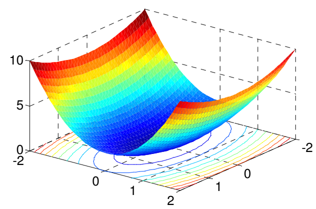
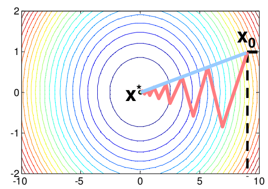
Bloc Math3
« Optimisation et traitement du signal avancé »

École associée
INSTITUT
Mines-Télécom

OPTIMISATION CONTINUE

télécom saint-étienne
école d'ingénieurs nouvelles technologies





S.ROBERT

2014-2015

FI2

6 CM 9 TD

Information pratique

Contenu

I Définitions et formulation

II Résolution d'équations

III Méthodes d'Optimisation sans contrainte

Evaluation du module

Evaluation écrite **40%**

Durée: *1h00*
Documents autorisés:
Polycopiés de Cours, Notes de cours
Calculatrice *fournie*

Evaluation par projet **60%**

Travail collectif *par binôme*
Durée: *9 h*

Chapitre I

Définitions et formulation — I

II

III

Formulation d'un problème d'optimisation

Fonction Objectif

Les contraintes …

I) Définitions et formulation

A) Formulation d'un problème d'optimisation

Optimisation ⇔ Identifier une configuration optimale ou **un optimal** d'un système

Trouver le minimum d'une fonction f(x)

I

II

III

Optimiser un processus physique

⇕

Modélisation Mathématique du processus

- Identifier **les variables de décision** sur lesquelles on peut agir $\vec{x} = \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$
- Définir la mesure de l'état du système par une **fonction objectif** $f(\vec{x})$ ou $f(\mathbf{x}) \in \mathbb{R}$
- Définir par des **contraintes** les valeurs autorisées que peuvent prendre les variables x $x \in X \subset \mathbb{R}$

remarque

Le choix des variables ***x*** *n'est pas forcement immédiat. De plus, il conditionne le résultat de l'optimisation*

I) Définitions et formulation

A) Formulation d'un problème d'optimisation

Exemples de problèmes d'optimisation

Exemple 1

Entreprise de Télécommunication ⟺ Optimisation du lieu d'implantation d'une nouvelle antenne

- 2 antennes existantes : (-5;10) et (5;0)
- Connexion de 4 nouveaux clients

(5;10) 200 heures

(0;12) 200 heures

(10;5) 150 heures

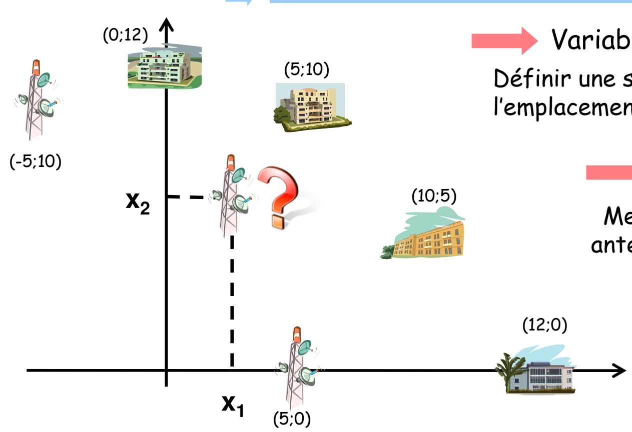
(12;0) 300 heures

- Positon idéale: la plus proche possible du client
- Priorités données aux meilleurs clients
- Respect de la loi : distance minimal entre chaque antenne de 10 km

I
II
III

I) Définitions et formulation

A) Formulation d'un problème d'optimisation



Variables de décision

Définir une solution possible par l'emplacement de l'antenne

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Fonction objectif

Mesurer la pertinence d'une solution potentielle: antenne la plus proche possible du client + priorité pour le meilleur d'entre eux

$$\begin{aligned} f(x_1,x_2) = {} & 200\sqrt{(x_1-5)^2+(x_2-10)^2} \\ & +150\sqrt{(x_1-10)^2+(x_2-5)^2} \\ & +200\sqrt{x_1^2+(x_2-12)^2} \\ & +300\sqrt{(x_1-12)^2+x_2^2} \end{aligned}$$

Contraintes

Respecter une distance minimal avec antennes existantes

$$\sqrt{(x_1+5)^2+(x_2-10)^2} \geq 10$$
$$\sqrt{(x_1-5)^2+x_2^2} \geq 10$$

Problème d'optimisation: $\min_{(x_1,x_2)} f(x_1,x_2)$ sous contraintes $\sqrt{(x_1+5)^2+(x_2-10)^2} \geq 10$, $\sqrt{(x_1-5)^2+x_2^2} \geq 10$

I
II
III

I) Définitions et formulation

A) Formulation d'un problème d'optimisation

Exemple 2

Fabriquant de jouet en bois (soldats + Train) ⟺ Optimisation du coût de production par semaine

	Soldat	Train
Prix de vente :	27 €	21 €
Matériel brut :	10 €	9 €
Coût généraux :	14 €	10 €
Menuiserie :	1 h	1 h
Finissage :	2 h	1 h

- 2 menuisiers disponibles + 2 spécialistes du finissage (40h/semaine chacun)
- Patron travaille 20h/sem au finissage

- Quelque soit la production de trains , elle sera vendue
- Vente maximale de 40 soldats / semaine

Problème d'optimisation: ?

