Rappels pour l'optimisation stochastique

Amélie Lambert

Cnam

MPRO - Mise à niveau

1. L'espérance conditionnelle d'un couple de variables aléatoires
   - Espace de probabilités
   - Variables aléatoires
   - Couple de variables aléatoires
     - Le cas discret
     - Le cas continu

2. Rappels d'optimisation
   - Optimisation sans contraintes
   - Optimisation sous contraintes
     - Optimisation sous contraintes d'égalité
     - Optimisation sous contraintes d'inégalité

1. L'espérance conditionelle d'un couple de variables aléatoires
   - Espace de probabilités
   - Variables aléatoires
   - Couple de variables aléatoires
     - Le cas discret
     - Le cas continu
2. Rappels d'optimisation
   - Optimisation sans contraintes
   - Optimisation sous contraintes
     - Optimisation sous contraintes d'égalité
     - Optimisation sous contraintes d'inégalité

Espace de probabilités $(\Omega, F, \mathbb{P})$

- Un *univers* Ω *:* Ensemble évènements-élémentaires ω (résultats possibles) de l'expérience considérée.
  Exemple : Jeter un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$; Observer la durée de vie d'une ampoule : $\Omega = [0, +\infty[$.

Espace de probabilités $(\Omega, F, \mathbb{P})$

- Un *univers* Ω *:* Ensemble évènements-élémentaires ω (résultats possibles) de l'expérience considérée.
  Exemple : Jeter un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$; Observer la durée de vie d'une ampoule : $\Omega = [0, +\infty[$.

- Des *évènements* F *:* Sous ensemble F de l'univers Ω.
  On ne considère pas tous les sous-ensemble de Ω comme évènement, mais seulement ceux qui ont un sens.

Espace de probabilités $(\Omega, F, \mathbb{P})$

- Un *univers* Ω *:* Ensemble évènements-élémentaires ω (résultats possibles) de l'expérience considérée.
  Exemple : Jeter un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$; Observer la durée de vie d'une ampoule : $\Omega = [0, +\infty[$.

- Des *évènements* F *:* Sous ensemble F de l'univers Ω.
  On ne considère pas tous les sous-ensemble de Ω comme évènement, mais seulement ceux qui ont un sens.

- Une *loi de probabilité* $\mathbb{P}$ *:* qui est une application de F dans dans $[0, 1]$ telle que :
  1. L'évènement certain est de probabilité 1 : $\mathbb{P}(\Omega) = 1$
  2. $\mathbb{P}$ est additive, pour toute famille finie ou dénombrable $F_1, F_2, \ldots, F_n$ d'évènements disjoints de F (i.e. $F_k \cap F_j = \emptyset$ si $k \neq j$) on a :

$$\mathbb{P}(\cup_{k=1}^{n} F_k) = \sum_{k=1}^{n} \mathbb{P}(F_k)$$

Propriétés de l'espace de probabilités $(\Omega, F, \mathbb{P})$

Propriété : Soit $(\Omega, F, \mathbb{P})$ un espace de probabilités alors :

i. $\mathbb{P}(\emptyset) = 0$
ii. $\mathbb{P}(\bar{F}) = 1 - \mathbb{P}(F)$
iii. Si A et B sont dans F et si $A \subseteq B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$

Propriétés de l'espace de probabilités $(\Omega, F, \mathbb{P})$

Propriété : Soit $(\Omega, F, \mathbb{P})$ un espace de probabilités alors :

i. $\mathbb{P}(\emptyset) = 0$
ii. $\mathbb{P}(\bar{F}) = 1 - \mathbb{P}(F)$
iii. Si A et B sont dans F et si $A \subseteq B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$

Propriété : Soit $(\Omega, F, \mathbb{P})$ un espace de probabilité alors : Si A et B sont dans F, mais ne sont pas nécéssairement disjoints, alors :

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \text{ d'où } \mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

1. L'espérance conditionelle d'un couple de variables aléatoires
   - Espace de probabilités
   - Variables aléatoires
   - Couple de variables aléatoires
     - Le cas discret
     - Le cas continu
2. Rappels d'optimisation
   - Optimisation sans contraintes
   - Optimisation sous contraintes
     - Optimisation sous contraintes d'égalité
     - Optimisation sous contraintes d'inégalité

Variables aléatoires

Exemple : Une urne contenant des boules blanches et des boules rouges, ces dernières en proportion λ. On tire, au hasard, une boule dans l'urne, et on considère les deux événements élémentaires :
$\omega_1 = \{\text{"la boule est blanche"}\}$, $\omega_2 = \{\text{"la boule est rouge"}\}$.
Pour calculer les probabilités des divers résultats, on introduit une variable aléatoire X qui prend la valeur 0 si ω_1 est réalisé, et la valeur 1 si c'est ω_2.
On a alors $\mathbb{P}(X = 1) = \lambda$ et $\mathbb{P}(X = 0) = 1 - \lambda$

Variables aléatoires

Exemple : Une urne contenant des boules blanches et des boules rouges, ces dernières en proportion λ. On tire, au hasard, une boule dans l'urne, et on considère les deux événements élémentaires :
$\omega_1 = \{\text{"la boule est blanche"}\}$, $\omega_2 = \{\text{"la boule est rouge"}\}$.
Pour calculer les probabilités des divers résultats, on introduit une variable aléatoire X qui prend la valeur 0 si ω_1 est réalisé, et la valeur 1 si c'est ω_2.
On a alors $\mathbb{P}(X = 1) = \lambda$ et $\mathbb{P}(X = 0) = 1 - \lambda$

Définition : Soit un univers Ω associé à une expérience aléatoire, sur lequel on a défini une mesure de probabilité $\mathbb{P}$ et des évènements F. Une *variable aléatoire* X est l'application :

$$(\Omega, F, \mathbb{P}) \to \mathbb{R}$$

$$\omega \to X(\omega)$$

Variables aléatoires continues

Définition : Une *variable aléatoire* est dite *continue* si elle peut prendre toutes les valeurs dans un intervalle donné (borné ou non borné). En règle générale, toutes les variables qui résultent d'une mesure sont de type continu.

