

Informatique & Mathématiques Appliquées
Optimisation



J. Gergaud & S. Gratton & D. Ruiz

30 avril 2013

Table des matières

1 Exemples et définitions	3
I Exemples	3
I.1 Cas continu et de dimension finie	3
I.2 Problèmes en nombres entiers	10
I.3 Problème en dimension infinie	11
II Problème d'optimisation	13
II.1 Définitions	13
II.2 Classification	16
III Exercices	18
2 Outils mathématiques	23
I Introduction	23
II Algèbre linéaire	23
II.1 rappels	23
III Calcul différentiel	24
III.1 Notations	24
III.2 Théorème des fonctions composées	25
III.3 Formule de Taylors	26
III.4 Courbes de niveau	26
III.5 Surfaces et plan tangent dans \mathbf{R}^3	27
IV Convexité des applications	27
IV.1 Ensembles convexes - applications convexes	27
IV.2 Convexité et dérivée première	28
IV.3 Convexité et dérivée seconde	29
IV.4 Illustrations	30
V Exercices	30
V.1 Avec corrections	30
3 Existence de solution, unicité de solution	33
I Introduction	33
II Existence de solution	33
II.1 Problèmes avec contraintes	33
II.2 Problème sans contraintes	34
III Cas convexe	34
IV Exercices	36
IV.1 Corrigés des exercices	36
IV.2 Exercices sans corrections	36
4 Condition nécessaire, condition suffisante de solution	39
Cas sans contraintes et cas de contrainte convexe	39
I Condition du premier ordre	39
I.1 Cas sans contraintes	39
I.2 Cas de contraintes convexes	39
I.3 Problèmes convexes	40
II Conditions du deuxième ordre	40
II.1 Condition nécessaire	40
II.2 Condition suffisante	41
III Exercices	42

5	Problèmes aux moindres carrés	45
I	Introduction	45
II	Les moindres carrés linéaires	45
	II.1 Rappels	45
	II.2 Approximation d'une fonction au sens des moindres carrés	46
III	La méthode de Newton	47
	III.1 Introduction	47
	III.2 Algorithme de Newton pour résoudre $f(x) = 0$	47
	III.3 Résolution d'équations : cas de la dimension n	48
	III.4 Convergence	49
	III.5 Application aux problèmes d'optimisation	50
IV	Résolution des problèmes aux moindres carrés non linéaires	51
	IV.1 Algorithme de Newton	51
	IV.2 Algorithme de Gauß-Newton	51
	IV.3 Exemples	51
6	Problèmes avec contraintes	55
I	Introduction	55
II	Conditions du premier ordre	56
	II.1 Conditions du premier ordre : contraintes sous forme égalités	56
	II.2 Qualification des contraintes	59
	II.3 Théorème de Karuch, Kuhn et Tucker	61
	II.4 Cas convexe	62
III	Conditions du second ordre	63
	III.1 Conditions Nécessaires du second ordre	63
	III.2 Conditions suffisantes	64
IV	Exercices	65
	IV.1 Avec corrections	65
7	Introduction à la dualité	67
I	Point selle	67
	I.1 Définition	67
	I.2 Caractérisation	68
II	Condition suffisante de solution	68
	II.1 Point selle de lagrangien	68
	II.2 Condition suffisante de solution	69
III	Dualité	69
	III.1 Définition du problème dual et théorèmes	69
	III.2 Cas d'un problème linéaire	70
IV	Exercice	70

Introduction

Optimiser, c'est rechercher parmi un ensemble C de choix possibles le meilleur (s'il existe !). Si f est une application d'un ensemble E dans F . On note le problème

$$(P) \begin{cases} \min f(x) \\ x \in C \subset E. \end{cases}$$

Il faut donc pour cela pouvoir comparer 2 choix et donc avoir une structure d'ordre sur l'ensemble F . On prendra toujours $F = \mathbf{R}$. Suivant les domaines d'applications :

- E s'appelle l'ensemble des stratégies, des états, des paramètres, l'espace ;
- C est l'ensemble des contraintes ;
- f est la fonction coût, économique ou le critère, l'objectif.

Une fois le problème bien défini, il se pose deux questions. La première est de savoir si (P) admet une solution. Si la réponse est positive, il nous faut trouver la ou les solutions. Suivant la nature de l'ensemble E les réponses sont plus ou moins faciles. Si E est fini, l'existence de solution est évidente, mais le calcul est difficile si le nombre d'éléments est grand. Par contre si $E = \mathbf{R}^n$ ou est de dimension infinie la question de l'existence de solution est moins triviale, mais si les fonctions sont dérivables il est "plus" facile de calculer la solution.

Chapitre 1

Exemples et définitions

I Exemples

I.1 Cas continu et de dimension finie



FIGURE 1.1 – Pierre de Fermat, né vers 1601, à Beaumont-de-Lomagne, près de Montauban, et mort le 12 janvier 1665 à Castres.

Exemple I.1.1 (Principe de Fermat)

Pierre de Fermat est un juriste et mathématicien français, surnommé « le prince des amateurs ». On lui doit entre autre le principe de Fermat qui dit que la lumière se propage d'un point à un autre sur des trajectoires telles que la durée du parcours soit minimale. Il imagina aussi pour la solution des problèmes, une méthode, dite de maximis et minimis, qui le fait regarder comme le premier inventeur du calcul différentiel dont il est un précurseur : il est le premier à utiliser la formule (sinon le concept) du nombre dérivé !¹

On s'intéresse ici à la trajectoire d'un rayon lumineux d'un point $A(0, a)$ vers un point $B(k, b)$ situés dans deux milieux homogènes différents (cf. la figure 1.2). Nous allons grâce au principe de Fermat retrouver la loi de la réfraction. On suppose pour cela que la trajectoire d'un rayon lumineux dans un milieu homogène est un segment de droite (ce qui peut aussi se démontrer grâce au principe de Fermat via le calcul des variations, cf. l'exemple I.3.1, qui est un problème d'optimisation en dimension infinie !).

On note P , de coordonnées $(x, 0)$, le point d'impact du rayon lumineux sur la surface du changement de milieu et c_1 et c_2 les vitesses de la lumière dans l'air et dans l'eau. Le temps de parcours entre les point A et B est donc

$$T(x) = \frac{1}{c_1} \sqrt{a^2 + x^2} + \frac{1}{c_2} \sqrt{b^2 + (k - x)^2}.$$

1. http://fr.wikipedia.org/wiki/Pierre_de_Fermat.

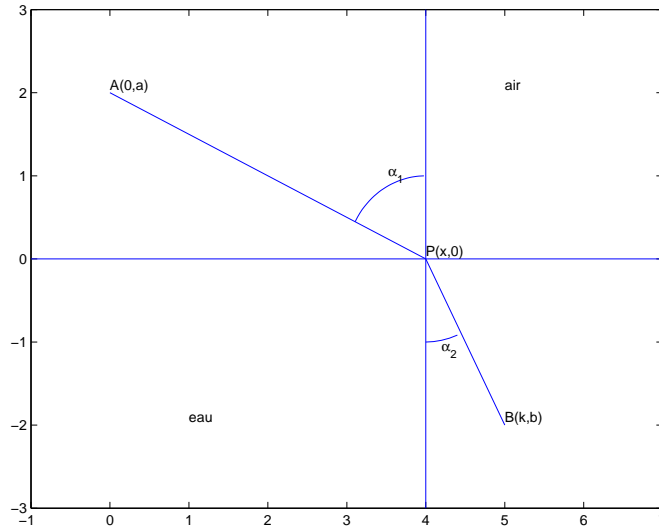


FIGURE 1.2 – Principe de Fermat.

Le problème est alors ici de trouver le point P (c'est-à-dire $x^* \in \mathbf{R}$) tel que

$$T(x^*) \leq T(x) \forall x \in \mathbf{R} \iff (P) \begin{cases} \text{Min } f(x) \\ x \in \mathbf{R}. \end{cases}$$

On peut ici tracer cette fonction (cf. la figure 1.3).

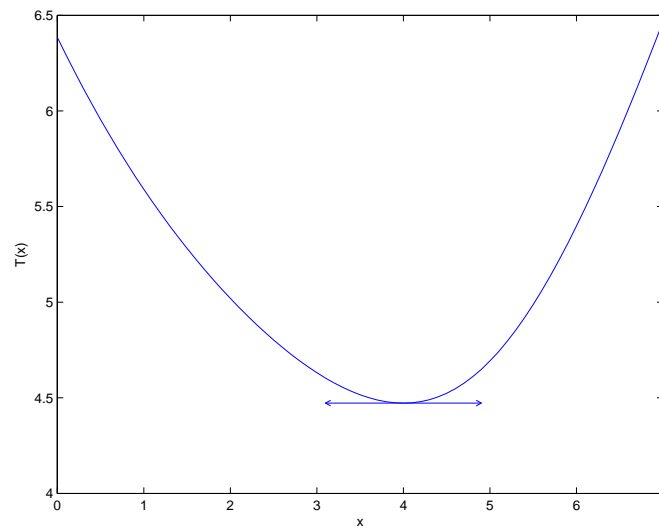


FIGURE 1.3 – Fonction T .