I) Définitions et formulation

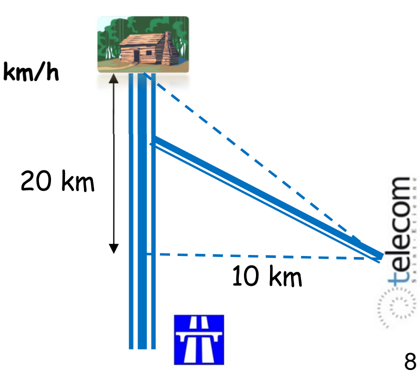
A) Formulation d'un problème d'optimisation

Exemple 3

Enseignant Mr  part en retard de son domicile pour dispenser son cours d'optimisation ⟺ Optimisation du trajet domicile / travail afin d'éviter … le drame !!!

- **Domicile situé à RAUBAIRVILLE rue de l'Electraumague près de l'autoroute**
- **Domicile à 20 km du point sur l'autoroute le plus proche de TELECOM Saint-Etienne**
- **Il peut sortir à tout moment pour emprunter une voie départementale qui le mènera tout droit à son travail**
- **Vitesse moyenne sur autoroute: 110 km/h**
- **Vitesse moyenne sur voie départementale : 70 km/h**

Problème d'optimisation: ?



I) Définitions et formulation

B) Reformulation !!!

Algorithmes et/ou logiciels d'optimisation impose une formulation propre ⟺ Transformation de la fonction objectif

Equivalence

I
II
III

2 problèmes d'optimisation P1 et P2 sont dits équivalents si l'on peut construire un **point admissible** de P2 à partir d'un point admissible de P1 (et réciproquement) avec la même valeur pour la fonction objectif

Transformations simples

Composition par une fonction s : $\mathbb{R} \to \mathbb{R}$
strictement croissante sur $\mathrm{Im}(f) = \{z \mid \exists x \in X \text{ tq } z = f(x)\}$

$$\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} s(f(\mathbf{x})) = s\left(\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x})\right)$$

$$\arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x}) = \arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} s(f(\mathbf{x}))$$

Opérateur Min et argmin

Min identifie le minimum de la fonction
Argmin identifie la valeur des variables qui correspondent au minimum de la fonction

I) Définitions et formulation

B) Reformulation !!!

En particulier …

$\forall c \in \mathbb{R}$ $\arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x}) = \arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} (f(\mathbf{x})+c)$ $\quad$ $\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} (f(\mathbf{x})+c) = c + \min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} (f(\mathbf{x}))$

si $f(x) > 0$ $\arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x}) = \arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} \log(f(\mathbf{x}))$ $\quad$ $\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} \log(f(\mathbf{x})) = \log\left(\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x})\right)$

si $f(x) \geq 0$ $\arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x}) = \arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} (f(\mathbf{x}))^2$ $\quad$ $\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} (f(\mathbf{x}))^2 = \left(\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x})\right)^2$

I
II
III

Etc …

Minimum ⟺ maximum

$$\max_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x}) = -\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} (-f(\mathbf{x}))$$

$$\arg\max_{\mathbf{x}\in X\subseteq\mathbb{R}^n} f(\mathbf{x}) = \arg\min_{\mathbf{x}\in X\subseteq\mathbb{R}^n} (-f(\mathbf{x}))$$

remarque

Dans la suite, on se ramènera souvent à une minimisation plutôt qu'une maximisation

I) Définitions et formulation

B) Reformulation !!!

Transformation des contraintes

remarque *Généralement utilisé lorsque les contraintes sont linéaires*

- Inégalité inférieure ⟺ Inégalité supérieure $\quad g(x) \le 0 \Leftrightarrow -g(x) \ge 0$
- Egalité ⟺ Inégalités $\quad g(x) = 0 \Leftrightarrow \begin{cases} g(x) \le 0 \\ g(x) \ge 0 \end{cases}$
- Changement de variable $x = \tilde{x} + a \quad \text{avec } a \in \mathbb{R}$ $\quad x \ge a \quad \Leftrightarrow \quad \tilde{x} \ge 0$

I II III

Considérons le problème d'optimisation suivant:

$$\max_{x,y} \left(-x^2 + \sin y\right)$$

sous contraintes $\begin{cases} 6x^2 - y^2 \ge 1 \\ x^2 + y^2 = 10 \\ x \ge 2 \\ y \in \mathbb{R} \end{cases}$

Reformuler le problème de manière à obtenir un problème de minimisation dont toutes les variables de décision sont positives ou nulles et les contraintes définies par des inégalités inférieures

I) Définitions et formulation

B) Reformulation !!!