Remarque : Pour une variable aléatoire continue, la probabilité associée à l'évènement $\{X = a\}$ est nulle, car il est impossible d'observer exactement cette valeur. On considère alors la probabilité que la variable aléatoire X prenne des valeurs comprises dans un intervalle $[a, b]$ tel que $\mathbb{P}(a \leq X \leq b)$.

Variables aléatoires discrètes

Définition : Une *variable aléatoire discrète* est une variable aléatoire qui ne prend que des valeurs entières, en nombre fini ou dénombrable :

$$(\Omega, F, \mathbb{P}) \to \mathbb{N}$$

$$\omega \to X(\omega)$$

On note $X(\Omega) = \{x_i, i \in I\}$ l'ensemble des valeurs ordonnées prises par X.

On associe à chacune des valeurs possibles de la variable aléatoire la probabilité qui lui correspond. Pour calculer la probabilité que la variable X soit égale a x_i, on cherche tous les événements élémentaires ω_j pour lesquels $X(\omega_j) = x_i$, et on a :

$$\mathbb{P}(X = x_i) = \sum_{j=1}^{k} \mathbb{P}(\{\omega_j\}) \text{ si } X = x_i \text{ sur les évts élémentaires } \omega_1, \ldots, \omega_k$$

1. L'espérance conditionelle d'un couple de variables aléatoires
   - Espace de probabilités
   - Variables aléatoires
   - Couple de variables aléatoires
     - Le cas discret
     - Le cas continu
2. Rappels d'optimisation
   - Optimisation sans contraintes
   - Optimisation sous contraintes
     - Optimisation sous contraintes d'égalité
     - Optimisation sous contraintes d'inégalité

Le cas discret

Couple aléatoire discret

On appelle *couple aléatoire discret* un couple (X, Y) de deux variables aléatoires discrètes définies sur le même espace de probabilité $(\Omega, F, \mathbb{P})$.

On note $X(\Omega) = \{x_i, i \in I\}$ et $Y(\Omega) = \{y_j, j \in J\}$, l'ensemble des valeurs ordonnées prises respectivement par X et Y.

$\Longrightarrow$ Etude de ce qui se passe simultanément pour X et Y.

Couple aléatoire discret

On appelle *couple aléatoire discret* un couple (X, Y) de deux variables aléatoires discrètes définies sur le même espace de probabilité $(\Omega, F, \mathbb{P})$.

On note $X(\Omega) = \{x_i, i \in I\}$ et $Y(\Omega) = \{y_j, j \in J\}$, l'ensemble des valeurs ordonnées prises respectivement par X et Y.

$\Longrightarrow$ Etude de ce qui se passe simultanément pour X et Y.

Exemple : On tire n fois une boule avec remise dans une urne contenant une boule rouge, une boule blanche et une boule verte. X est le nombre de boules rouges obtenues et Y le rang d'apparition de la première boule rouge. ($Y = n + 1$ si aucune boule rouge tirée).
Pour $n \geq 2$, on ne peut pas avoir $X = n$ et $Y = n$, et si $Y = n + 1$, alors $X = 0$.

Loi conjointe ou loi du couple aléatoire discret (X, Y)

Soient X et Y deux variables aléatoires discrètes définies sur le même espace de probabilité $(\Omega, F, \mathbb{P})$.

On note $X(\Omega) = \{x_i, i \in I\}$ et $Y(\Omega) = \{y_j, j \in J\}$, l'ensemble des valeurs ordonnées prises respectivement par X et Y, et $\mathbb{P}(X = x_i \cap Y = y_j)$ la probabilité de tomber sur le couple (x_i, y_j).

Définition : On appelle *loi conjointe* ou *loi du couple aléatoire discret* (X, Y) l'application p_{ij} :

$$p_{ij} : \begin{array}{rcl} X(\Omega) \times Y(\Omega) & \rightarrow & [0,1] \\ (x, y) & \rightarrow & \mathbb{P}(X = x_i \cap Y = y_j) \end{array}$$

Il est clair que :

$$\left\{ \begin{array}{l} 0 \leq p_{ij} \leq 1 \\ \displaystyle\sum_{i \in I, j \in J} p_{ij} = 1 \end{array} \right.$$

Exemple : loi jointe d'un couple aléatoire discret

On tire deux chiffres au hasard, indépendamment et de façon équiprobable entre 1 et 3. Soient :

- X la variable aléatoire qui représente le maximum des chiffres obtenus
- Y la variable aléatoire qui représente la somme des chiffres obtenus

La loi jointe du couple (X, Y) pour le maximum et la somme se représente sous forme du tableau suivant :

choix	probabilité
(1,1)	$\frac{1}{9}$
(1,2)	$\frac{1}{9}$
(1,3)	$\frac{1}{9}$
(2,1)	$\frac{1}{9}$
(2,2)	$\frac{1}{9}$
(2,3)	$\frac{1}{9}$
(3,1)	$\frac{1}{9}$
(3,2)	$\frac{1}{9}$
(3,3)	$\frac{1}{9}$

X \ Y	2	3	4	5	6
1	$\frac{1}{9}$	0	0	0	0
2	0	$\frac{2}{9}$	$\frac{1}{9}$	0	0
3	0	0	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Lois marginales

Soit (X, Y) un couple aléatoire. Les variables aléatoires X et Y sont appelées *variables marginales* du couple (X, Y).