Une condition nécessaire de solution de (P) est $T'(x) = 0$ (cf. la figure 1.3). Ce qui donne ici

$$\begin{aligned} T'(x) &= \frac{x}{c_1 \sqrt{a^2 + x^2}} + \frac{-(k-x)}{c_2 \sqrt{b^2 + (k-x)^2}} = 0 \\ \iff \frac{x}{c_1 \sqrt{a^2 + x^2}} &= \frac{(k-x)}{c_2 \sqrt{b^2 + (k-x)^2}} \\ \iff \frac{\sin \alpha_1}{c_1} &= \frac{\sin \alpha_2}{c_2} \\ \iff n_1 \sin \alpha_1 &= n_2 \sin \alpha_2. \end{aligned}$$

Remarque I.1.2

Nous retrouvons dans ce cas les lois de Descartes² ou de Snell.

Remarque I.1.3

La condition $T'(x) = 0$ n'est qu'une condition nécessaire, en effet si nous considérons la fonctionnelle réelle $f(x) = x^3$ nous avons $f'(0) = 0$ mais 0 n'est pas un minimum de f (1.4).

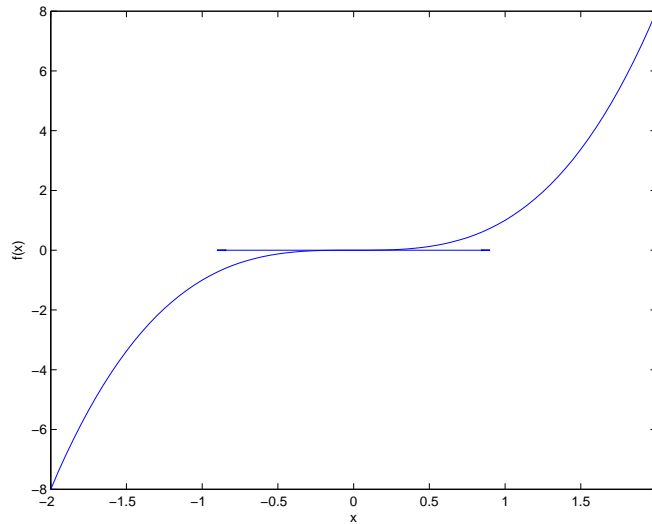


FIGURE 1.4 – $f'(0) = 0$ et 0 n'est pas un minimum.

Exemple I.1.4 (Datation par le carbone 14)

Le carbone radioactif ^{14}C est produit dans l'atmosphère par l'effet des rayons cosmiques sur l'azote atmosphérique. Il est oxydé en $^{14}CO_2$ et absorbé sous cette forme par les organismes vivants qui, par suite, contiennent un certain pourcentage de carbone radioactif relativement aux carbones ^{12}C et ^{13}C qui sont stables. On suppose que la production de carbone ^{14}C atmosphérique est demeurée constante durant les derniers millénaires. On suppose d'autre part que, lorsqu'un organisme meurt, ses échanges avec l'atmosphère cessent et que la radioactivité due au carbone ^{14}C décroît suivant la loi exponentielle suivante :

$$A(t) = A_0 e^{-\lambda t}$$

où λ est une constante positive, t représente le temps en année et $A(t)$ est la radioactivité exprimée en nombre de désintégrations par minute et par gramme de carbone. On désire estimer les paramètres A_0 et λ par la méthode des moindres carrés. Pour cela on analyse les troncs (le bois est un tissu mort) de très vieux arbres Sequoia gigantea et Pinus aristata. Par un prélèvement effectué sur le tronc, on peut obtenir (cf. table 1.1) :

- son âge t en année, en comptant le nombre des anneaux de croissance,
- sa radioactivité A en mesurant le nombre de désintégration.

t	500	1000	2000	3000	4000	5000	6300
A	14.5	13.5	12.0	10.8	9.9	8.9	8.0

TABLE 1.1 – Données.

Notre but est donc de trouver les valeurs des paramètres A_0 et λ pour que la fonction $A(t)$ "colle" au mieux à nos données. Si on donne des valeurs à ces constantes, nous pouvons calculer les quantités appelées résidus (Fig. 1.5)

$$r_i(A_0, \lambda) = A_i - A(t_i) = A_i - A_0 e^{-\lambda t_i}.$$

2. Associer les noms de Fermat et Descartes est surprenant pour qui connaît les confrontations scientifiques virulentes qui les opposèrent. Les étudiants intéressés peuvent voir la vidéo ([4]) où se rendre au musée Pierre de Fermat de Beaumont de Lomagne, ville natale de P. de Fermat près de Toulouse.

Par suite nous pouvons calculer la quantité

$$f(A_0, \alpha) = \frac{1}{2} \sum_{i=1}^n (A_i - A_0 e^{-\lambda t_i})^2.$$

Cette quantité est la somme des carrés des longueurs des résidus.

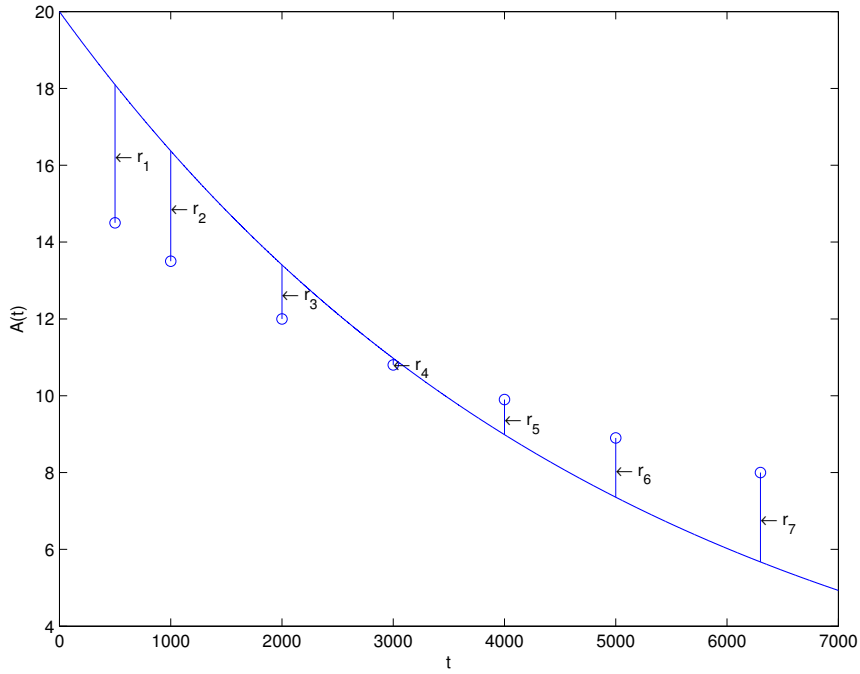


FIGURE 1.5 – Résidus $r(20, 0.0002)$ pour le problème de datation par le carbone 14.

Plus cette quantité sera faible, plus notre courbe sera proche de nos points expérimentaux. Estimer les paramètres A_0 et λ par les moindres carrés, c’est rechercher la valeur solution du problème d’optimisation suivant :

$$(P) \begin{cases} \text{Min} f(A_0, \lambda) = \frac{1}{2} \sum_{i=1}^n (A_i - A_0 e^{-\lambda t_i})^2 \\ (A_0, \lambda) \in \mathbf{R}^2. \end{cases}$$

Attention dans le problème (P) ci-dessus, les instants t_i et les valeurs A_i sont connus. Ce sont les valeurs des paramètres A_0 et λ que l’on cherche.

Remarque I.1.5

– Dans l’exemple précédent on peut aussi écrire : $f(\beta) = \frac{1}{2} \| r(\beta) \|^2$ où

$$\beta = (A_0, \lambda), \quad r(\beta) = \begin{pmatrix} r_1(\beta) \\ \vdots \\ r_n(\beta) \end{pmatrix}, \quad r_i(\beta) = A_i - A_0 e^{-\lambda t_i} \quad \text{et} \quad \|\cdot\| \quad \text{est la norme euclidienne.}$$

- Minimiser $f(\beta)$ est équivalent à minimiser $\alpha f(\beta)$ avec $\alpha > 0$. Le terme $\frac{1}{2}$ est mis ici afin de ne pas avoir le terme 2 lorsque l’on dérive la fonction $f(\beta)$.
- On peut aussi prendre comme critère :
 - $f(\beta) = \|r(\beta)\|_1 = \sum_{i=1}^n |r_i(\beta)|$;
 - $f(\beta) = \|r(\beta)\|_\infty = \text{Max}_{i=1, \dots, n} |r_i(\beta)|$.

Exemple I.1.6 (Modèle de Kaplan)

On désire étudier la diffusion d’une drogue dans un organe d’un corps donné. La drogue est injectée par intraveineuse dans le sang à l’instant $t_0 = 0$. On modélise le système par un modèle à compartiments (cf. la figure 1.6).