- Inégalité ⟺ Egalité (+ positivité) → Introduction d'une variable de décision: **variable d'écart** z (ou y)

$$g(x) \le 0 \quad \Leftrightarrow \quad g(x) + z^2 = 0 \quad \text{ou} \quad \begin{cases} g(x) + y = 0 \\ y \ge 0 \end{cases}$$

remarque *La fonction objectif n'est pas affectée par l'introduction des variables de d'écart*

I II III

Considérons le problème d'optimisation suivant:

$$\min_{x_1, x_2} \left(x_1^2 - x_2^2\right)$$

sous contraintes $\begin{cases} \sin x_1 \le \dfrac{\pi}{2} \\ \ln\left(e^{x_1} + e^{x_2}\right) \ge \sqrt{e} \end{cases}$

Reformuler le problème de manière à obtenir des contraintes définies par des égalités et/ou des contraintes de positivité par l'introduction de variables d'écart

I) Définitions et formulation

C) Hypothèses restrictives

Domaine de l'optimisation très vaste !!! ⟺ Restrictions du cours

I
II
III

Hypothèse de continuité — sinon — **Optimisation combinatoire ou programmation en nombres entiers**

(Cf cours Option math TSE2)

Exemple 2 ne vérifie pas l'hypothèse

Fonction objectif et fonctions définissant les contraintes ⟺ Fonctions continues des variables de décision

Hypothèse de différentiabilité (implique l'hypothèse de continuité) — sinon — **Optimisation non différentiable**

Fonction objectif et fonctions définissant les contraintes ⟺ Fonctions différentiables des variables de décision

Hypothèse de déterminisme — sinon — **Optimisation stochastique** / **Optimisation robuste**

Ignorer les erreurs de modèles (bruit de mesure, …)

(Cf cours Option math TSE2)

I) Définitions et formulation

D) Définitions complètes

Problème d'optimisation

à n variables de décision, m contraintes d'égalités et p contraintes d'inégalités

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \quad \text{sous contraintes} \quad \begin{cases} h(\mathbf{x})=0 \\ g(\mathbf{x})\le 0 \end{cases} \quad \text{et} \quad \mathbf{x}\in X$$

I
II
III

- $f:\mathbb{R}^n \to \mathbb{R}$
- $h:\mathbb{R}^n \to \mathbb{R}^m$
- $g:\mathbb{R}^n \to \mathbb{R}^p$
- $X \subseteq \mathbb{R}^n$ ensemble convexe des solutions (admissible ou non)

Ensemble convexe

$X \subseteq \mathbb{R}^n$ est convexe si $\forall x, y \in X$
on a $\lambda x + (1-\lambda) y \in X$ avec $0 \le \lambda \le 1$

L'ensemble Y des **points admissibles** est défini par :

$$Y = \left\{ \mathbf{x}\in\mathbb{R}^n \,\middle|\, h(\mathbf{x})=0,\ g(\mathbf{x})\le 0 \text{ et } \mathbf{x}\in X \right\}$$

D) Définitions complètes

Types de solutions

Minimum local
Aucun de ses voisins immédiats admissibles n'est meilleur ⟺ $x^* \in Y$ est minimum local s'il existe $\varepsilon > 0$ tq

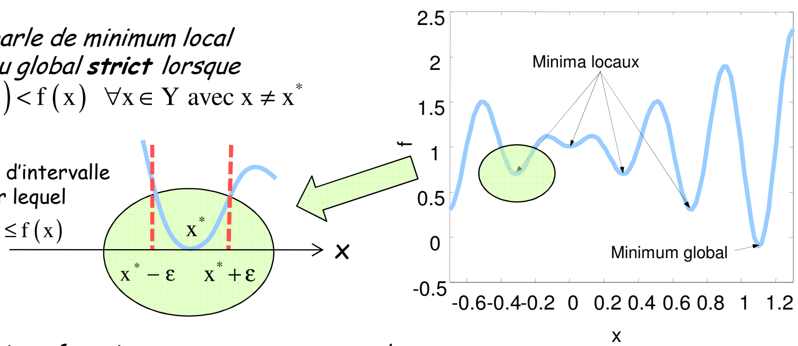
$$f(x^*) \le f(x) \quad \forall x \in Y \text{ tq } \|x - x^*\| < \varepsilon$$

Minimum global
Aucun des points admissibles n'est meilleur ⟺ $x^* \in Y$ est minimum global si

$$f(x^*) \le f(x) \quad \forall x \in Y$$

remarque — *On parle de minimum local et/ou global **strict** lorsque* $f(x^*) < f(x) \quad \forall x \in Y \text{ avec } x \ne x^*$

