}{\partial x_i}\right)_{\substack{i=1,\ldots,n\\ j=1,\ldots,m}}} = \begin{pmatrix} \frac{\partial c_1(x)}{\partial x_1} & \ldots & \frac{\partial c_m(x)}{\partial x_1} \\ \ldots & \ldots & \ldots \\ \frac{\partial c_1(x)}{\partial x_n} & \ldots & \frac{\partial c_m(x)}{\partial x_n} \end{pmatrix}$$

Matrice jacobienne (« jacobien » = déterminant de J_c)

$$\boxed{J_c(x) = \nabla c(x)^T} = \begin{pmatrix} \nabla c_1(x)^T \\ \ldots \\ \nabla c_m(x)^T \end{pmatrix} = \left(\frac{\partial c_i(x)}{\partial x_j}\right)_{\substack{i=1,\ldots,n\\ j=1,\ldots,m}} = \begin{pmatrix} \frac{\partial c_1(x)}{\partial x_1} & \ldots & \frac{\partial c_1(x)}{\partial x_n} \\ \ldots & \ldots & \ldots \\ \frac{\partial c_m(x)}{\partial x_1} & \ldots & \frac{\partial c_m(x)}{\partial x_n} \end{pmatrix}$$

Méthodes de Newton

- **La méthode de Newton sert à résoudre numériquement un système d'équations non linéaires. Elle est à la base de la majorité des algorithmes d'optimisation.**

Considérons un système de n équations non linéaires à n inconnues.

$$g(x) = 0 \quad \text{avec} \quad g : x \in \mathbb{R}^n \mapsto g(x) \in \mathbb{R}^n$$

La fonction g est supposée différentiable. Écrivons son développement de Taylor à l'ordre 1 en un point x_0.

$$g(x) = g(x_0) + \nabla g(x_0)^T (x - x_0) + o\left(\|x - x_0\|\right)$$

En approximant au voisinage de x_0 la fonction g par la fonction linéaire $\hat{g}_0$:

$$\hat{g}_0(x) \underset{\text{def}}{=} g(x_0) + G_0 (x - x_0) \quad \text{avec} \quad G_0 = \nabla g(x_0)^T$$

et en résolvant le système linéaire :

$$\hat{g}_0(x) = 0$$

on obtient le point x_1 :