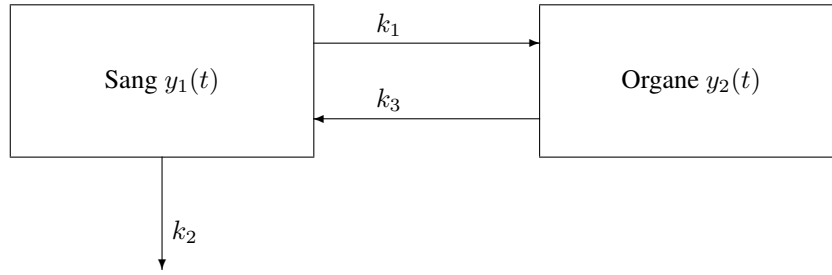


FIGURE 1.6 – Modèle par compartiments.

Les concentrations dans le sang, mesurées à différents instants, sont données à la table 1.2.

t_i	y_{i1}	t_i	y_{i1}
0.25	215.6	3.00	101.2
0.50	189.2	4.00	88.0
0.75	176.0	6.00	61.6
1.00	162.8	12.00	22.0
1.50	138.6	24.00	4.4
2.00	121.0	48.00	0.0

TABLE 1.2 – Données pour l'exemple de Kaplan.

Le système d'équations différentielles décrivant le modèle est alors

$$(EDO) \begin{cases} \frac{dy_1}{dt} = \dot{y}_1(t) = -(k_1 + k_2)y_1(t) + k_3y_2(t) \\ \frac{dy_2}{dt} = \dot{y}_2(t) = k_1y_1(t) - k_3y_2(t) \\ y_1(0) = c_0 \\ y_2(0) = 0. \end{cases}$$

On désire estimer les paramètres c_0, k_1, k_2 et k_3 par les moindres carrés. Posons $\beta = (c_0, k_1, k_2, k_3)$, alors pour toute valeur de β , on peut intégrer le système d'équations différentielles ordinaires à condition initiale (EDO). Notons $(y_1(t, \beta), y_2(t, \beta))$ cette solution. Par suite on peut calculer les n résidus

$$r_i(\beta) = y_{i1} - y_1(t_i, \beta).$$

Ces résidus sont visualisés sur la figure 1.7. Nous estimerons alors le paramètre β en résolvant le problème d'optimisation aux moindres carrés

$$(P) \begin{cases} \text{Min}f(\beta) = \frac{1}{2} \sum_{i=1}^n r_i^2(\beta) = \frac{1}{2} \|r(\beta)\|^2 \\ \beta \in \mathbf{R}^4. \end{cases}$$

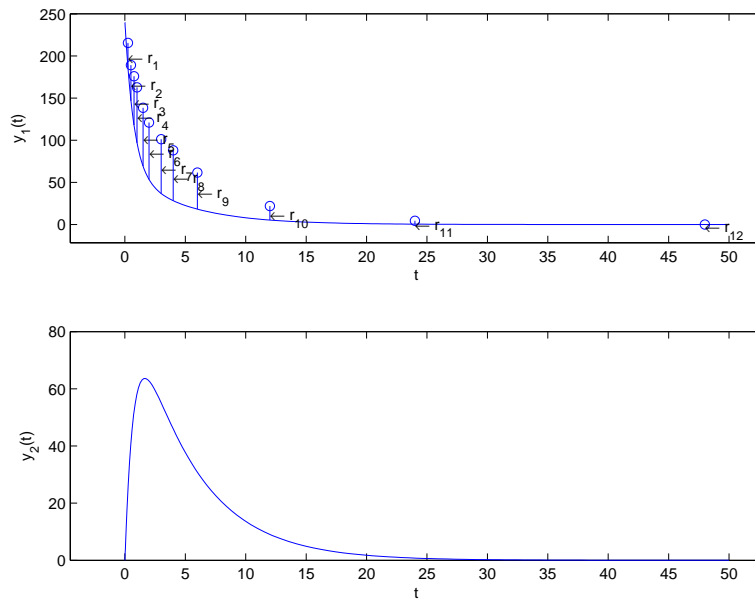


FIGURE 1.7 – Critère des moindres carrés pour le modèle de Kaplan.

Exemple I.1.7

On veut mesurer la liaison entre 2 gènes dominants, l'un contrôlant la couleur d'une fleur, rouge (R) est dominant sur blanc (b), et l'autre la taille, grand (G) est dominant sur petit (p). Dans la descendance F_2 , issu de deux populations homozygotes de phénotype $[RG]$ et $[bp]$, on a étudié $n = 3839$ plantes. On a obtenu les résultats de la table 1.3.

Phénotypes	$[RG]$	$[Rp]$	$[bG]$	$[bp]$
Effectifs observés	1997	906	904	32

TABLE 1.3 – Données de Sir R.A. Fisher.

Le problème est d'estimer, à partir de ces données le taux de recombinaison r . Ici la population F_1 est hétérozygote de génotype Rb, Gp . Nous avons donc les probabilités de la table 1.4 pour les différents gamètes possibles et les différents croisements possibles.

$\varphi : \sigma$	RG	bp	Rp	bG
	$\frac{1}{2}(1-r)$	$\frac{1}{2}(1-r)$	$\frac{1}{2}r$	$\frac{1}{2}r$
RG $\frac{1}{2}(1-r)$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r(1-r)$
bp $\frac{1}{2}(1-r)$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}(1-r)^2$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r(1-r)$
Rp $\frac{1}{2}r$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{4}r^2$
bG $\frac{1}{2}r$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r(1-r)$	$\frac{1}{4}r^2$	$\frac{1}{4}r^2$

TABLE 1.4 – Probabilités pour la descendance F_2 .

Par suite nous avons dans la population F_2 la loi suivante pour la variable aléatoire phénotype X

$$\begin{aligned}
 X : F_2 &\longrightarrow \{[RG], [Rp], [bG], [bp]\} \\
 1 \text{ plante} &\longmapsto \text{son phénotype,}
 \end{aligned}$$

$$\begin{aligned}
 P(X = [RG]) &= \frac{1}{4}(3 - 2r + r^2) = \frac{2 + \theta}{4} \\
 P(X = [Rp]) &= \frac{1}{4}(2r - r^2) = \frac{1 - \theta}{4} \\
 P(X = [bG]) &= \frac{1}{4}(2r - r^2) = \frac{1 - \theta}{4} \\
 P(X = [bp]) &= \frac{1}{4}(1 - r)^2 = \frac{\theta}{4}
 \end{aligned}$$

où $\theta = (1 - r)^2 \in [\frac{1}{4}; 1]$.

Définissons maintenant le vecteur aléatoire de dimension 4

$$\begin{aligned}
 (A, B, C, D) : F_2^n &\longrightarrow \mathbf{R}^4 \\
 n \text{ plantes} &\longmapsto \begin{aligned} &(\text{nb de plantes de phénotypes } [RG], \\ &\text{nb de plantes de phénotypes } [Rp], \\ &\text{nb de plantes de phénotypes } [bG], \\ &\text{nb de plantes de phénotypes } [bp]). \end{aligned}
 \end{aligned}$$

On suppose la population F_2 de taille infinie, donc la loi de ce vecteur aléatoire est une loi multinomiale

$$\begin{aligned}
 L(a, b, c, d; \theta) &= P((A, B, C, D) = (a, b, c, d)) \\
 &= \frac{n!}{a!b!c!d!} P(X = [RG])^a P(X = [Rp])^b P(X = [bG])^c P(X = [bp])^d \\
 &= \frac{n!}{a!b!c!d!} \left(\frac{2 + \theta}{4}\right)^a \left(\frac{1 - \theta}{4}\right)^{b+c} \left(\frac{\theta}{4}\right)^d.
 \end{aligned}$$

L s'appelle la vraisemblance³. L'estimation de θ par le maximum de vraisemblance consiste alors à rechercher la valeur de θ solution du problème de maximisation suivant

$$(P) \begin{cases} \text{Max } L(1997, 906, 904, 32; \theta) \\ \theta \in [\frac{1}{4}; 1]. \end{cases}$$

Exemple I.1.8

Un fermier désire déterminer les quantités de lisier de porc et d'engrais composé à étendre sur 20 ha de prairie de façon à optimiser le coût total de la fertilisation. Le coût et la composition du lisier et de l'engrais sont donnés à la table 1.5.

	coût (par tonne)	composition chimique ($kg t^{-1}$)		
		azote	phosphate	potasse
lisier	25 francs	6	1.5	4
engrais	1300 francs	250	100	100

TABLE 1.5 – Coûts et compositions des engrais.

Le fermier veut appliquer au moins 75 kg ha^{-1} d'azote, 25 kg ha^{-1} de phosphate et 35 kg ha^{-1} de potasse. Il ne peut appliquer le lisier qu'à un taux maximum de 8 t/heure et l'engrais qu'à un taux maximum de 0.4 t/heure. Il ne peut de plus consacrer pour ce travail qu'un maximum de 25 heures.