Exemple d'intervalle pour lequel $f(x^*) \le f(x)$



remarque — *Certaines fonctions peuvent comporter des minima locaux sans qu'un minimum global n'existe*

D) Définitions complètes

Infimum

$f : \mathbb{R}^n \to \mathbb{R}$ non bornée

Possibilité de problème sans optimum

$$\min_{x \in \mathbb{R}} f(x) = \min_{x \in \mathbb{R}} x$$

f bornée inférieurement sur $Y \subseteq \mathbb{R}^n$

$$\min_{x \in \mathbb{R}} f(x) = \min_{x \in \mathbb{R}} e^x$$

Infimum de f sur Y

$$\inf_{y \in Y} f(y) \text{ tq } \begin{cases} \inf_{y \in Y} f(y) \le f(x) \quad \forall x \in Y \\ \text{et} \\ \forall M > \inf_{y \in Y} f(y) \quad \exists x_0 \in Y \text{ tq } f(x_0) < M \end{cases}$$

Théorème de Weierstrass

Si l'infimum d'une fonction sur ensemble est atteint par au moins un point de l'ensemble alors il constitue un minimum global du problème d'optimisation correspondant

II) Fonction objectif

A) Convexité de la fonction objectif

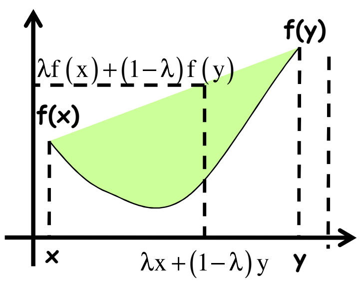
Caractéristique très importante en optimisation ⟺ Pratiquement impossible d'identifier un optimum global pour une fonction non convexe !!!

Définition formelle

$f : \mathbb{R}^n \to \mathbb{R}$ est **convexe**
si pour tout $x, y \in \mathbb{R}^n$ et pour tout $\lambda \in [0,1]$, on a
$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

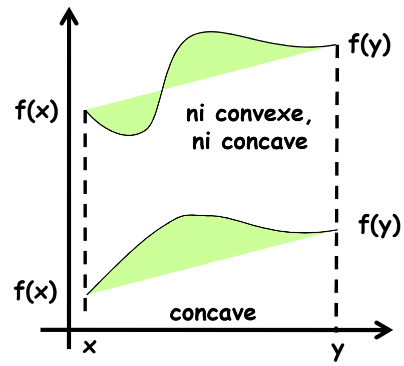
*On parle de convexité **stricte** lorsque* $f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$ *pour tout* $x \neq y$

Interprétation graphique (n=1)



convexe

Segment au dessus de la courbe de f entre x et y



La notion de concavité est rattachée à des problèmes de maximisation

II) Fonction objectif

B) Dérivabilité

1^er ordre $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction continue

$i^{\text{ème}}$ Dérivée partielle de f

$$\nabla_i f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{\alpha \to 0} \frac{f(x_1, x_2, \ldots x_i + \alpha, \ldots x_n) - f(x_1, x_2, \ldots x_i, \ldots x_n)}{\alpha}$$

Gradient de f (Cf cours Electromagnétisme …)
si les limites précédentes existent…

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial f(\mathbf{x})}{\partial x_1} \\ \dfrac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

De façon plus générale si
$f : \mathbb{R}^n \to \mathbb{R}^m$ telle que $f_i : \mathbb{R}^n \to \mathbb{R}$ est différentiable

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial f_1(\mathbf{x})}{\partial x_1} & \dfrac{\partial f_2(\mathbf{x})}{\partial x_1} & \cdots & \dfrac{\partial f_m(\mathbf{x})}{\partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_1(\mathbf{x})}{\partial x_n} & \dfrac{\partial f_2(\mathbf{x})}{\partial x_n} & \cdots & \dfrac{\partial f_m(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

Matrice gradient

La matrice Jacobienne est définie par: $J(\mathbf{x}) = \nabla f(\mathbf{x})^T$

II) Fonction objectif

B) Dérivabilité

Dérivée directionnelle de f en x dans la direction $\mathbf{d} \in \mathbb{R}^n$
ou **dérivée de Gateaux**

$$\lim_{\alpha \to 0} \frac{f(\mathbf{x}+\alpha\mathbf{d})-f(\mathbf{x})}{\alpha} = \nabla f(\mathbf{x})^T \mathbf{d} = \mathbf{d}^T \nabla f(\mathbf{x})$$

I
II
III

remarque

*Dérivée directionnelle pour **d=e**$_i$* ⇔ *dérivée partielle* $\nabla_i f(\mathbf{x})$

Si pour tout $\mathbf{d} \in \mathbb{R}^n$, la dérivée directionnelle de f dans la direction **d** existe ⇒ f est **différentiable** ou **Gateaux-différentiable**