La *loi marginale* de X est entièrement déterminée par les probabilités $p_{i.}$ de tomber sur les points x_i :

$$p_{i.} = \mathbb{P}(X = x_i) = \sum_{j \in J} \mathbb{P}(X = x_i \cap Y = y_j) = \sum_{j \in J} p_{ij}$$

De même pour la *loi marginale* de Y et les probabilités $p_{.j}$ de tomber sur les points y_j :

$$p_{.j} = \mathbb{P}(Y = y_j) = \sum_{i \in I} \mathbb{P}(X = x_i \cap Y = y_j) = \sum_{i \in I} p_{ij}$$

Exemple (suite)

On tire deux chiffres au hasard, indépendamment et de façon équiprobable entre 1 et 3.

- X la variable aléatoire qui représente le maximum des chiffres obtenus
- Y la variable aléatoire qui représente la somme des chiffres obtenus

On calcule aisément les lois marginales de X et Y : sur le tableau de la loi jointe du couple (X, Y) pour le max et la somme, il suffit de sommer sur chaque ligne pour la loi de X et sur chaque colonne pour la loi de Y :

choix	probabilité
(1,1)	$\frac{1}{9}$
(1,2)	$\frac{1}{9}$
(1,3)	$\frac{1}{9}$
(2,1)	$\frac{1}{9}$
(2,2)	$\frac{1}{9}$
(2,3)	$\frac{1}{9}$
(3,1)	$\frac{1}{9}$
(3,2)	$\frac{1}{9}$
(3,3)	$\frac{1}{9}$

X \ Y	2	3	4	5	6	$p_{i.}$	
1	$\frac{1}{9}$	0	0	0	0	$p_{1.}$	$\frac{1}{9}$
2	0	$\frac{2}{9}$	$\frac{1}{9}$	0	0	$p_{2.}$	$\frac{3}{9}$
3	0	0	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$p_{3.}$	$\frac{5}{9}$
$p_{.j}$	$p_{.2}$	$p_{.3}$	$p_{.4}$	$p_{.5}$	$p_{.6}$		
	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$		

Loi marginale et indépendance

Les variables aléatoires X et Y sont *indépendantes* si et seulement si pour tout i, j, les évènements $\{X = x_i\}$ et $\{Y = y_j\}$ sont indépendants, c'est-à-dire si :

$$\forall (i,j) \in I \times J \quad \mathbb{P}(X = x_i \cap Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)$$

ou bien encore

$$p_{ij} = p_{i.}p_{.j}$$

Exemple (suite)

On tire deux chiffres au hasard, indépendamment et de façon équiprobable entre 1 et 3.

- X la variable aléatoire qui représente le maximum des chiffres obtenus
- Y la variable aléatoire qui représente la somme des chiffres obtenus

Il est clair que X et Y ne sont pas indépendantes puisque par exemple :

$$p_{12} = \frac{1}{9} \neq p_{1.} \times p_{.2} = \frac{1}{9} \times \frac{1}{9} = \frac{1}{81}$$

Exemple : Jeu de cartes

On tire une carte au hasard dans un jeu de 32 cartes. Le résultat de ce tirage est représenté par le couple aléatoire (X, Y), où :

- X est la variable aléatoire représentant la couleur : X appartient à l'ensemble {Pique, Coeur, Carreau, Trèfle},
- Y est la variable aléatoire représentant la valeur : Y appartient à l'ensemble {7, 8, 9, 10, Valet, Dame, Roi, As }

Nous avons que :

$$\forall (i,j) \in I \times J \quad \mathbb{P}(X = x_i \cap Y = y_i) = \frac{1}{32} = \frac{1}{4} \times \frac{1}{8} = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_i)$$

donc X et Y sont indépendantes.

Remarque

Soit $i \in I$ fixé. Notons qu'on peut avoir $p_{ij} = 0$, c'est-à-dire que l'événement $\{X = x_i \cap Y = y_j\}$ ne se réalise jamais.

Par contre, on exclut le cas où $p_{i.} = 0$: ceci signifierait que X ne prend jamais la valeur x_i , auquel cas cette valeur n'aurait rien à faire dans $X(\Omega)$.

Probabilité conditionnelle

Puisque chacune des probabilités $p_{i.}$ est non nulle, on peut définir la probabilité conditionnelle de $Y = y_j$ sachant $X = x_i$:

Soit $x_i \in X(\Omega)$, la *probabilités conditionnelle* de Y sachant $X = x_i$ est la loi discrète prenant les valeurs y_j avec les probabilités :

$$\begin{aligned} p_{j|i} &= \mathbb{P}(Y = y_j | X = x_i) \\ &= \frac{\mathbb{P}(Y=y_j \cap X=x_i)}{\mathbb{P}(X=x_i)} = \frac{p_{ij}}{p_{i.}} \end{aligned}$$

Espérance conditionnelle

Si X est figée à x_i, il est naturel de considérer la valeur moyenne de la variable aléatoire Y lorsque $X = x_i$: c'est ce qu'on appelle l'*espérance conditionnelle* de Y sachant $X = x_i$. Elle s'écrit :

$$\mathbb{E}[Y|X = x_i] = \sum_{j \in J} p_{j|i} y_j$$

et on la note $\mathbb{E}[Y|X]$.