Appelons x_1 (respectivement x_2) la quantité en tonnes de lisier (respectivement d'engrais) étendu. Le problème est alors d'obtenir un coût minimum, c'est-à-dire que l'on cherche à minimiser $25x_1 + 1300x_2$. Mais nous avons aussi les contraintes suivantes :

$$\begin{array}{ll}
 x_1 \geq 0 & \text{non négativité de } x_1 \\
 x_2 \geq 0 & \text{non négativité de } x_2 \\
 6x_1 + 250x_2 \geq 75 \times 20 = 1500 & \text{contrainte sur l'azote} \\
 1.5x_1 + 100x_2 \geq 500 & \text{contrainte sur le phosphate} \\
 4x_1 + 100x_2 \geq 700 & \text{contrainte sur la potasse} \\
 (1/8)x_1 + (1/0.4)x_2 \leq 25 & \text{contrainte de temps.}
 \end{array}$$

3. likelihood en anglais.

En résumé nous avons le problème suivant à résoudre :

$$(P) \begin{cases} \text{Min } f(x) = 25x_1 + 1300x_2 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ 6x_1 + 250x_2 \geq 75 \times 20 = 1500 \\ 1.5x_1 + 100x_2 \geq 500 \\ 4x_1 + 100x_2 \geq 700 \\ (1/8)x_1 + (1/0.4)x_2 \leq 25. \end{cases}$$

Exemple I.1.9 (Gestion de portefeuille[2])

La théorie de la sélection optimale de portefeuille a été développée par Harry Markowitz, prix Nobel d'économie en 1990, dans les années 1950. On considère un investisseur qui a une quantité fixée d'argent à sa disposition pour investir dans n actifs différentes (actions, stocks, ...) dont le retour est aléatoire. Pour chaque actif, on suppose connu son espérance mathématique μ_i , sa variance σ_i^2 . On suppose aussi connu pour deux actifs i et j leur coefficient de corrélation linéaire ρ_{ij} . On note x_i la proportion investie dans l'actif i . On peut donc calculer l'espérance mathématique et la variance résultant d'un portefeuille $x = (x_1, \dots, x_n)$

$$E(x) = \mu^T x$$

$$Var(x) = x^T Q x,$$

où Q est la matrice des covariances, $q_{ij} = \rho_{ij}\sigma_i\sigma_j$. Le portefeuille sera dit efficace si, pour une variance fixée, il a la plus grande espérance mathématique. C'est à dire s'il est solution du problème d'optimisation

$$(P) \begin{cases} \text{Max } E(x) \\ Var(x) = V \\ \sum_{i=1}^n x_i = 1 \\ x \geq 0. \end{cases}$$

On peut aussi s'intéresser au problème (MV0)⁴ de Markowitz.

$$(MVO) \begin{cases} \text{Min } Var(x) \\ E(x) \geq R \\ \sum_{i=1}^n x_i = 1 \\ x \geq 0. \end{cases}$$

Ces deux formulations sont en fait équivalentes.

I.2 Problèmes en nombres entiers

Exemple I.2.1 (Problème du sac à dos de Knapsack)

Un alpiniste veut mettre dans son sac à dos un maximum de 16 kg de ravitaillement. Il peut choisir un certain nombre d'unités de trois produits différents. Le poids unitaire en kilogrammes et la valeur énergétique unitaire de ces produits sont connus et donnés dans la table (1.6).

Produits	I	II	III
Poids	2	5	7
Valeurs	4	10	15

TABLE 1.6 – Poids unitaires et valeurs énergétiques unitaires.

Le problème pour l'alpiniste est de savoir ce qu'il doit emporter pour avoir une valeur totale en calories maximale sans dépasser les 16 kg.

Si nous notons x_1, x_2 et x_3 les nombres d'unités à emporter des articles I, II et III, le problème s'écrit

$$(P) \begin{cases} \text{Max } 4x_1 + 10x_2 + 15x_3 \\ 2x_1 + 5x_2 + 7x_3 \leq 16 \\ (x_1, x_2, x_3) \in \mathbb{N}^3. \end{cases}$$

4. mean-variance optimization.

Exemple I.2.2

([1]) Dans un service hospitalier, les malades i attendent d'être opérés. Le malade i a besoin d'une durée d'opération D_i . D'autre part, compte tenu des disponibilités des chirurgiens, la somme des durées des opérations possibles chaque jours j de la période étudiée est connue et égale à T_j . On veut minimiser la somme des pénalités d'attente pour les différents malades.

On note :

- $x_{ij} = 1$ si le malade i est opéré le jour j ;
- $x_{ij} = 0$ si le malade i n'est pas opéré le jour j ;
- c_{ij} la pénalité du malade i s'il est opéré le jour j . c_{ij} est une fonction croissante de j .

Le problème s'écrit alors :

$$(P) \begin{cases} \text{Min } f(x) = \sum_i \sum_j c_{ij} x_{ij} \\ \sum_i D_i x_{ij} \leq T_j \quad \forall j \text{ limitation des possibilités opératoire du jour } j \\ \sum_j x_{ij} = 1 \quad \forall i \text{ Le malade } i \text{ est opéré une fois et une seule} \\ x_{ij} = 0 \text{ ou } 1 \text{ l'opération est effectuée en une fois.} \end{cases}$$

Exemple I.2.3 (Alignement de séquences)

Soit 2 séquences *CTGTATC* et *CTATAATCCC*. On désire trouver le "meilleur" alignement possible. À chaque alignement, est associé un score (simple ici) suivant : pour chaque position on associe 0 si les 2 bases sont identiques, +1 si les deux bases sont différentes et +3 s'il y a un "trou". On effectue ensuite la somme. La figure (1.8) donne un exemple de la fonction score S .

C	T	A	T	-	A	A	-	T	C	C	C	
-	-	C	T	G	T	A	T	C	-	-	-	
3	3	1	0	3	1	0	3	1	3	3	3	= 24

FIGURE 1.8 – Exemple de calcul d'un score.

Le problème est alors de résoudre le problème d'optimisation suivant

$$(P) \begin{cases} \text{Min } S(\text{alignement}) \\ \text{pour tous les alignements possibles.} \end{cases}$$

Remarque I.2.4

la difficulté est ici de construire l'ensemble de tous les alignements possibles. Ceci se fait de la façon suivante. Supposons que l'on soit à la position i , alors pour aller à la position $i + 1$, nous avons trois possibilités :

- avancer d'un nucléotide pour les 2 séquences ;
- avancer d'un nucléotide pour la séquence S_1 et mettre un "trou" pour la séquence S_2 ;
- avancer d'un nucléotide pour la séquence S_2 et mettre un "trou" pour la séquence S_1 .

Nous pouvons ainsi construire un arbre permettant d'avoir tous les alignements possibles.

I.3 Problème en dimension infinie

Exemple I.3.1 (Problème de la brachistochrone)

Le problème de la brachistochrone⁵ fut posé par Jean Bernoulli⁶ (cf. la figure 1.9) en 1696 et est considéré comme le problème fondateur du calcul des variations.

Ce problème consiste en la recherche dans un plan vertical du chemin reliant 2 points P_0 et P_f de ce plan, suivant lequel un corps M entraîné par son propre poids effectuera le trajet de P_0 à P_f en un temps minimum. On suppose qu'il n'y a pas de frottement. Introduisons dans le plan un système de coordonnées (x, y) pour lequel P_0 ait comme coordonnées $(0, 0)$ et $P_f (x_f, y_f), x_f > 0$ et $y_f < 0$. Supposons que $y(\cdot)$ est la fonction qui donne l'équation de la courbe joignant les points P_0 et P_f . Les lois de la mécanique nous disent que le module de la vitesse v en $(x, y(x))$ ne dépend pas de la forme de la courbe $y(\cdot)$ sur $[0, x]$, mais seulement de l'ordonnée $y(x)$, et que cette vitesse est égale à $\sqrt{2g(-y(x))}$, où g est l'accélération gravitationnelle. Si on note s l'abscisse curviligne, le temps pour parcourir l'élément $ds = \sqrt{dx^2 + dy^2}$ est alors $ds/\sqrt{2g(-y(x))}$. Posons

$$T : C^1([0, x_f], \mathbf{R}) \longrightarrow \mathbf{R}$$

$$y(\cdot) \longmapsto \int_0^{x_f} \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2g(-y(x))}} dx.$$

5. Le mot brachistochrone vient du grec *brakhisto* qui signifie le plus court et de *chronos* qui signifie temps.

6. http://fr.wikipedia.org/wiki/Jean_Bernoulli

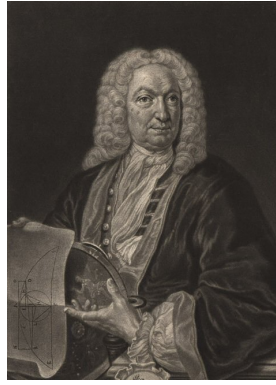


FIGURE 1.9 – Jean Bernoulli 27 juillet 1667 – 1er janvier 1748.