Direction de descente

Soient
- f: $\mathbb{R}^n \to \mathbb{R}$ une fonction différentiable
- $\mathbf{x}, \mathbf{d} \in \mathbb{R}^n$

d est une direction de descente si $\mathbf{d}^T \nabla f(\mathbf{x}) < 0$

II) Fonction objectif

B) Dérivabilité

Parmi toutes les directions **d** à partir d'un point x, celle où la **descente est la plus forte** est la direction opposée au gradient de f

$\mathbf{d} \in \mathbb{R}^n$ tel que $\|\mathbf{d}\| = \|\nabla f(\mathbf{x})\|$
$-\nabla f(\mathbf{x})^T \nabla f(\mathbf{x}) \geq \mathbf{d}^T \nabla f(\mathbf{x})$

I
II
III

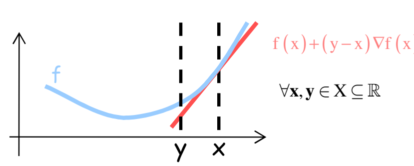
Convexité par le calcul du gradient

Soit $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable sur un ensemble convexe ouvert X

f **est convexe sur X** ssi
$f(\mathbf{y}) - f(\mathbf{x}) \geq (\mathbf{y}-\mathbf{x})^T \nabla f(\mathbf{x}) \qquad \forall \mathbf{x}, \mathbf{y} \in X$

⇔ $f(\mathbf{y}) \geq \underbrace{f(\mathbf{x}) + (\mathbf{y}-\mathbf{x})^T \nabla f(\mathbf{x})}_{\text{Eq. hyperplan tangent à f en x}} \qquad \forall \mathbf{x}, \mathbf{y} \in X$

Une fonction f est convexe ssi son graphe se trouve au dessus des hyperplans tangents



II) Fonction objectif

B) Dérivabilité

2nd ordre

$j^{\text{ième}}$ Dérivée partielle de $\nabla_i f(\mathbf{x})$

(Cf cours Electromagnétisme …)

$$\boxed{\frac{\partial \nabla_i f(\mathbf{x})}{\partial x_j} = \frac{\partial\left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right)}{\partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}}$$

Hessien de f

Les dérivées seconde précédentes sont regroupées dans la **matrice Hessienne** de f

$f: \mathbb{R}^n \to \mathbb{R}$ une fonction 2 fois différentiable

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ & & & \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

Matrice Hessienne

remarque

C'est le gradient de la matrice Jacobienne

II) Fonction objectif

B) Dérivabilité

Convexité par le calcul du hessien

Comme le gradient, le hessien donne des informations sur la convexité éventuelle de f

Soit $f: X \subseteq \mathbb{R}^n \to \mathbb{R}$ une fonction 2 fois différentiable sur un ensemble convexe ouvert X

f est **convexe (resp. strictement convexe)** dans X si
$\nabla^2 f(\mathbf{x})$ est semi définie positive (resp. définie positive) pour tout **x** dans X

Matrice semi-définie positive

Matrice carrée $A \in \mathbb{R}^{n\times n}$ telle que $\mathbf{x}^T A\mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$

Si A symétrique $\Rightarrow$ $\lambda \geq 0$

Matrice définie positive

Matrice carrée $A \in \mathbb{R}^{n\times n}$ telle que $\mathbf{x}^T A\mathbf{x} > 0 \quad \forall \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x} \neq 0$

Si A symétrique $\Rightarrow$ $\lambda > 0$

Si A définie positive $\Rightarrow$ A^{-1} existe et est définie positive

II) Fonction objectif

B) Dérivabilité

I
II
III

Courbure

Dérivée seconde $\Leftrightarrow$ Information sur la courbure

Considérons $g : \mathbb{R} \to \mathbb{R}$ tq $g(\alpha) = f(\mathbf{x} + \alpha \mathbf{d})$

Dérivons 1 fois $g'(\alpha) = \mathbf{d}^T \nabla f(\mathbf{x} + \alpha \mathbf{d})$ $\Leftrightarrow$ $g'(\alpha = 0)$ Dérivée directionnelle de f en x

Dérivons 2 fois $g''(\alpha) = \mathbf{d}^T \nabla^2 f(\mathbf{x} + \alpha \mathbf{d}) \mathbf{d}$ $\Leftrightarrow$ $g''(\alpha = 0)$ Proportionnelle à la courbure de f en x

