Université Paris VI
Master 1 : Modèles stochastiques pour la finance

**TD 18 - 19 : Optimisation sous contraintes - Filtrage linéaire - Vecteurs gaussiens - Filtres de Kalman Bucy**

On admettra ici que le théorème du cours donné dans le paragraphe Multiplicateurs de Lagrange est valable pour des fonctions définies sur un ouvert de $\mathbb{R}^n$ au lieu de $\mathbb{R}^n$ : si $U$ est un ouvert de $\mathbb{R}^n$ et $f_1, \dots, f_d, G$ sont des fonctions de $U$ vers $\mathbb{R}$, si $V = \{x \in U \,;\, f_1(x) = \cdots = f_d(x) = 0\}$, si $x_0 \in V$ est tel que

- $f_1, \dots, f_d, G$ sont différentiables au voisinage de $x_0$,
- $\nabla f_1(x_0), \dots, \nabla f_d(x_0)$ sont linéairement indépendants,
- $G_{|V}$ admet un extremum en $x_0$,

alors $\nabla G(x_0)$ est une combinaison linéaire des vecteurs $\nabla f_1(x_0), \dots, \nabla f_d(x_0)$.

Les nombres $\lambda_1, \dots, \lambda_d$ tels que

$$\nabla G(x_0) = \sum_{i=1}^{d} \lambda_i \nabla f_i(x_0)$$

sont alors appelés *multiplicateurs de Lagrange*, en $x_0$, du problème d'optimisation de $G$ sous la contrainte $f_1(x) = \cdots = f_d(x) = 0$.

**1. Nécessité du fait que l'ensemble soit ouvert.** Montrer que le résultat précédent n'est pas valable si $U$ n'est pas ouvert.

**2. Optimisation sous contraintes avec lagrangien (1).** Soit $n \in \mathbb{N}^*$, $\alpha_1, \dots, \alpha_n > 0$, $p = (p_1, \dots, p_n) \in (\mathbb{R}^{+*})^n$, et $m > 0$.

On considère, à une date fixée (qui ne variera pas) $n$ actifs de prix unitaires respectifs $p_1, \ \dots, p_n$. On a une somme de $m$ euros que l'on cherche à répartir dans des quantités positives ou nulles $x_1, \dots, x_n$ de ces actifs. L'utilité associée à une telle répartition est donnée par la fonction

$$f \ : \ x \in (\mathbb{R}^+)^n \quad \mapsto \quad x_1^{\alpha_1} \cdots x_n^{\alpha_n} \in \mathbb{R}.$$

a) Traduire le problème en un problème d'optimisation sous contrainte.
b) Montrer que ce problème a au moins une solution.
c) En utilisant les multiplicateurs de Lagrange, donner cette solution.

**3. Optimisation sous contraintes avec lagrangien (2).** Soit $n \in \mathbb{N}^*$ et $A > 0$. On considère, à la date 0, $n$ actifs de prix unitaires respectifs $p_1, \dots, p_n$ pour lesquels les retours sur investissement, à la date 1 (actualisés à la date 0), sont des variables aléatoires $Y_1, \ \dots, Y_n$. On pose, pour tout $i$,

$$X_i = Y_i - p_i = \text{ le rendement effectif unitaire de l'actif } i.$$

À la date 0, on achète des quantités positives ou nulles $x_1, \dots, x_n$ de ces actifs. Le rendement total est donc $\langle x, X \rangle$, où $x = (x_1, \dots, x_n)$ et $X = (X_1, \dots, X_n)$. La somme totale des investissements n'est pas limitée, mais le risque associé à cet investissement, dont on ne précise pas la nature, est donné par

$$Z_x = \sum_{i=1}^{n} x_i \ln(x_i) X_i,$$

avec la convention $0 \ln(0) = 0$.

On suppose que $X$ est d'espérance $m = (m_1, \dots, m_n)$ connue avec, pour tout $i$, $m_i > 0$, et de matrice de covariance $K \in M_n(\mathbb{R})$ connue.

On se fixe un rendement moyen égal à $A$. On cherche les $x_i \geq 0$ pour lesquels le rendement moyen vaudra $A$ et pour lesquels le risque moyen sera minimal.

a) Traduire le problème en un problème d'optimisation sous contrainte.
b) Montrer que ce problème a au moins une solution.
c) On admettra que aucune solution n'a de coordonnée nulle (cela peut facilement être démontré en considérant le comportement de la fonction à minimiser au voisinage des points ayant une coordonnée nulle). En utilisant les multiplicateurs de Lagrange, donner cette solution.