Le cas continu

Fonction de densité jointe

On appelle *fonction de densité* une fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ telle que :

$$\begin{cases} f(x,y) \geq 0 \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dxdy = 1 \end{cases}$$

Soient X et Y deux variables aléatoires définies sur le même espace de probabilité $(\Omega, F, \mathbb{P})$. Le couple aléatoire (X, Y) est dit à densité si il existe une fonction $f(x,y)$, ayant les propriétés ci-dessus, telle que :

$$\mathbb{P}(X \leq a \cap Y \leq b) = \int_{-\infty}^{b} [\int_{-\infty}^{a} f(x,y)dx]dy$$

La fonction $f(x,y)$ est appelée fonction de densité jointe du couple (X, Y).

Fonction de repartition jointe

Soient X et Y deux variables aléatoires définies sur le même espace de probabilité $(\Omega, F, \mathbb{P})$. La *fonction de répartition jointe* du couple aléatoire (X, Y) est définie par :

$$\begin{aligned} F(x,y) &= \mathbb{P}((X < a) \cap (Y < b)) \\ &= \int_{-\infty}^{b} [\int_{-\infty}^{a} f(x,y)dx]dy \\ &= \int_{-\infty}^{a} [\int_{-\infty}^{b} f(x,y)dy]dx \end{aligned}$$

et en tout point où $f(x,y)$ est continue, on a :

$$f(x,y) = \frac{\partial^2 F}{\partial y \partial x}(x,y)$$

La fonction $F(X, Y)$ est appelée fonction de répartition du couple (X, Y).

Lois de densité marginales

Si l'on s'intéresse à un événement sur X quelle que soit la valeur prise par Y, on obtient la loi de densité marginale de la variable aléatoire X :

Pour le couple (X, Y), les *lois marginales* de X et Y sont :

$$f_X(a) = \mathbb{P}(X < a) = \mathbb{P}(X < a \cap Y \in \mathbb{R}) = \int_{-\infty}^{+\infty} \Big(\int_{-\infty}^{a} f(x, y) dx \Big) dy$$

$$f_Y(b) = \mathbb{P}(Y < b) = \mathbb{P}(X \in \mathbb{R} \cap Y < b) = \int_{-\infty}^{+\infty} \Big(\int_{-\infty}^{b} f(x, y) dy \Big) dx$$

Indépendance

Les variables aléatoires X et Y sont *indépendantes* si et seulement si la fonction de densité du couple est égale au produit des fonctions de densité marginales, i.e. si :

$$f(a,b) = f_X(a)f_Y(b)$$

Probabilité conditionnelle

Soit (X, Y) un couple aléatoire absolument continu, de densité de probabilité $f(x, y)$. Soit f_X la densité de probabilité de X et un réel a tel que $f_X(a) \neq 0$.

La *probabilité conditionnnelle* de Y liée par la condition $X = a$ est définie par sa densité de probabilité :

$$f(y|X < a) = \frac{f(a, y)}{f_X(a)}$$

Espérance conditionnelle

On veut maintenant définir l'*espérance conditionnelle* d'une variable aléatoire continue par rapport à une autre variable aléatoire continue. Pour a fixé, l'espérance conditionnelle de Y sachant $X < a$ est :

$$\mathbb{E}[Y|X < a] = \int_{-\infty}^{+\infty} yf(y|X < a)dy$$

1 L'espérance conditionelle d'un couple de variables aléatoires
- Espace de probabilités
- Variables aléatoires
- Couple de variables aléatoires
  - Le cas discret
  - Le cas continu

2 Rappels d'optimisation
- Optimisation sans contraintes
- Optimisation sous contraintes
  - Optimisation sous contraintes d'égalité
  - Optimisation sous contraintes d'inégalité

Idée generale

D'un point de vue mathématique, l'optimisation consiste à rechercher le minimum ou le maximum d'une fonction avec ou sans contraintes.

Cependant, on limite souvent l'optimisation à une recherche de minimum. En effet, maximiser une fonction f revient à minimiser la fonction $-f$.

On ne parlera donc ici que de minimisation.

1. L'espérance conditionelle d'un couple de variables aléatoires
   - Espace de probabilités
   - Variables aléatoires
   - Couple de variables aléatoires
     - Le cas discret
     - Le cas continu

2. Rappels d'optimisation
   - Optimisation sans contraintes
   - Optimisation sous contraintes
     - Optimisation sous contraintes d'égalité
     - Optimisation sous contraintes d'inégalité

Optimisation sans contraintes

Définition du problème

Définition : Soit X une partie de $\mathbb{R}$, et une fonction $f : X \to \mathbb{R}$, on appelle *problème d'optimisation sans contraintes* le problème suivant (P) :

$$\text{Trouver } x^* = \underset{x \in X}{\operatorname{argmin}} \quad f(x)$$

Afin de pouvoir calculer ou approcher facilement la solution de notre problème, il est intéressant de connaître les hypothèses garantissant l'existence et l'unicité de cette solution.

Définitions : Minimum local ou global

Minimum local :

f possède un minimum local u si

$$\exists \delta > 0 \text{ tel que } f(u) \leq f(x), \forall x, ||x - u|| \leq \delta$$

Minimum global :

f possède un minimum global u si

$$\forall x, \quad f(u) \leq f(x)$$

Existence d'un minimum

Théorème : Si f est continue et si X est un ensemble compact, alors le problème d'optimisation de f admet une solution $x^* \in X$ telle que :

$$\forall x \in X, \quad f(x^*) \leq f(x)$$

Idée de démonstration :

Un espace est compact s'il peut être couvert par un nombre fini d'ouvert, il est donc borné. f étant continue et X étant compact, f atteint ses bornes et $x^* = \min_{x \in X} \quad f(x)$ existe.