Le problème s'écrit alors

$$(P) \begin{cases} \text{Min } T(y(\cdot)) \\ y(0) = 0 \\ y(x_f) = y_f. \end{cases}$$

La solution de ce problème est visualisée à la figure 1.10.

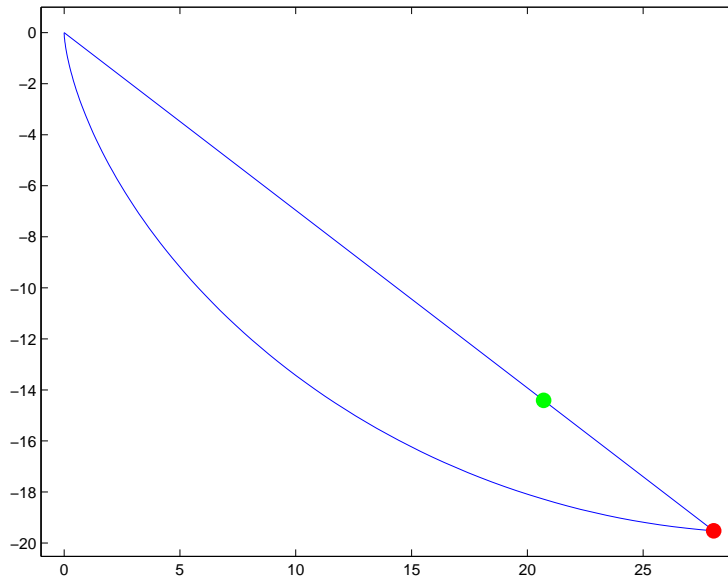


FIGURE 1.10 – Brachistochrone.

Exemple I.3.2 (Transfert orbital)

On désire transférer un satellite S d'une orbite initiale (celle où la fusée Ariane l'a "posé") vers l'orbite géostationnaire (cf. la figure 1.11), le moteur du satellite étant un moteur à poussée faible.

Le satellite est considéré comme un point matériel et on note $r(t) \in \mathbf{R}^3$ la position, $v(t) \in \mathbf{R}^3$ la vitesse, $m(t)$ la masse du satellite et $T(t) \in \mathbf{R}^3$ la poussée du moteur. L'équation du mouvement, provenant des équations de Newton est alors

$$\begin{aligned} \dot{r}(t) &= v(t) \\ \dot{v}(t) &= -\frac{\mu r(t)}{\|r(t)\|^3} + \frac{T(t)}{m(t)} \\ \dot{m}(t) &= -\beta \|T(t)\|, \end{aligned}$$

où μ est la constante gravitationnelle de la Terre et $\beta = 1/g_0 I_{sp}$ est une constante positive (g_0 est l'accélération gravitationnelle terrestre à la surface de la Terre et I_{sp} est une constante caractéristique du moteur appelé impulsion spécifique). À l'instant initial les position, vitesse et masse du satellite sont connues et à l'instant terminal t_f le satellite doit être sur l'orbite géostationnaire à une position et vitesse $(r(t_f), v(t_f)) = (r_f, v_f)$ fixés. Bien évidemment la poussée du moteur est bornée

$$\|T(t)\| \leq T_{max}.$$

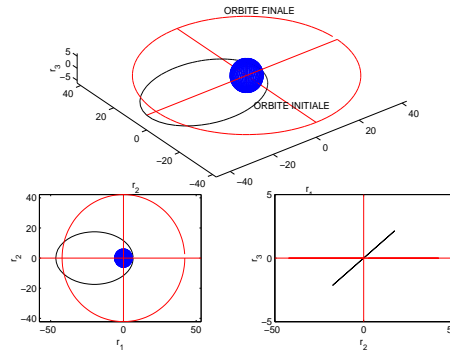


FIGURE 1.11 – Transfert orbital.

L'objectif est alors de trouver une loi de commande du moteur qui réalise le transfert et qui minimise le temps de transfert. On peut aussi s'intéresser à la maximisation de la masse finale (dans ce cas le temps de transfert doit-être fixé). Si on normalise le contrôle $u(t) = T(t)/T_{max}$ alors le problème s'écrit pour la maximisation de la masse finale (ou la minimisation de la consommation)

$$(P) \begin{cases} \text{Min } J(u) = \int_0^{t_f} \|u(t)\| dt \\ \dot{r}(t) = v(t) \quad \text{pp. in } [0, t_f] \quad t_f \text{ fixed} \\ \dot{v}(t) = -\mu r(t)/\|r(t)\|^3 + \frac{T_{max}}{m(t)} u(t) \\ \dot{m}(t) = -\beta T_{max} \|u(t)\| \\ (r(t), v(t), m(t)) \in A \\ \|u(t)\| \leq 1 \\ r(0), v(0), m(0) \text{ fixé} \\ r(t_f), v(t_f) \text{ fixé,} \end{cases}$$

ce problème est un problème de contrôle optimal et l'inconnue est la commande, donc une fonction u , ici de $[0, t_f]$ à valeurs dans \mathbf{R}^3 .

Remarque I.3.3

Ces problèmes d'optimisation en dimension infinie seront traités dans le cours de contrôle optimal en deuxième année majeure mathématiques appliquées.

II Problème d'optimisation

II.1 Définitions

Définition II.1.1 (Ensemble convexe) Un sous ensemble C d'un espace vectoriel est dit convexe si pour tout $(x, y) \in C$ le segment $[x, y] = \{\alpha x + (1 - \alpha)y, \alpha \in [0, 1]\}$ est inclus dans C .

Définition II.1.2 (Fonction convexe) Une fonction f de $C \subset E$ à valeurs dans \mathbf{R} , E espace vectoriel, est convexe si et seulement si elle vérifie :

- (i) C est convexe ;
- (ii)

$$\forall (x, y) \in C^2, \quad \forall \alpha \in [0, 1], \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

Dans le cas $n = 1$, ceci signifie que le graphe de la fonction f est toujours sous la corde, cf. la FIGURE (1.12).

Définition II.1.3 (Problème d'optimisation sans contraintes) On appelle problème d'optimisation sans contraintes en dimension finie tout problème (P) consistant en la recherche d'un minimum d'une fonctionnelle f définie sur \mathbf{R}^n . On notera ce problème sous la forme suivante :

$$(P) \begin{cases} \text{Min } f(x) \\ x \in \mathbf{R}^n \end{cases}$$

où $f : \mathbf{R}^n \rightarrow \mathbf{R}$ sera donnée.

Remarque II.1.4

Résoudre le problème (P) revient à rechercher le point x^* de \mathbf{R}^n tel que $f(x^*) \leq f(x) \forall x \in \mathbf{R}^n$.

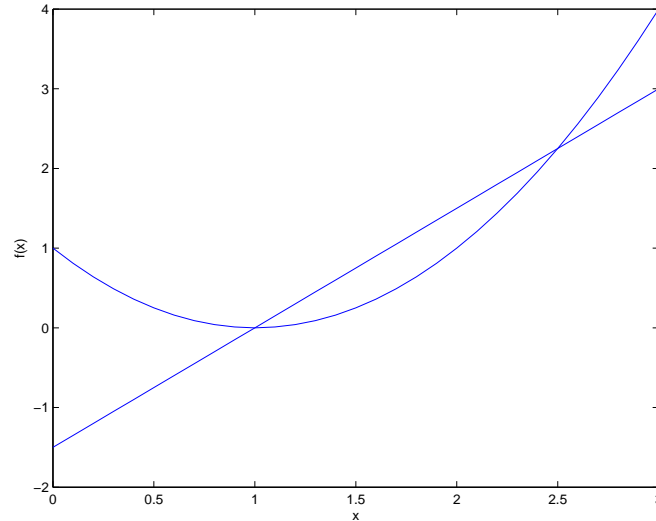


FIGURE 1.12 – Fonction convexe.

Remarque II.1.5

Un problème de maximisation se ramène très facilement à un problème de minimisation :

$$\text{Max}f(x) \iff \text{Min}(-f(x))$$

Définition II.1.6 (Problème d’optimisation avec contraintes) On appelle problème d’optimisation avec contraintes tout problème (P) consistant en la recherche d’un minimum sur un ensemble C inclus dans \mathbf{R}^n d’une fonctionnelle f définie sur \mathbf{R}^n . On notera ce problème sous la forme suivante :

$$(P) \begin{cases} \text{Min}f(x) \\ x \in C \subset \mathbf{R}^n \end{cases}$$

où $f : \mathbf{R}^n \rightarrow \mathbf{R}$ sera donnée.