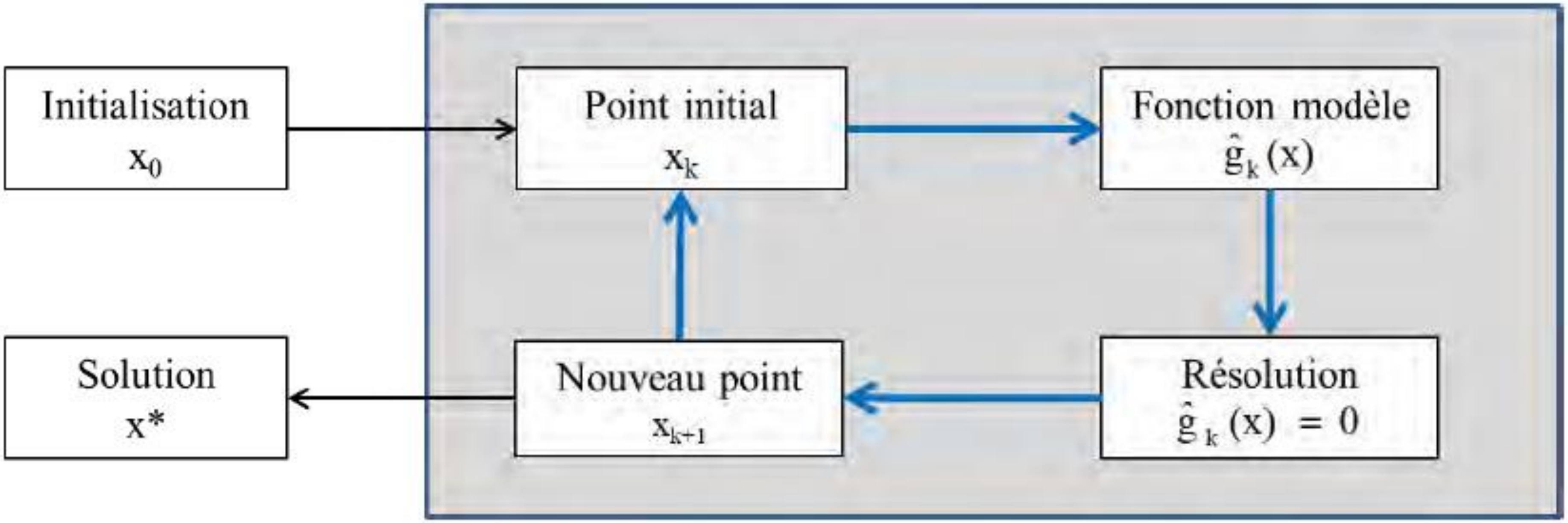
$$x_1 = x_0 - G_0^{-1} g(x_0)$$

La méthode de Newton consiste à itérer le processus jusqu’à obtenir une solution satisfaisante.

Les itérations de Newton sont définies par :

$$x_{k+1} = x_k - G_k^{-1} g(x_k) \quad \text{avec} \quad G_k = \nabla g(x_k)^T$$

Principe de la méthode de Newton



En dimension 1, la méthode de Newton est aussi appelée méthode de la tangente et est donnée par :

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

Exemples

Équation 1 : Résoudre $g(x) = x^2 - 1 = 0$.

Le Tableau 3-1 montre les itérations de Newton à partir de $x_0 = 4$.

La méthode converge vers la solution $x^* = 1$

Itération	x_k	$g(x_k)$	$g'(x_k)$
0	4,00000000	1,5E+01	8,0000
1	2,12500000	3,5E+00	4,2500
2	1,29779412	6,8E-01	2,5956
3	1,03416618	6,9E-02	2,0683
4	1,00056438	1,1E-03	2,0011
5	1,00000016	3,2E-07	2,0000
6	**1,00000000**	**2,5E-14**	**2,0000**

Équation 2 : Résoudre $g(x) = (x-1)^2 = 0$.

Le Tableau montre les itérations de Newton à partir de $x_0 = 4$.

La méthode converge vers la solution $x^* = 1$

Itération	x_k	$g(x_k)$	$g'(x_k)$
0	4,00000000	9,0E+00	6,0000
1	2,50000000	2,3E+00	3,0000
2	1,75000000	5,6E-01	1,5000
3	1,37500000	1,4E-01	0,7500
4	1,18750000	3,5E-02	0,3750
5	1,09375000	8,8E-03	0,1875
6	1,04687500	2,2E-03	0,0938
7	1,02343750	5,5E-04	0,0469
8	1,01171875	1,4E-04	0,0234
9	1,00585938	3,4E-05	0,0117
10	1,00292969	8,6E-06	0,0059
15	1,00009155	8,4E-09	0,0002
20	**1,00000286**	**8,2E-12**	**0,0000**

Équation 3 : Résoudre $g(x) = \operatorname{Arc}\tan x = 0$.

Les deux tableaux montre les itérations de Newton à partir de
x_0 =1,3 ou $_{0}x_0$ =1,5.

Itération	x_k	$g(x_k)$	$g'(x_k)$
0	**1,300**	**0,915**	**0,372**
1	-1,162	-0,860	0,426
2	0,859	0,710	0,575
3	-0,374	-0,358	0,877
4	0,034	0,034	0,999
5	0,000	0,000	1,000
6	**0,000**	**0,000**	**1,000**

Converge x_0 =1,3

Itération	x_k	$g(x_k)$	$g'(x_k)$
0	**1,500**	**0,983**	**0,308**
1	-1,694	-1,038	0,258
2	2,321	1,164	0,157
3	-5,114	-1,378	0,037
4	32,296	1,540	0,001
5	-1575,317	-1,570	0,000
6	**3894976,008**	**1,571**	**0,000**

Diverge pour x_0 =1,5

Ces exemples en dimension 1 montrent que la convergence de la méthode de Newton n'est pas garantie

Problème de minimisation

Considérons à présent un problème de minimisation sans contraintes.