Unicité du minimum

Théorème : Soit $f : X \subset \mathbb{R}^n \to \mathbb{R}$ strictement convexe sur X convexe. Si il existe, le minimum de f sur X est unique.

Idée de démonstration :
On suppose qu'il existe deux minima u et v distincts et on démontre avec la convexité que cela est impossible

Conditions nécessaires d'optimalité

Théorème *Conditions nécessaires d'optimalité du 1er ordre :*
Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable sur V un voisinage ouvert de u, si u est un minimum (local ou global) de f, alors $\nabla f(u) = 0$.

Conditions nécessaires d'optimalité

Théorème *Conditions nécessaires d'optimalité du 1er ordre :*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable sur V un voisinage ouvert de u, si u est un minimum (local ou global) de f, alors $\nabla f(u) = 0$.

Definition : Un point u de $\mathbb{R}^n$ vérifiant $\nabla f(u) = 0$ est appelé *point critique (ou stationnaire)*.

Conditions nécessaires d'optimalité

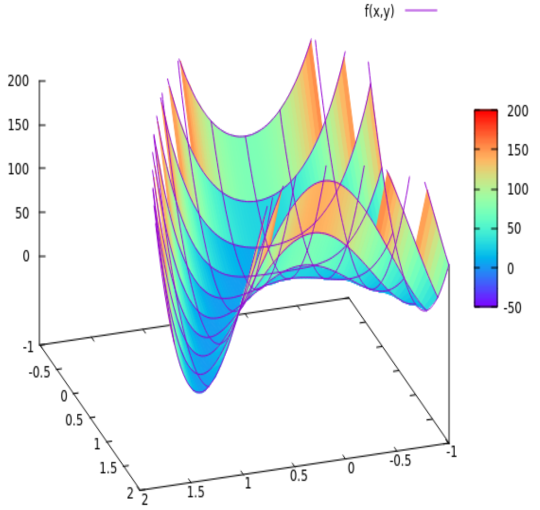
Théorème *Conditions nécessaires d'optimalité du 1^{er} ordre :*
Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable sur V un voisinage ouvert de u, si u est un minimum (local ou global) de f, alors $\nabla f(u) = 0$.

Definition : Un point u de $\mathbb{R}^n$ vérifiant $\nabla f(u) = 0$ est appelé *point critique (ou stationnaire)*.

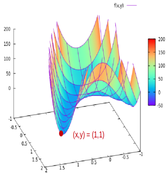
Théorème *Conditions nécessaires d'optimalité du 2^{nd} ordre :*
Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction deux fois différentiable sur V un voisinage ouvert de u, si u est un minimum (local ou global) de f, alors $\nabla f(u) = 0$ et $\nabla^2 f(u)$ est semi définie positive et f est localement convexe en u.

Exemple 1 (1/2)

Soit la fonction $f(x, y) = 100(y - x^2)^2 + (1 - x)^2$, avec $-1 \leq x \leq 2$ et $-1 \leq y \leq 2$:

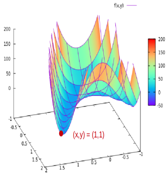


Exemple 1 (2/2)



Soit $(x^*, y^*) = (1, 1)$ un min. local de
$f(x, y) = 100(y - x^2)^2 + (1 - x)^2$ sur $-1 \leq x, y \leq 2$

Exemple 1 (2/2)



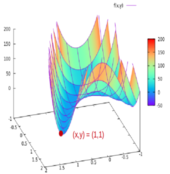
Soit $(x^*, y^*) = (1, 1)$ un min. local de
$f(x, y) = 100(y - x^2)^2 + (1 - x)^2$ sur $-1 \leq x, y \leq 2$

- Conditions nécessaires du 1^{er} ordre :

$$\nabla f(x,y) = \begin{pmatrix} 400x^3 - 400xy + 2x - 2 \\ 200y - 200x^2 \end{pmatrix}$$

et $\nabla f(1, 1) = 0$

Exemple 1 (2/2)



Soit $(x^*, y^*) = (1, 1)$ un min. local de
$f(x, y) = 100(y - x^2)^2 + (1 - x)^2$ sur $-1 \leq x, y \leq 2$

- Conditions nécessaires du 1^{er} ordre :

$$\nabla f(x, y) = \begin{pmatrix} 400x^3 - 400xy + 2x - 2 \\ 200y - 200x^2 \end{pmatrix}$$

  et $\nabla f(1, 1) = 0$

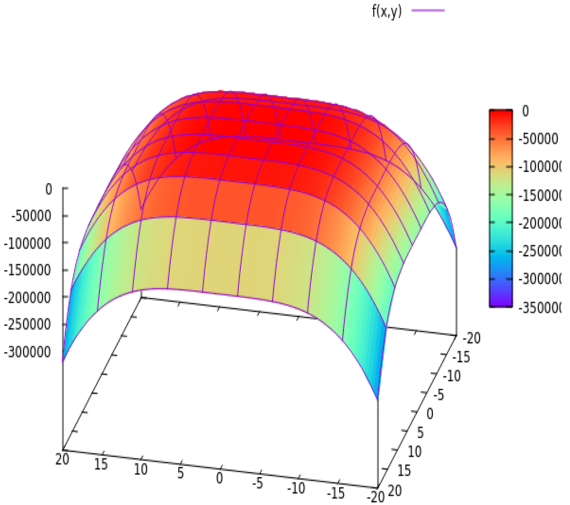
- Conditions nécessaires du 2^{nd} ordre :

$$\nabla^2 f(x, y) = \begin{pmatrix} 1200x^2 - 400y + 2 & -400x \\ -400x & 200 \end{pmatrix}$$