**4. Contrôle optimal linéaire.** On considère le système

$$X_{n+1} = aX_n + bU_n + W_n,$$

où les $W_i$ sont des vaiid gaussiennes $\mathcal{N}(0,\sigma^2)$.

On cherche à minimiser

$$\mathbb{E}_x[\sum_{k=0}^{N-1}(qX_k^2+rU_k^2)+\gamma X_N^2],$$

où $q,\gamma\geq 0$, $r>0$.

Donner, en appliquant le cours, en fonction de la formule de la transformation de Riccati correspondant à ce problème, le contrôle optimal et la valeur du minimum.

**5. Une définition alternative des vecteurs gaussiens.** On assimile les vecteurs de $\mathbb{R}^n$ à des matrices colonnes. Le produit scalaire de $x,y\in\mathbb{R}^n$ est alors $y^*x$ ($^*$ désignant la transposée).

On appelle, ici, *variable aléatoire réelle gaussienne* toute variable aléatoire réelle $Y$ telle qu'il existe $\mu\in\mathbb{R}$, $\sigma\geq 0$ tel que pour tout $t\in\mathbb{R}$,

$$\mathbb{E}[\exp(itY)]=\exp(it\mu-\frac{\sigma^2t^2}{2}).$$

a) Montrer que, dans ce cas, si $\sigma=0$, la loi de $Y$ est la masse de Dirac en $\mu$ et que si $\sigma\neq 0$, la loi de $Y$ est $\mathcal{N}(\mu,\sigma^2)$.

b) On rappelle que, d'après le cours, un vecteur aléatoire $Z=(Z_1,\dots,Z_n)$ est dit gaussien s'il existe $m\in\mathbb{R}^n$ et $C$ une matrice $n\times n$ symétrique semi-définie positive tels que pour tout $x\in\mathbb{R}^n$,

$$\mathbb{E}[\exp(ix^*Z)]=\exp(ix^*m-\frac{x^*Cx}{2}).$$

Soit $X=(X_1,\dots,X_n)$ un vecteur aléatoire réel tel que pour tous $\lambda=(\lambda_1,\dots,\lambda_n)\in\mathbb{R}^n$, la variable aléatoire $\lambda^*X=\sum_i\lambda_iX_i$ est une variable aléatoire réelle gaussienne. Montrer que $X$ est un vecteur gaussien.

c) En déduire que l'image, par une application linéaire, d'un vecteur gaussien est un vecteur gaussien.

**6. Un contre-exemple utile.** Soient $X,Y$ des va réelles indépendantes de lois respectives $\mathcal{N}(0,1)$ et $p\delta_1+(1-p)\delta_{-1}$, où $p\in[0,1]$. Soit $Z=XY$.

1. Quelle est la loi de $Z$ ? Pour quelles valeurs de $p$ le couple $(X,Z)$ est-il gaussien (on considérera $X+Z$ et $X-Z$) ?
2. Pour quelles valeurs de $p$ les variables aléatoires $X$ et $Z$ sont-elles indépendantes ? Décorrelées ?

**7. Un autre contre-exemple.** Soit $X\sim\mathcal{N}(0,1)$, $a>0$ et

$$Y:=X\mathbb{1}_{|X|\leq a}-X\mathbb{1}_{|X|>a}.$$

1. Quelle est la loi de $Y$ ?
2. Le couple $(X,Y)$ est-il gaussien ?
3. Montrer qu'il existe $a>0$ tel que

$$\frac{1}{\sqrt{2\pi}}\int_0^a t^2e^{-\frac{t^2}{2}}\,\mathrm{d}t=\frac{1}{4}.$$

Que vaut alors $\frac{1}{\sqrt{2\pi}}\int_a^{+\infty}t^2e^{-\frac{t^2}{2}}\,\mathrm{d}t$ ?

4. Pour cette valeur de $a$, calculer la covariance de $X$ et $Y$. Ces deux vecteurs sont-ils indépendants ? Doit-on y voir une contradiction ?