Normalisation pour éviter l'influence de la longueur de **d** (en divisant par d^2) $\Leftrightarrow$ **Courbure** de f en x dans la direction d

quotient de Rayleigh $\dfrac{\mathbf{d}^T \nabla^2 f(\mathbf{x}) \mathbf{d}}{\mathbf{d}^T \mathbf{d}}$

remarque

*Courbure de f en **-d** est identique à la courbure de f en **d***

II) Fonction objectif

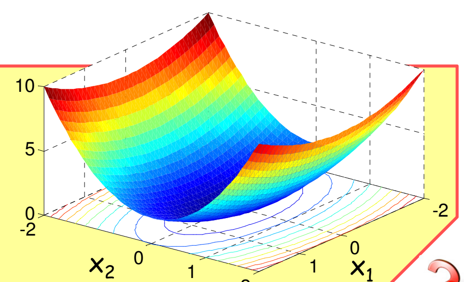
B) Dérivabilité

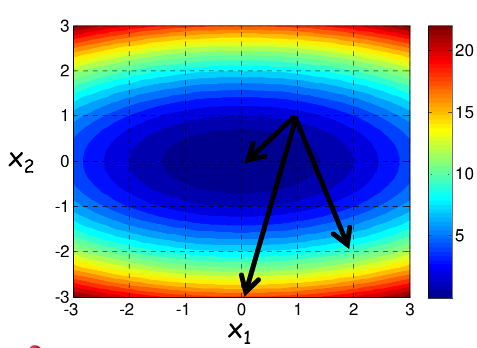
I
II
III

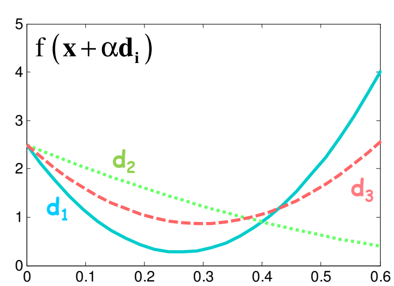
Soit $f(\mathbf{x}) = \frac{1}{2} x_1^2 + 2x_2^2$

Calculer les dérivées directionnelles et la courbure de f au point $(1 \quad 1)^T$ dans chacune des directions suivantes :

$d_1 = -\nabla f(\mathbf{x}), \quad d_2 = (-1 \quad -1)^T, \quad d_3 = (1 \quad -3)^T$







$f(\mathbf{x} + \alpha \mathbf{d}_i)$ ne diminue pas continuellement !!!

Propriété de descente est locale

remarque

On peut vérifier que la fonction est convexe sur un intervalle donné

II) Fonction objectif

C) Linéarité et non linéarité

Fonction linéaire

$f:\mathbb{R}^n \to \mathbb{R}$ linéaire si elle s'écrit
$f(x) = c^T\mathbf{x}$ où $c \in \mathbb{R}^n$ constant

$f:\mathbb{R}^n \to \mathbb{R}^m$ linéaire si chacune de ces composantes
$f_i:\mathbb{R}^n \to \mathbb{R}$ est linéaire
$f(x) = A\mathbf{x}$ où $A \in \mathbb{R}^{m\times n}$ constant

I
II
III

Fonction affine

$f:\mathbb{R}^n \to \mathbb{R}$ Affine si elle s'écrit
$f(x) = c^T\mathbf{x} + d$ où $c \in \mathbb{R}^n, d \in \mathbb{R}$ constant

$f:\mathbb{R}^n \to \mathbb{R}^m$ affine si chacune de ces composantes
$f_i:\mathbb{R}^n \to \mathbb{R}$ est affine
$f(x) = A\mathbf{x} + \mathbf{b}$ où $A \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$ constant

II) Fonction objectif

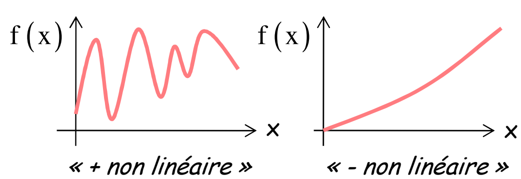
C) Linéarité et non linéarité

Fonction non linéaire

Toute fonction qui n'est pas affine est non linéaire

Définition

La continuité du gradient au sens de Lipschitz permet de quantifier le « degré » de non linéarité.



I
II
III

Soit $f: X \subseteq \mathbb{R}^n \to \mathbb{R}$. La matrice gradient est continue au sens de *Lipschitz sur X si*

$$\exists M \text{ tq } \left\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right\|_n \le M \left\|\mathbf{x} - \mathbf{y}\right\|_n \text{ pour tout } \mathbf{x},\mathbf{y} \in X$$

- M correspond à une borne supérieure de la courbure
- M est appelé constante de Lipschitz
- M = 0 ⇒ ∇f constant sur X ⇒ F linéaire
- M constante théorique très difficile à calculer

Norme vectorielle

La norme-p de $\mathbf{x}$ est définie par la fonction $\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}$ telle que

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^{p} |x_i|^p} \qquad x \in \mathbb{R}^n$$

II) Fonction objectif

C) Linéarité et non linéarité

Cas particulier : fonction quadratique

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ sera dite **quadratique** si elle peut s'écrire

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$$

Q matrice symétrique de $n \times n$

$c \in \mathbb{R},\ \mathbf{b} \in \mathbb{R}^n$

$\Rightarrow$ $$\nabla f(\mathbf{x}) = Q\mathbf{x} + \mathbf{b}$$

$\Rightarrow$ $$\nabla^2 f(\mathbf{x}) = Q$$

remarque

Le facteur $\frac{1}{2}$ sert seulement à simplifier les écritures du gradient

Exemple de fonction quadratique définie par : $Q = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}$ et $b = (1 \quad 0)^T$ et $c = -4$