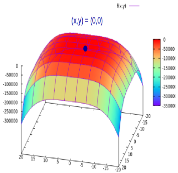
  et $\nabla^2 f(1, 1) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} \succeq 0$ (Val prop. : 0.39936 et 1001.6)

Exemple 2 (2/2)

Soit la fonction $f(x, y) = -x^4 - y^4$, avec $-20 \leq x \leq 20$ et $-20 \leq y \leq 20$:

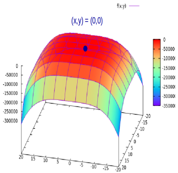


Exemple 2 (2/2)



Soient $f(x,y) = -x^4 - y^4$ et $(x^*, y^*) = (0,0)$
qui vérifie les conditions nécessaires d'optimalité

Exemple 2 (2/2)



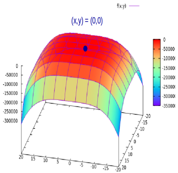
Soient $f(x,y) = -x^4 - y^4$ et $(x^*, y^*) = (0,0)$
qui vérifie les conditions nécessaires d'optimalité

- Conditions nécessaires du 1^{er} ordre :

$$\nabla f(x,y) = \begin{pmatrix} -4x^3 \\ -4y^3 \end{pmatrix}$$

  et $\nabla f(0,0) = 0$

Exemple 2 (2/2)



Soient $f(x,y) = -x^4 - y^4$ et $(x^*, y^*) = (0,0)$
qui vérifie les conditions nécessaires d'optimalité

- Conditions nécessaires du 1^{er} ordre :

$$\nabla f(x,y) = \begin{pmatrix} -4x^3 \\ -4y^3 \end{pmatrix}$$

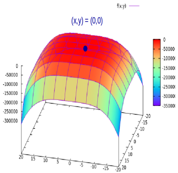
  et $\nabla f(0,0) = 0$

- Conditions nécessaires du 2^{nd} ordre :

$$\nabla^2 f(x,y) = \begin{pmatrix} -12x^2 & 0 \\ 0 & -12y^2 \end{pmatrix}$$

  et $\nabla^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ est semi-définie positif.

Exemple 2 (2/2)



Soient $f(x,y) = -x^4 - y^4$ et $(x^*, y^*) = (0,0)$
qui vérifie les conditions nécessaires d'optimalité

- Conditions nécessaires du 1^{er} ordre :

$$\nabla f(x,y) = \begin{pmatrix} -4x^3 \\ -4y^3 \end{pmatrix}$$

  et $\nabla f(0,0) = 0$

- Conditions nécessaires du 2^{nd} ordre :

$$\nabla^2 f(x,y) = \begin{pmatrix} -12x^2 & 0 \\ 0 & -12y^2 \end{pmatrix}$$

  et $\nabla^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ est semi-définie positif.

mais $(0,0)$ n'est pas un minimum local

Conditions nécessaires et suffisantes d'optimalité

Dans le cas où f est une fonction convexe, les conditions nécessaires deviennent nécessaires et suffisantes.

Theorème *Conditions nécessaires et suffisantes d'optimalité globale*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable sur $\mathbb{R}$ et convexe ($\nabla^2 f(u)$ est définie positif). Un point u est un minimum global de f si et seulement si $\nabla f(u) = 0$.

1. L'espérance conditionelle d'un couple de variables aléatoires
   - Espace de probabilités
   - Variables aléatoires
   - Couple de variables aléatoires
     - Le cas discret
     - Le cas continu

2. Rappels d'optimisation
   - Optimisation sans contraintes
   - **Optimisation sous contraintes**
     - Optimisation sous contraintes d'égalité
     - Optimisation sous contraintes d'inégalité

Optimisation sous contraintes d'égalité

Optimisation sous contraintes d'égalité

Définition : Soit X un ouvert de $\mathbb{R}^n$, $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^m$. On appelle *problème d'optimisation avec contraintes d'égalité* le problème (P) :

$$\text{Trouver } x^* = \underset{x \in X}{\operatorname{argmin}} \quad f(x) \text{ sous les contraintes } g_i(x) = 0 \; i = 1, \ldots, m$$

Dans la suite on note : $g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$ l'ensemble des contraintes, et $\mathcal{C} = \{x \in X : g_i(x) = 0\}$ la surface des contraintes.

Forme géométrique de l'optimalité au premier ordre

Definition 1 : Soit x un point de $\mathcal{C}$. On définit $T(x)$ l'espace tangeant à la surface $\mathcal{C}$ en x comme suit :

$$T(x) = \{v \in \mathbb{R}^m : \nabla g(x)v = 0\}$$

Forme géométrique de l'optimalité au premier ordre

Definition 1 : Soit x un point de $\mathcal{C}$. On définit $T(x)$ l'espace tangeant à la surface $\mathcal{C}$ en x comme suit :

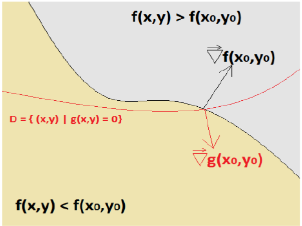
$$T(x) = \{v \in \mathbb{R}^m : \nabla g(x)v = 0\}$$