**8. Indépendance gaussienne : un exemple simple.** Soit $(X,Y)$ un vecteur gaussien centré de matrice de covariance

$$K=\begin{bmatrix}1&\rho\\ \rho&1\end{bmatrix},$$

où $\rho\in[0,1]$. Montrer que $A:=X+Y$ et $B:=X-Y$ sont des v.a. indépendantes, donner leur lois.

**9. Vecteurs gaussiens et combinaisons linéaires.** On rappelle qu'un vecteur gaussien admet une densité par rapport à la mesure de Lebesgue si et seulement si sa matrice de covariance est inversible. Soit $X=(X_1,X_2,X_3)$ un vecteur gaussien de $\mathbb{R}^3$ d'espérance $m=(1,0,2)$ et de covariance

$$K=\begin{bmatrix}1&1&0\\1&3&4\\0&4&2\end{bmatrix}.$$

a) $X$ possède-t-il une densité par rapport à la mesure de Lebesgue ?
b) Quels sont les couples $i, j$ tels que $X_i$ et $X_j$ sont indépendants. $X$ a-t-il une composante indépendante du vecteur formé par les deux autres ?
c) Quelle est la loi de $Y = X_1 + 2X_2 + 3X_3$ ?

**10. Gaussiennes : indépendance et projection orthogonale.** Soient $X, Y$ 2 v.a. indep. de loi $\mathcal{N}(0, 1)$. Pour $a > 0$, on définit $Z_a = X + aY$ et on note $\mathcal{H}_a = \mathrm{Vect}(Z_a)$ et $P_{\mathcal{H}_a}$ la projection orthogonale sur $\mathcal{H}_a$. On fixe $\lambda > 0$.
a) Calculer $P_{\mathcal{H}_a}(X)$.
b) Calculer l'espérance conditionnelle $\mathbb{E}[X \,|\, \mathcal{H}_a]$.
c) Calculer l'espérance conditionnelle $\mathbb{E}[e^{\lambda X} \,|\, \mathcal{H}_a]$.
d) Calculer $P_{\mathcal{H}_a}(e^{\lambda X})$.

**11. Conditionnement gaussien.** Soit $(X, Y)$ un vecteur gaussien centré de matrice de convariance

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$

Pourquoi cette matrice est-elle la matrice de covariance d'un vecteur gaussien ? Donner $E(X|Y)$. On note, pour $m \in \mathbb{R}$, $\sigma^2 > 0$, $A_{m,\sigma^2}$ la loi image de la loi $\mathcal{N}(m, \sigma^2)$ par la fonction arctan. Donner la loi conditionnelle de $\arctan(X)$ sachant $Y$ (on ne demande pas d'expliciter la formule donnée, on pourra donner le résultat en fonction d'une loi $A_{m,\sigma^2}$).

**12. Filtre de Kalman Bucy : un premier exemple d'approximation.** Soit $X$ une v.a. d'espérance $m$ et de variance $\sigma^2$.

Donner l'approximation de $X$ par une constante qui soit la meilleure pour la moyenne quadratique (i.e. pour la norme $L^2$).

**13. Filtre de Kalman Bucy : approximation en dimension 2.** Soit $(X, Y)$ un vecteur gaussien centré de matrice de covariance

$$K = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix},$$

avec $\sigma_Y^2 > 0$.

Donner la meilleure approximation (i.e. l'espérance) de $X$ sachant $Y$ et donner la variance de l'erreur commise.

**14. Filtre de Kalman Bucy sur un exemple simple : le cas de la cible constante. I.** On considère le problème très simple suivant : on veut estimer une constante $m$ à partir de plusieurs mesures de celle ci. L'idée est, après chaque mesure, de donner une nouvelle approximation de $m$ construite à partir de la précédente approximation (sauf s'il s'agit de la première mesure !) et de la nouvelle mesure. Les mesures sont supposées toutes sans biais (elles ont toutes pour moyenne $m$), mais de variances différentes, connues (l'écart quadratique moyen de chaque mesure autours de $m$ lui est propre, et il est supposé connu). On suppose donc connaître l'erreur quadratique moyenne commise dans la première approximation (puisque ce sera tout simplement la première mesure), et notre protocole nous permettra de donner celle commise dans la nouvelle approximation.

La mise en place d'un tel protocole implique donc deux étapes :
(i) le choix d'une approximation initiale de $m$ : on prendra naturellement le résultat de la première mesure,
(ii) la définition, étant données une approximation $\hat{X}_n$ de $m$ et une nouvelle mesure $Y_{n+1}$ de $m$, d'une nouvelle approximation $\hat{X}_{n+1}$ de $m$.