Donner l'expression analytique de : $f(x, y)$

II) Fonction objectif

D) Conditionnement

Conditionnement de f en x $=$ **Nombre de conditionnement** de la matrice $\nabla^2 f(\mathbf{x})$

Définition formelle

Soit $A \in \mathbb{R}^{n\times n}$ matrice symétrique non singulière

Nb de conditionnement de A $\kappa(A) = \|A\|\|A^{-1}\|$

si norme-2 euclidienne

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$$

remarque

Si A est symétrique semi définie positive alors $\sigma_i = \lambda_i$

Matrice singulière

Matrice singulière A telle que $\det(A) = 0$

Matrice non singulière = matrice inversible ou régulière

Valeurs/vecteurs propres

valeurs propres λ_i de $A \in \mathbb{R}^{n\times n}$

racines du polynôme caractéristique $\det(\lambda I - A) = 0$

vecteurs propres **x** tq : $A\mathbf{x} = \lambda\mathbf{x}$

Valeurs singulières

valeurs singulières σ_i de $A \in \mathbb{R}^{m\times n}$

si il existe 2 matrices orthogonales

$U \in \mathbb{R}^{m\times m}$ et $V \in \mathbb{R}^{n\times n}$ tel que :

$U^T AV = \mathrm{diag}(\sigma_1, \sigma_2, \cdots \sigma_p)$

avec $p = \min(m, n)$ et $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$

II) Fonction objectif

D) Conditionnement

Théorème de Rayleigh-Ritz

$A \in \mathbb{R}^{n\times n}$ matrice réelle symétrique
$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ vp de A

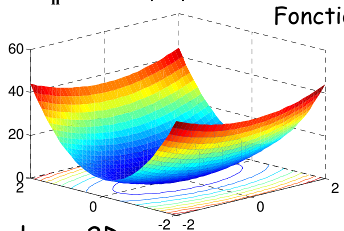
$$\lambda_1 = \max_{x\neq 0} \frac{x^T A x}{x^T x} \quad \text{et} \quad \lambda_n = \min_{x\neq 0} \frac{x^T A x}{x^T x}$$

Interprétation géométrique

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction non linéaire et $\mathbf{x} \in \mathbb{R}^n$
Supposons la matrice $\nabla^2 f(\mathbf{x})$ (forcement symétrique) définie positive

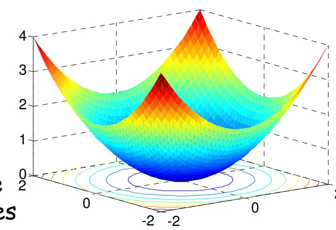
- λ_1 la plus grande vp
  $\mathbf{d}_1$: un vect propre
  ⇒ $A\mathbf{d}_1 = \lambda_1 \mathbf{d}_1$ $\xrightarrow[\text{normalisation}]{\mathbf{d}_1^T \times}$ $\lambda_1 = \frac{\mathbf{d}_1^T A \mathbf{d}_1}{\mathbf{d}_1^T \mathbf{d}_1}$ $\xrightarrow[\text{Rayleigh-Ritz}]{\text{Théorème de}}$ **+ grande courbure** est dans la direction $\mathbf{d}_1$

- λ_n la plus petite vp
  $\mathbf{d}_n$: un vect propre
  ⇒ $A\mathbf{d}_n = \lambda_n \mathbf{d}_n$ $\xrightarrow[\text{normalisation}]{\mathbf{d}_n^T \times}$ $\lambda_n = \frac{\mathbf{d}_n^T A \mathbf{d}_n}{\mathbf{d}_n^T \mathbf{d}_n}$ $\xrightarrow[\text{Rayleigh-Ritz}]{\text{Théorème de}}$ **+ petite courbure** est dans la direction $\mathbf{d}_n$



Exemple en 2D

Fonction **mal conditionnée**
grande différence de courbure entre 2 directions
⇒ Ellipses allongées



Fonction **bien conditionnée**
Homogénéité des courbures dans diverses directions
⇒ courbes de niveau circulaire ($\kappa_2 = 1$)

II) Fonction objectif

D) Conditionnement

Changement de variable

Permet la diminution du conditionnement

$\boxed{\mathbf{x}' = M\mathbf{x}}$ M: matrice inversible
$\mathbf{x} \in \mathbb{R}^n$

$$\tilde{f}(\mathbf{x}') = f(M^{-1}\mathbf{x}') = f(\mathbf{x}) \Rightarrow \begin{cases} \nabla \tilde{f}(\mathbf{x}') = M^{-T} \nabla f(M^{-1}\mathbf{x}') \\ \nabla^2 \tilde{f}(\mathbf{x}') = M^{-T} \nabla^2 f(M^{-1}\mathbf{x}') M^{-1} = M^{-T} \nabla^2 f(\mathbf{x}) M^{-1} \end{cases}$$