Remarque II.1.7

Dans la pratique C sera défini de la façon suivante :

$$C = \{x \in \mathbf{R}^n / g_i(x) \leq 0 \quad i = 1, \dots, m \text{ et } h_l(x) = 0 \quad l = 1, \dots, p\} \tag{1.1}$$

et nous écrirons (P) sous la forme

$$(P) \begin{cases} \text{Min}f(x) \\ g_i(x) \leq 0 \quad i = 1, \dots, m \\ h_l(x) = 0 \quad l = 1, \dots, p \end{cases}$$

Définition II.1.8 (Optimisation non différentiable) On appelle problème d’optimisation non différentiable un problème d’optimisation où les fonctions qui interviennent ne sont pas dérivables.

Remarque II.1.9

On ne traitera dans ce cours que des problèmes d’optimisation différentiables.

Définition II.1.10 (Problème d’optimisation convexe) Un problème d’optimisation est dit convexe si et seulement si la fonction f est convexe et l’ensemble des contrainte C est convexe.

Remarque II.1.11

Si C est définie par (1.1) et si les fonctions g_i sont convexes et les fonctions h_l sont affines, alors C est convexe. Attention, la réciproque est fausse.

Définition II.1.12 (Problème aux moindres carrés) On appelle problème aux moindres carrés un problème d’optimisation sans contraintes où la fonctionnelle f est de la forme suivante :

$$f(\beta) = \frac{1}{2} \|r(\beta)\|^2 = \frac{1}{2} (r(\beta)|r(\beta)) = \frac{1}{2} \sum_{i=1}^n r_i^2(\beta)$$

Le problème est dit aux moindres carrés linéaires si la fonction r est affine :

$$\begin{aligned} r : \mathbf{R}^p &\longrightarrow \mathbf{R}^n \\ \beta &\longmapsto y - X\beta \end{aligned}$$

où X matrice de type (n, p) et y un élément de \mathbf{R}^n .

Exemple II.1.13

L'exemple (??) est un problème aux moindres carrés non linéaire.

Définition II.1.14 (Problème linéaire) Un problème d'optimisation est dit linéaire si et seulement si les fonctions $f, g_i,$ et h_l sont affines.

Exemple II.1.15

L'exemple (I.1.8) est un problème linéaire.

Définition II.1.16 (Optimum global, optimum local) Soit (P) un problème d'optimisation sans contraintes.

- (i) x^* est un minimum global $\iff x^*$ est la solution de (P)
- (ii) x^* est un minimum local faible \iff il existe $\varepsilon > 0$ tel que x^* est la solution de (P') où

$$(P') \begin{cases} \text{Min } f(x) \\ \|x - x^*\| < \varepsilon \end{cases}$$

- (iii) x^* est un minimum local fort si

$$\forall \mathbf{x} \in B(\mathbf{x}^*, \varepsilon) = \{\mathbf{x} \in \mathbf{R}^n / \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon\}, \mathbf{x} \neq \mathbf{x}^*, f(\mathbf{x}^*) < f(\mathbf{x}).$$

Dans le cas où $n = 1, \|x - x^*\|$ devient $|x - x^*|$ et par suite nous avons

$$\|x - x^*\| < \varepsilon \iff |x - x^*| < \varepsilon \iff x^* - \varepsilon < x < x^* + \varepsilon,$$

(cf. la figure 1.13).

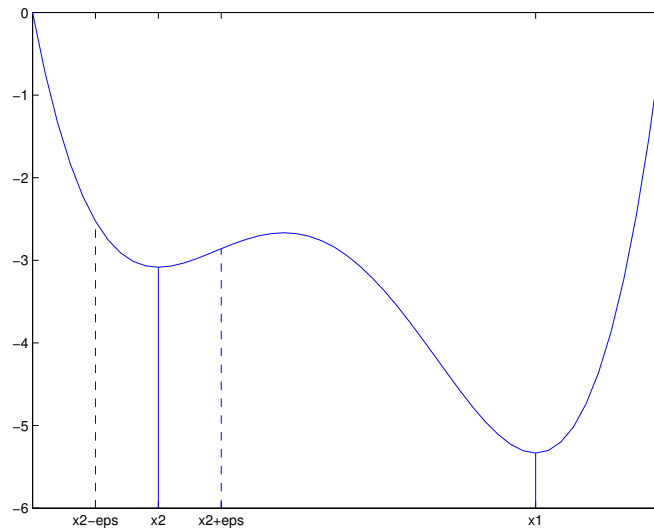


FIGURE 1.13 – x^2 est un minimum local fort, x^1 est un minimum global

Remarque II.1.17

On dit que x^* est un minimum alors que c'est $f(x^*)$ qui est un minimum. Il s'agit d'un abus de langage que nous emploierons systématiquement.

Remarque II.1.18

On appelle optimisation globale la recherche d'un optimum global. Un algorithme globalement convergent est lui un algorithme qui converge vers un minimum local quel que soit le point de départ.

II.2 Classification

Considérons le problème d'optimisation suivant :

$$(P) \begin{cases} \text{Min } f(x) \\ x \in C \subset E \end{cases}$$

Suivant la nature des ensembles C et E et de la fonction f nous avons différents types de problème d'optimisation. La figure 1.14 donne une classification des problèmes d'optimisation (nous n'étudierons dans ce cours que les parties en bleu).

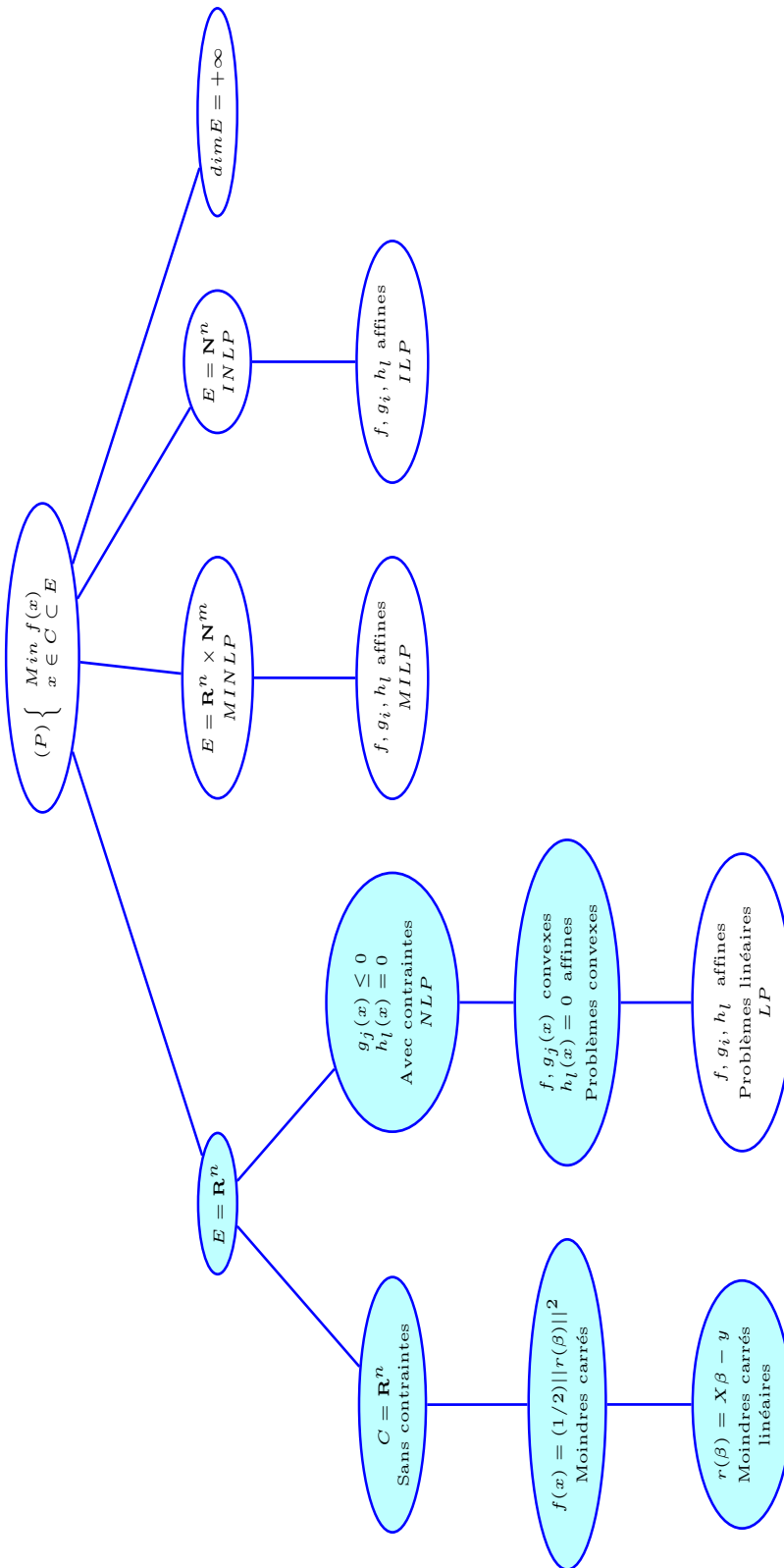


FIGURE 1.14 – Classification des problèmes d'optimisation, on a en bleu ce qui sera vu en cours.