Afin d'alléger les notations, nous supprimons les indices. On a donc $m \in \mathbb{R}$, une approximation $\hat{X}$ ($= \hat{X}_n$ dans le texte précédent) de $m$ de la forme

$$\hat{X} = m + W,$$

avec $W$ centrée de variance $\sigma_W^2$ connue, et enfin une nouvelle mesure ($= Y_{n+1}$ dans le texte précédent)

$$Y = m + V,$$

avec $V$ centrée de variance $\sigma_V^2$ connue, indépendante de $W$.

On cherche une nouvelle approximation $\hat{X}'$ ($= \hat{X}_{n+1}$ dans le texte précédent) de $m$ de la forme

$$\hat{X}' = \hat{X} + g(Y - \hat{X}),$$

avec $g \in \mathbb{R}$. On va chercher $g$ de façon à ce que

$$\mathbb{E}[(\hat{X}' - m)^2]$$

soit minimal.

a) Calculer, en fonction de $g$, $\mathbb{E}[(\hat{X}' - m)^2]$.

b) Déterminer la valeur optimale $g^*$, et exprimer alors $\hat{X}'$ et la nouvelle erreur quadratique moyenne $\mathbb{E}[(\hat{X}' - m)^2]$ en fonction des données connues (i.e. $\sigma_W^2$, $\sigma_V^2$, $\hat{X}$ et $Y$).

**II.** Traduisons les considérations simples de la partie I en un problème de filtre de Kalman. On considère le système suivant :

$$\begin{cases} X_{n+1} = X_n = \cdots = X_1 = m \in \mathbb{R} & \text{(équation d'évolution d'état, cible fixe)} \\ Y_n = b_n X_n + W_n & \text{(équation d'observation)} \end{cases}$$

où $X_1$ est une variable aléatoire gaussienne, les $W_i$ sont des variables aléatoires indépendantes, indépendantes de $X_1$, de loi $\mathcal{N}(0, \sigma_i^2)$. Les $b_i$ devraient valoir 1 par analogie avec la partie I, nous les laissons arbitraires pour permettre une analogie plus forte avec la théorie générale. Par ailleurs, comme ils sont connus, quitte à diviser les mesures, on pourrait supposer qu'ils valent 1. Cela dit, en dimension supérieure, ils pourraient être remplacés par des matrices non inversibles : ils exprimeraient alors le caractère partiel des mesures.

a) On pose $\hat{X}_1 = \frac{Y_1}{b_1}$. Calculer la variance $P_1$ de l'erreur $e_1 = X_1 - \hat{X}_1$.

b) Supposons connue $\hat{X}_n$ l'estimation de $X_n$ à la fin de l'étape $n$. On suppose également connue la variance $P_n = \mathrm{Var}(e_n)$, avec $e_n = X_n - \hat{X}_n$. A l'étape $n+1$, on dispose d'une nouvelle observation : $y_{n+1}$. On pose alors, pour tout $g \in \mathbb{R}$,

$$\widetilde{X}_{n+1}^g = \hat{X}_n + g[Y_{n+1} - b_{n+1}\hat{X}_n],$$

Déterminer $g^* = g_{n+1} \in \mathbb{R}$ qui minimise $\mathbb{E}[(X_n - \widetilde{X}_{n+1}^g)^2]$.

c) On pose alors $\hat{X}_{n+1} = \widetilde{X}_{n+1}^{g_{n+1}}$. Exprimer $\hat{X}_{n+1}$, $e_{n+1}$ et $P_{n+1} = \mathrm{Var}(e_{n+1})$ en fonction des données du problème (comme $b_{n+1}$) et des informations connues depuis l'étape $n$ (comme $\hat{X}_n, P_n$).

d) On suppose maintenant que pour tout $i$, $b_i = 1$ et que pour tout $i \geq 2$, $\sigma_i = 1$. Que valent alors $\hat{X}_1$ et $P_1$ ? Montrer que pour tout $n \geq 1$,

$$g_{n+1} = \frac{\sigma_1^2}{1 + \sigma_1^2} = P_{n+1}$$

et que

$$\hat{X}_{n+1} = \frac{1}{1 + n\sigma_1^2}(Y_1 + \sigma_1^2 \sum_{k=1}^{n+1} Y_k).$$