Conditionnement de $\tilde{f}$ en $\mathbf{x}'$ est le nb de conditionnement de de la matrice $\nabla^2 \tilde{f}(\mathbf{x}')$ soit $M^{-T}\nabla^2 f(\mathbf{x}) M^{-1}$

Choisir M tq le conditionnement le plus proche possible de 1 ⇔ **préconditionnement**

remarque

Le meilleur conditionnement possible ($\kappa_2 = 1$) est obtenu en effectuant le changement de variable sur la base de la factorisation de Cholesky de la matrice $\nabla^2 f(x)$ soit $M = L^T$

Décomposition de Cholesky

Soit $A \in \mathbb{R}^{n\times n}$
matrice symétrique définie positive
Décomposition de A tq : $A = LL^T$
$L \in \mathbb{R}^{n\times n}$:matrice triangulaire inférieure

Soit $f(\mathbf{x}) = \frac{1}{2}x_1^2 + \frac{25}{2}x_2^2 + 3x_1x_2 - 12x_1 - \sqrt{\pi}x_2 - 6$
Transformer f de façon à obtenir un conditionnement de 1

III) Les contraintes …

A) Contraintes actives

Définition formelle Soient $g:\mathbb{R}^n \to \mathbb{R}^p$ et $h:\mathbb{R}^n \to \mathbb{R}^m$

une contrainte d'inégalité $g_i(\mathbf{x}) \le 0$ est dite **active** en **x*** si $g_i(\mathbf{x}^*) = 0$ pour i=1,…p

est dite **inactive** en **x*** si $g_i(\mathbf{x}^*) < 0$ pour i=1,…p

une contrainte d'égalité $h_i(\mathbf{x}) = 0$ est dite **active** en **x*** si $h_i(\mathbf{x}^*) = 0$ pour i=1,…m

L'ensemble des indices i des contraintes actives en **x*** est noté $A(\mathbf{x}^*)$

I

II

III

Interprétation géométrique Soit le problème d'optimisation $\min_{x\in\mathbb{R}} x^2$ sous contraintes:

$x \le 4$ ⟺ $g_1(x) = x - 4 \le 0$

$x \ge -10$ ⟺ $g_2(x) = -x - 10 \le 0$

sol : x*=0 $g_1(x^*) = -4 < 0$ et $g_2(x^*) = -10 < 0$

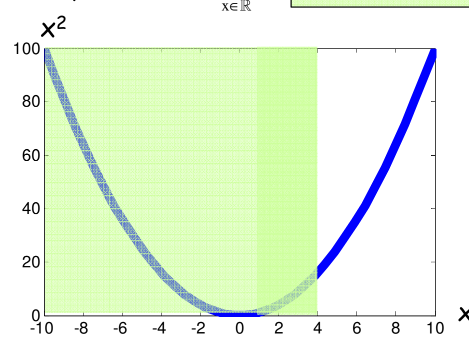
Possibilité d'ignorer les contraintes sans incidence

$x \le 4$ ⟺ $g_1(x) = x - 4 \le 0$

$x \ge 1$ ⟺ $g_2(x) = 1 - x \le 0$

sol : x*=1 $g_1(x^*) = -3 < 0$ et $g_2(x^*) = 0$

Seul g_1 peut être ignorée sans incidence



III) Les contraintes …

A) Contraintes actives

Théorème des contraintes actives

Permet de supprimer les contraintes inactives

Réduire les contraintes du problème

Soit $g:\mathbb{R}^n \to \mathbb{R}^m$ une fonction non linéaire et **x*** un point admissible

I

II

III

x* est un minimum local du problème

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

sous contraintes

$$g(\mathbf{x}) \le 0$$

$$\mathbf{x} \in Y \subseteq \mathbb{R}^n$$

ssi

x* est un minimum local du problème

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

sous contraintes

$$g_i(\mathbf{x}) = 0 \text{ pour } i \in A(\mathbf{x}^*)$$

$$\mathbf{x} \in Y \subseteq \mathbb{R}^n$$

A(**x***): Ensemble des indices des contraintes actives en **x***

remarque

*Intérêt limité car il faudrait connaitre la solution **x*** !!!*

III) Les contraintes …

B) Gestion des contraintes

Gestion des contraintes dans les algorithmes ⟺ complexe !!!

- Hypothèses nécessaires : Indépendance linéaire
  Etc …

I
II
III

- Nombreuses méthodes :
  - Elimination des contraintes redondantes
  - Pb d'optimisation avec contraintes ⇒ Pb d'optimisation sans contraintes
    + simple à résoudre

Limites du cours : Optimisation sans contraintes

remarque

La gestion des contraintes dans le cadre de l'optimisation combinatoire fait l'objet du cours « programmation par contraintes » du bloc option Fi-OptMath 1