III Exercices

Exercice III.1

Les données ci-dessous représentent le rendement de la repousse d'une prairie en fonction du temps écoulé depuis la première fauche. Les unités ne sont pas précisées.

temps	9	14	21	28	42	57	63	70	79
rendements	8.93	10.80	18.59	22.33	39.35	56.11	61.73	64.62	67.08

On choisit d'ajuster ces données à un modèle de croissance de Weibull de la forme :

$$y(t) = x_1 - x_2 \exp(-x_3 t^{x_4})$$

1 On veut estimer les paramètres par les moindres carrés. Écrire le problème sous la forme :

$$\begin{cases} \text{Min} & f(\beta) = \frac{1}{2} \|r(\beta)\|^2 \\ \beta & \in \mathbf{R}^p \end{cases}$$

Que représente β ? On donnera l'expression de $r(\beta)$.

Exercice III.2 (Régression linéaire simple)

Soit n points expérimentaux $M_i = (x_i, y_i)$ pour $i = 1, \dots, n$. On considère le modèle suivant : $y(x) = \beta_0 + \beta_1 x$. On estime les paramètres par les moindres carrés.

1 On veut estimer les paramètres par les moindres carrés. Écrire le problème sous la forme :

$$\begin{cases} \text{Min} & f(\beta) = \frac{1}{2} \|r(\beta)\|^2 = \frac{1}{2} \|y - X\beta\|^2 \\ \beta & \in \mathbf{R}^p \end{cases}$$

On donnera les valeurs de X et de y et à quoi correspond β .

2 Pour β fixé visualiser sur le graphique des données les résidus $r_i(\beta)$.

3 On souhaite maintenant trouver la meilleure droite au sens des moindres carrés qui passe par l'origine. Écrire le problème d'optimisation.

Exercice III.3 (Datation par le carbone 14)

Le carbone radioactif ^{14}C est produit dans l'atmosphère par l'effet des rayons cosmiques sur l'azote atmosphérique. Il est oxydé en $^{14}\text{CO}_2$ et absorbé sous cette forme par les organismes vivants qui, par suite, contiennent un certain pourcentage de carbone radioactif relativement aux carbones ^{12}C et ^{13}C qui sont stables. On suppose que la production de carbone ^{14}C atmosphérique est demeurée constante durant les derniers millénaires. On suppose d'autre part que, lorsqu'un organisme meurt, ses échanges avec l'atmosphère cessent et que la radioactivité due au carbone ^{14}C décroît suivant la loi exponentielle suivante :

$$A(t) = A_0 e^{-\lambda t}$$

où λ est une constante positive, t représente le temps en année et $A(t)$ est la radioactivité exprimée en nombre de désintégrations par minute et par gramme de carbone. On désire estimer les paramètres A_0 et λ par la méthode des moindres carrés. Pour cela on analyse les troncs (le bois est un tissu mort) de très vieux arbres Sequoia gigantea et Pinus aristata. Par un prélèvement effectué sur le tronc, on peut obtenir :

- son âge t en année, en comptant le nombre des anneaux de croissance,
- sa radioactivité A en mesurant le nombre de désintégration.

t	500	1000	2000	3000	4000	5000	6300
A	14.5	13.5	12.0	10.8	9.9	8.9	8.0

1 Sur le graphique des données visualiser les résidus pour $A_0 = 20$ et $\lambda = 0.0002$;

2 Écrire le problème sous la forme :

$$\begin{cases} \text{Min} & f(\beta) = \frac{1}{2} \|r(\beta)\|^2 \\ \beta & \in \mathbf{R}^p. \end{cases}$$

On donnera l'expression de $r(\beta)$.

3 On pose maintenant $y(t) = \ln A(t)$, $\theta_0 = \ln A_0$, $\theta_1 = -\lambda$ et $y_i = \ln(A_i)$. Écrire le problème d'estimation par les moindres carrés de θ_0 et θ_1 .

Exercice III.4

On donne ci-dessous la population des États Unis pour les années 1900 à 2000⁷.

années	1900	1910	1920	1930	1940	1950	1960	1970	1980
pop.	75.995	91.972	105.711	123.203	131.669	150.697	179.323	203.212	226.505
années	1990	2000							
pop.	249.633	281.422							

TABLE 1.7 – Données provenant de "U.S. Census"

On désire ajuster ces données sur le modèle

$$y(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$

On utilise la méthode des moindres carrés.

- 1 Visualiser les résidus sur le graphe des données.
- 2 Écrire le problème sous la forme :

$$\begin{cases} \text{Min } f(\beta) = \frac{1}{2} \|r(\beta)\|^2 = \frac{1}{2} \|y - X\beta\|^2 \\ \beta \in \mathbf{R}^p \end{cases}$$

On donnera les valeurs de X et de y et à quoi correspond β .

- 3 Pour β fixé visualiser sur le graphique des données les résidus $r_i(\beta)$.

Exercice III.5 (Maintenance d'un réseau de distribution)

8

Un ingénieur responsable de la maintenance d'un réseau de distributeurs de boissons aimerait prédire le temps nécessaire pour l'approvisionnement. Le service d'approvisionnement comprend le remplissage des machines et leurs réglages éventuels. Deux variables influencent ce temps : le nombre de caisses à charger et la distance parcourue par l'employé pour approvisionner l'ensemble des machines. Le responsable dispose de 25 observations, qui sont résumées dans le tableau suivant :

Obs.	Temps	Nb caisses	Dist.	Obs.	Temps	Nb. caisses	Dist.
1	16.68	7	560	13	13.50	4	255
2	11.50	3	220	14	19.75	6	462
3	12.03	3	340	15	24.00	9	448
4	14.88	4	80	16	29.00	10	776
5	13.75	6	150	17	13.35	6	200
6	18.11	7	330	18	19.00	7	132
7	8.00	2	110	19	9.50	3	36
8	17.83	7	210	20	35.10	17	770
9	79.24	30	1460	21	17.90	10	140
10	21.50	5	605	22	52.32	26	810
11	40.33	16	688	23	18.75	9	450
12	21.00	10	215	24	19.83	8	685
				25	10.75	4	150

On désire ajuster à cet ensemble de données un modèle de régression multiple

$$y(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

où y est le temps requis, x_1 est le nombre de caisses utilisées et x_2 est la distance parcourue.

7. Exemple provenant du cours de Cleve Moler page 4 chap. 5 "Numerical computing with Matlab"

8. Exemple provenant du livre d'A. Antoniadis, J. Berruyer et R. Carmona, régression non linéaire et applications", éditions Economica, p. 45

1 Écrire le problème sous la forme :

$$\begin{cases} \text{Min} & f(\beta) = \frac{1}{2} \|r(\beta)\|^2 = \frac{1}{2} \|y - X\beta\|^2 \\ \beta \in & \mathbf{R}^p \end{cases}$$

On donnera les valeurs de X et de y et à quoi correspond β .

2 Pour β fixé visualiser sur le graphique des données les résidus $r_i(\beta)$.

Exercice III.6 (Courbe étalon ⁹⁾

La première étape d'un dosage radioimmunologique consiste à établir une courbe étalon. Un dosage repose sur l'hypothèse qu'une hormone et son isotope marqué se comportent de façon équivalente vis-à-vis de leur anticorps spécifique : lorsque l'on met en présence une quantité déterminée d'anticorps, une quantité déterminée d'hormone radioactive et une quantité variable d'hormone froide, la dose de complexe anticorps-hormone marquée en fin de réaction est d'autant plus faible que la quantité d'hormone froide est importante. Néanmoins, la relation qui existe entre la dose d'hormone froide mise en réaction et la radioactivité de complexe extrait n'est pas stable et doit être appréciée dans chaque situation expérimentale. C'est l'objet de l'établissement de la courbe étalon, à partir d'une gamme de dilutions connues d'une quantité déterminée de l'hormone à doser. La table 1.8 donne les données recueillies pour une telle courbe dans le cas d'un dosage du cortisol : on a mesuré la radioactivité du complexe (en coups par minutes ou cpm). On considère le modèle suivant :

$$y(x) = \beta_2 + \frac{\beta_1 - \beta_2}{(1 + \exp(\beta_3 + \beta_4 x))^{\beta_5}} \tag{1.2}$$

Dose en ng/.1 ml	Réponse en c.p.m.			
0	2868	2785	2849	2805
0	2779	2588	2701	2752
0.02	2615	2651	2506	2498
0.04	2474	2573	2378	2494
0.06	2152	2307	2101	2216
0.08	2114	2052	2016	2030
0.1	1862	1935	1800	1871
0.2	1364	1412	1377	1304
0.4	910	919	855	875
0.6	702	701	689	696
0.8	586	596	561	562
1	501	495	478	493
1.5	392	358	399	394
2	330	351	343	333
4	250	261	244	242
100	131	135	134	133

TABLE 1.8 – Données pour un dosage de Cortisol

On désire estimer les paramètres par les moindres carrés (attention, il y a pour chaque dose 4 observations de y). On notera $(x_i)_{i=1,\dots,16}$ (respectivement $(y_{i,j})_{i=1,\dots,16;j=1,\dots,4}$) les éléments de la première colonne (respectivement des 4 dernières colonnes) de la table 1.8 et $r_{i,j}(\beta)$ le résidu liés au point $(x_i, y_{i,j})$.