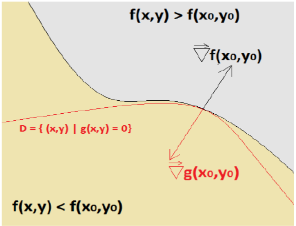
Definition 2 : Soit x un point de $\mathcal{C}$. On définit $N(x)$ l'espace normal à la surface $\mathcal{C}$ en x comme l'espace vectoriel orthogonal à $T(x)$. C'est le sous espace vectoriel engendré par $\nabla g_1(x), \ldots, \nabla g_m(x)$:

$$N(x) = \mathrm{vect}\ \{\nabla g_1(x), \ldots, \nabla g_m(x)\}$$

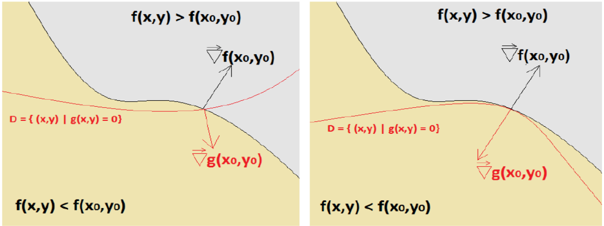
On suppose ici que les vecteurs $\nabla g_1(x), \ldots, \nabla g_m(x)$ forment une partie libre.

Optimisation locale sous contraintes d'égalité





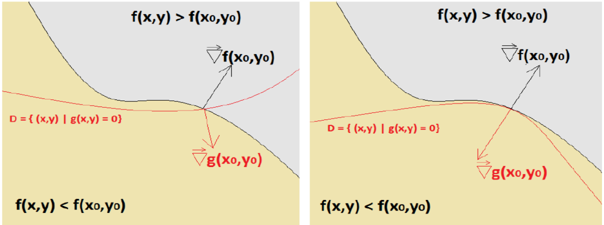
Optimisation locale sous contraintes d'égalité



Comme $N(x^*)$ est l'espace vectoriel engendré par les vecteurs $\nabla g_1(x^*), \ldots, \nabla g_m(x^*)$, il existe des coefficients réels $\alpha_1, \ldots, \alpha_m$ tels que $\nabla f(x^*) = \alpha_1 \nabla g_1(x^*) + \ldots + \alpha_m \nabla g_m(x^*)$.

En posant $\lambda_1 = -\alpha_1, \ldots, \lambda_m = -\alpha_m$, on déduit le théorème suivant :

Optimisation locale sous contraintes d'égalité



Comme $N(x^*)$ est l'espace vectoriel engendré par les vecteurs $\nabla g_1(x^*), \ldots, \nabla g_m(x^*)$, il existe des coefficients réels $\alpha_1, \ldots, \alpha_m$ tels que $\nabla f(x^*) = \alpha_1 \nabla g_1(x^*) + \ldots + \alpha_m \nabla g_m(x^*)$.
En posant $\lambda_1 = -\alpha_1, \ldots, \lambda_m = -\alpha_m$, on déduit le théorème suivant :

Théorème (multiplicateurs de Lagrange) : Soit x^* un minimiseur local de $f(x)$ sous la contrainte $g(x) = 0$, alors il existe des réels $\lambda_1^*, \ldots, \lambda_m^*$, appelés multiplicateurs de Lagrange, tels que :

$$\nabla f(x^*) + \lambda_1^* \nabla g_1(x^*) + \ldots + \lambda_m^* \nabla g_m(x^*) = 0$$

Fonction de lagrange (ou Lagrangien)

Définition : On introduit la fonction de Lagrange (ou Lagrangien) associé au problème (P) la fonction obtenu comme suit, en associant un multiplicateur de Lagrange λ_i à chaque contrainte g_i :

$$L : (x, \lambda) \in X \times \mathbb{R}^m \to L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

Conditions nécessaires d'optimalité globale du 1^{er} ordre

Théorème de Lagrange : (*Conditions nécessaires du 1^{er} ordre*) Soit $f : X \subset \mathbb{R}^n \to \mathbb{R}$ différentiable et $g : \mathbb{R}^n \to \mathbb{R}^m$ telle que les g_i sont différentiables. Soit $\mathcal{C} = \{x \in X : g_i(x) = 0\}$. Si x^* est un minimum local de f sur $\mathcal{C}$ et si les ∇g_i en x* sont linéairement indépendants alors :

$$\exists ! \lambda^* \in \mathbb{R}^m \text{ tel que } \nabla L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$$

Conditions nécessaires d'optimalité globale du 1$^{\text{er}}$ ordre

Théorème de Lagrange : (*Conditions nécessaires du 1er ordre*) Soit $f : X \subset \mathbb{R}^n \to \mathbb{R}$ différentiable et $g : \mathbb{R}^n \to \mathbb{R}^m$ telle que les g_i sont différentiables. Soit $\mathcal{C} = \{x \in X : g_i(x) = 0\}$. Si x^* est un minimum local de f sur $\mathcal{C}$ et si les ∇g_i en x* sont linéairement indépendants alors :

$$\exists ! \lambda^* \in \mathbb{R}^m \text{ tel que } \nabla L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$$

Démonstration : On a $\forall i = 1, \ldots, n \ \frac{\partial L(x^*,\lambda^*)}{\partial x_i} = \frac{\partial f}{\partial x_i}(x^*) + \sum_{i=1}^{m} \lambda_i^* \frac{\partial g_i}{\partial x_i}(x^*)$ et $\forall j = 1, \ldots, m \ \frac{\partial L(x^*,\lambda^*)}{\partial \lambda_j} = g_j(x^*)$, donc on a bien

$$\nabla L(x, \lambda) = 0 \Longleftrightarrow \begin{cases} \nabla f(x^*) + \lambda_1 \nabla g_1(x^*) + \ldots + \lambda_m \nabla g_m(x^*) = 0 \\ g(x^*) = 0 \end{cases}$$

Conditions suffisantes d'optimalité globale du 1^{er} ordre

Théorème : (*Conditions suffisantes du 1^{er} ordre*) Supposons que le lagrangien associé au problème de départ :$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$ possède un minimum local x^* sur X lorsque le vecteur de multiplicateurs $\lambda = \lambda^*$ et $g_i(x^*) = 0$ pour tout $i = 1, \ldots, m$, si de plus f est convexe et les g_i sont des fonctions linéaires, alors x^* est un optimum global.