Problème sans contrainte

$$\min_{x \in R^n} f(x) \quad \rightarrow \quad \begin{array}{ll} \text{gradient :} & g(x) = \nabla f(x) \\ \text{hessien :} & H(x) = \nabla^2 f(x) \end{array}$$

Condition nécessaire de minimum local

$$x^* \text{ minimum local} \quad \Rightarrow \quad \begin{cases} g(x^*) = 0 \\ H(x^*) \geq 0 \quad \text{(hessien semi-défini positif)} \end{cases}$$

Recherche des points stationnaires

Application de la méthode de Newton au système d'équations non linéaires : $\boxed{g(x) = 0}$

Modèle linéaire de g en x_k : $\hat{g}_k(x) = g(x_k) + G_k(x - x_k)$ avec $\begin{cases} g(x_k) = \nabla f(x_k) \\ G_k = \nabla g(x_k)^T = \nabla^2 f(x_k) = H_k \end{cases}$

Méthode de **Newton** : G = H $\rightarrow$ calcul explicite du hessien à chaque itération

Méthode de **quasi-Newton** : G = approximation de H
construite à partir des itérations précédentes
sans calcul explicite du hessien

Méthode de Newton

Modèle linéaire de $g=\nabla f$ en x_k

$$\hat{g}_k(x) = g(x_k) + G_k(x - x_k) \quad \text{avec} \quad \begin{cases} g(x_k) = \nabla f(x_k) \\ G_k = \nabla g(x_k)^T = \nabla^2 f(x_k) = H_k \end{cases}$$

- Itération : $x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$ $\rightarrow$ équations de Newton
- Condition d'arrêt : $\|\nabla f(x_k)\| \le \varepsilon$

Difficultés

- Calcul explicite et inversion du hessien $\nabla^2 f(x_k)$ à chaque itération $\rightarrow$ coûteux
- Convergence non garantie même près de la solution
  $\rightarrow$ mêmes difficultés que pour la résolution d'équations

- 1ère condition nécessaire de minimum : $\nabla f(x^*)=0$
  $\rightarrow$ point stationnaire x^* = minimum local, maximum local ou point selle
  $\rightarrow$ 2ème condition nécessaire de minimum à vérifier : $\nabla^2 f(x^*) \ge 0$

Modèle quadratique de f en x_k

- Développement de Taylor à l'ordre 2 de f en x_k

$$f(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k) + o\left(\|x - x_k\|^2\right)$$

- Modèle quadratique en x_k

$$\hat{f}_k(x) = f_k(x_k) + g_k^T (x - x_k) + \frac{1}{2}(x - x_k)^T H_k (x - x_k)$$

- Lien entre le modèle de f et le modèle de g $=\nabla f$

$$\nabla \hat{f}_k(x) = g_k + H_k(x - x_k) = \hat{g}_k(x)$$

Minimisation du modèle de f en x_k

- Conditions suffisantes de minimum local : $\min_{x \in R^n} \hat{f}_k(x) \Leftarrow \begin{cases} \nabla \hat{f}_k(x^*) = \hat{g}_k(x^*) = 0 \\ \nabla^2 \hat{f}_k(x^*) = H_k > 0 \end{cases}$

- Si le hessien de f en x_k est défini positif : $\nabla^2 f(x_k) > 0$

  Minimisation du modèle quadratique de f en x_k

  $\Leftrightarrow$ Méthode de Newton en x_k pour résoudre $\nabla f(x)=0$

- Sinon la méthode de Newton n'est pas directement applicable pour une minimisation