1 Écrire le résidu lié au point $(0.04, 2378)$.

2 (i) Quelle est la dimension du vecteur des paramètres β .

(ii) Quel est le nombre de points n ?

3 Écrire le problème d'optimisation des paramètres par les moindres carrés.

4 Pour β fixé visualiser sur le graphique des données les résidus $r_{i,j}(\beta)$.

Exercice III.7 (Géoréférence d'une image satellite ¹⁰⁾

On dispose d'une image satellite que l'on désire recaler par rapport à une carte géographique que l'on a à notre disposition. Pour cela on définit n points, appelés points d'amer, que l'on peut parfaitement faire correspondre sur la carte et sur l'image satellite. On prend par exemple un croisement de route, un point particulier sur une rivière, ... Concrètement on a donc à notre disposition n coordonnées (x_i, y_i) des n points d'amer sur la carte et n coordonnées (x'_i, y'_i) de ces mêmes points sur l'image satellite. On choisit d'exprimer ces coordonnées :

- en pixels pour les (x'_i, y'_i) (coordonnées (0,0) pour le coin inférieur gauche) ;
- en mètres relativement à un référentielle géodésique particulier pour les (x_i, y_i) , via une carte IGN par exemple.

On a par exemple les données suivantes :

Numéros	x_i	y_i	x'_i	y'_i
1	252	2661	458805	1831634
2	235	2603	458157	1830577
⋮	⋮	⋮	⋮	⋮
23	1021	2254	471301	1819574

En pratique l'image satellite est déformée par rapport à la réalité. Cette déformation a plusieurs origines : satellite non verticale par rapport à la prise de vue, présence de nuages dans l'atmosphère, ... En conséquence on suppose que l'on peut écrire :

$$\begin{cases} x = \gamma_0 + \gamma_1 x' + \gamma_2 y' + \gamma_3 x'^2 + \gamma_4 x' y' + \gamma_5 y'^2 \\ y = \delta_0 + \delta_1 x' + \delta_2 y' + \delta_3 x'^2 + \delta_4 x' y' + \delta_5 y'^2 \end{cases}$$

On désire estimer les paramètres par les moindres carrées

- 1 Pour l'estimation des paramètres $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_5)$ quelles sont les données ?
- 2 Écrire le problème d'estimation par les moindres carrés linéaires de γ .
- 3 Mêmes questions pour δ .

Exercice III.8

On désire construire un réservoir de forme cylindrique de volume maximum dont la surface latérale est inférieure à S_{lat} et la surface totale est inférieure à S_{tot} .

- 1 Formaliser le problème.

Exercice III.9

Soit $B(a, \delta)$ la boule de \mathbf{R}^2 de centre a et de rayon $\delta > 0$ fixée. Soient p frères ennemis. On désire enfermer ces frères ennemis dans $B(a, \delta)$ en maximisant la distance minimale ξ qui les sépare deux à deux.

- 1 Formaliser le problème.

Exercice III.10 (Réseaux de neurones)

On s'intéresse ici à la modélisation via les réseaux de neurones.

Un neurone formel est une fonction paramétrée par $n + 1$ paramètres w_1, \dots, w_n, θ

$$g : \mathbf{R}^n \longrightarrow \mathbf{R}$$

$$x \longmapsto g(x, w, \theta) = \sigma\left(\sum_{i=1}^n w_i x_i + \theta\right)$$

où σ est une fonction donnée qui s'appelle une fonction d'activation. Chaque paramètre w_i s'appelle le poids synaptique associé au signal d'entrée x_i .

On prendra dans la suite, sauf mention contraire, comme fonction σ la fonction sigmoïde (cf. la figure 1.15)

$$\sigma : \mathbf{R} \longrightarrow \mathbf{R}$$

$$x \longmapsto \frac{1}{1 + e^{-x}}$$

On a à notre disposition K points $x^k \in \mathbf{R}^n$ et $y^k \in \mathbf{R}$, on appelle apprentissage du neurone l'estimation par les moindres carrés des paramètres du neurone.

- 1 Écrire le problème au moindres carrés qui définit l'apprentissage. On donnera en particulier la fonction résidus r en précisant clairement l'espace de départ et l'espace d'arrivée.
- 2 Ce problème est-il un problème aux moindres carrés linéaires ? Si oui, on donnera la matrice X .
- 3 Si on prend comme fonction d'activation σ l'identité le problème au moindres carrés devient-il linéaire ? Si oui, on donnera la matrice X .

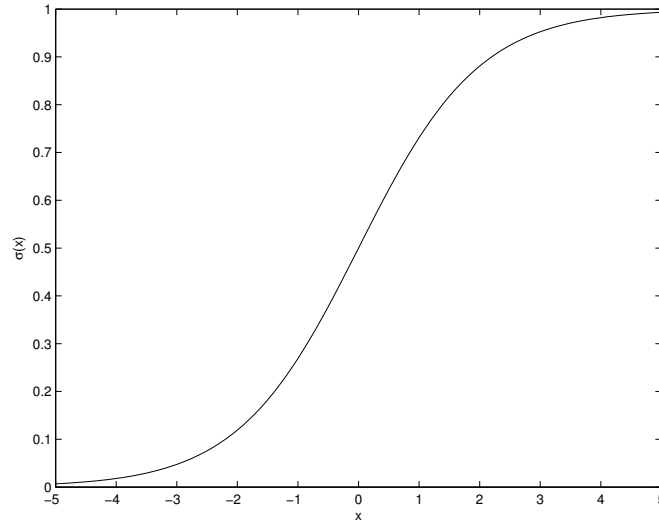


FIGURE 1.15 – Fonction sigmoïde.

4 On considère maintenant le modèle d'une couche de m neurones, c'est-à-dire un ensemble de m neurones g_i ayant la même fonction d'activation σ . Une couche est donc une fonction de \mathbf{R}^n à valeurs dans \mathbf{R}^m dépendant de paramètres w_{ij}, θ_j pour $i = 1, \dots, n$ et $j = 1, \dots, m$.

On a à notre disposition K points $x^k \in \mathbf{R}^n$ et $y^k \in \mathbf{R}^m$, on appelle apprentissage l'estimation par les moindres carrés des paramètres du réseau de neurones formé d'une couche de m neurones.

Écrire le problème au moindres carrés qui définit l'apprentissage. On donnera en particulier la fonction résidus r en précisant clairement l'espace de départ et l'espace d'arrivée.

Bibliographie

- [1] Carpentier. Cours de 3^{ème} année enseiht, filière informatique et mathématiques appliquées. INPT-ENSEEIH, 1983.
- [2] Gerard Cornuejols and Reha Tütüncü. *Optimization Methods in Finance*. Cambridge University Press, 2007.
- [3] J.E. Dennis and Jr. Robert B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM, 1996.
- [4] J.M. Devos. Fermat "le premier homme du monde". France 3, IREM de Toulouse, CRDP Midi-Pyrénées, 1995. Casette vidéo.
- [5] Jean-Baptiste Hiriart-Urruty. *L'Optimisation*. Que sais-je. Presses Universitaires de France, 1996. ISBN : 2 13 047981 2.
- [6] Jean-Baptiste Hiriart-Urruty. *Analyse convexe et optimisation*. Presses Universitaires de France, 2000. ISBN :
- [7] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of Convex Analysis*. Presses Universitaires de France, 2001. ISBN : 978-3-540-42205-1.
- [8] Sylvie Huet, Annie Bouvier, Marie-Anne Gruet, and Emmanuel Jolivet. *Statistical Tools for Nonlinear Regression, A Practical Guide with S-PLUS Examples*. Springer, 1996. ISBN : 0-387-94727-2.
- [9] Sylvie Huet, Emmanuel Jolivet, and Antoine Messéan. *La régression non-minéaire : méthodes et applications à la biologie*. INRA, 1992.
- [10] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer, second edition, 2006.
- [11] R. Tyrrell Rockafellar. *Convec Analysis*. Princeton University Press, 1996.

Index

équation d'Euler, 39
équations normales, 40

cône tangent, 59
condition nécessaire du deuxième ordre, 40
condition suffisante du deuxième ordre, 41

direction tangente, 59

formule de Taylor-Young, 26

hypothèse de qualification des contraintes, 59

inéquation d'Euler, 39

KKT (conditions de —), 61

Lagrangien, 61
lagrangien, 57
lemme de Farkas et Minkowski, 62

multiplicateur de Karush-Kuhn-Tucker, 61
multiplicateur de Lagrange, 61

point critique, 39

théorème de Karush, Kuhn et Tucker, 61
théorème des fonctions implicites, 26

valeur critique, 39