Exemple

Soit le problème

$$(P)\begin{cases} \min & f(x,y) = x^2 + 2y^2 \\ s.c. & g_1(x,y) = x + y - b = 0 \end{cases}$$

Exemple

Soit le problème

$$(P)\left\{\begin{array}{ll} \min & f(x,y) = x^2 + 2y^2 \\ s.c. & g_1(x,y) = x + y - b = 0 \end{array}\right.$$

Son lagrangien est $L(\lambda, x, y) = (x^2 + 2y^2) + \lambda(x + y - b)$

Exemple

Soit le problème

$$(P)\begin{cases} \min & f(x,y) = x^2 + 2y^2 \\ s.c. & g_1(x,y) = x + y - b = 0 \end{cases}$$

Son lagrangien est $L(\lambda, x, y) = (x^2 + 2y^2) + \lambda(x + y - b)$

L convexe en (x, y). Donc minimum atteint lorsque $\nabla_{x,y} L(\lambda, x, y) = 0$

$$\nabla_{x,y} L(\lambda, x, y) = \begin{bmatrix} 2x + \lambda \\ 4y + \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longleftrightarrow \left\{ \begin{array}{l} \lambda = -2x \\ \lambda = -4y \end{array} \right\} \Longleftrightarrow x = 2y$$

D'après la contrainte, on a $x + y - b = 2y + y - b = 3y - b = 0$

Donc le minimum est atteint quand $y = \frac{b}{3}$, $x = \frac{2b}{3}$ et $\lambda = -2x = -\frac{4b}{3}$

Optimisation sous contraintes d'inégalité

Optimisation sous contraintes d'inégalité

Définition : Soit X un ouvert de $\mathbb{R}^n$, $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^m$. On appelle problème d'optimisation avec contraintes d'inégalité le problème suivant (P) :

$$\text{Trouver } x^* = \underset{x \in X}{\operatorname{argmin}} \quad f(x) \text{ sous les contraintes } g_i(x) \leq 0 \; i = 1, \ldots, m$$

Conditions de Karush-Kuhn-Tucker du premier odre

Hypothèses :
Si X est convexe, si f et g_i sont différentiables et convexes dans le problème :

$$\text{Trouver } x^* = \min_{x \in X} \quad f(x) \text{ sous les contraintes } g_i(x) \leq 0 \; i = 1, \ldots, m$$

et si $\lambda_i \geq 0$, $i = 1, \ldots, m$, alors le lagrangien :

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

est aussi une fonction convexe sur X
(puisque $\lambda_i \geq 0$ et $g_i(x)$ convexe $\Longrightarrow \lambda_i g_i(x)$ convexe, et qu'une somme de fonctions convexes est convexe)

Conditions de Karush-Kuhn-Tucker du premier odre

Si f et g_i sont differentiables et convexes et si $\lambda \geq 0$, alors le lagrangien $L(\lambda, x)$ est une fonction différentiable et convexe en x, et il s'ensuit qu'il possède un minimum global en x^* sur X lorsque $\lambda = \lambda^*$ si :

$$\nabla_x L(x, \lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x) = 0$$

Théorème :

S'il existe un λ^* tel que :

$$\begin{aligned}
&\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0 \\
&\lambda_i^* g_i(x^*) = 0 \qquad i = 1, \ldots, m \\
&g_i(x^*) \leq 0 \qquad i = 1, \ldots, m \\
&\lambda_i^* \geq 0 \qquad i = 1, \ldots, m
\end{aligned}$$

Alors x^* est une solution optimale globale du problème de départ.

Exemple

Soit le problème

$$(P) \begin{cases} \min & f(x) = x^2 \\ s.c. & g_1(x) = 2x + 5 \leq 0 \end{cases}$$

Exemple

Soit le problème

$$(P)\begin{cases} \min & f(x) = x^2 \\ s.c. & g_1(x) = 2x + 5 \leq 0 \end{cases}$$

Son lagrangien est $L(\lambda, x) = x^2 + \lambda(2x + 5)$

Exemple

Soit le problème

$$(P)\begin{cases} \min & f(x) = x^2 \\ s.c. & g_1(x) = 2x + 5 \leq 0 \end{cases}$$

Son lagrangien est $L(\lambda, x) = x^2 + \lambda(2x + 5)$

L convexe en (x). Donc minimum atteint lorsque $\nabla_x L(\lambda, x) = 0$

$$\nabla_x L(\lambda, x) = 2x + 2\lambda = 0 \Longleftrightarrow x = -\lambda$$

A l'optimum, il faut que $(2x + 5)\lambda = 0$, comme $g_1(x)$ est une contrainte d'inégalité, on a $\lambda \geq 0$, de plus pour que $g_1(x)$ soit satisfaite il faut que $x \neq 0$, on a donc $\lambda > 0$ et $2x + 5 = 0$.

$$2x + 5 = -2\lambda + 5 = 0 \Longleftrightarrow \lambda = \frac{5}{2} \text{ et } x = -\frac{5}{2}$$