Contrôle des connaissances du cours «Bruits et Signaux» de l'Ecole Doctorale d'Astrophysique.

Énoncé et Corrigé

Le mardi 16 décembre 2003 de 10h à 12h amphi de l'IAP à Paris.

Le contrôle est noté sur 20, le barème est indiqué à la fin des problèmes. Le total des points est de 24 points. Le polycopié et les notes de cours sont les seuls documents autorisés. En règle générale on ne demande pas de calculs compliqués mais un peu de raisonnement.

---

1) La méthode des multiplicateurs de Lagrange : rappels et exercice.

La méthode des multiplicateurs de Lagrange (1788) est une méthode d'optimisation permettant de déterminer l'extremum local d'une fonction f sous contrainte que les variables de f sont liées par une relation g. A deux dimensions, par exemple, cette méthode permettrait de trouver les valeurs x et y où $f(x,y)$ est extremum sous contrainte que $g(x,y) = 0$. La fonction f est la *fonction objectif* et g est la *contrainte*, un point (x,y) tel que $g(x,y) = 0$ est un *point admissible*.

Pour résoudre le problème, on forme la fonction auxiliaire L appelée *lagrangien* :

$$L(x,y,\lambda) = f(x,y) + \lambda g(x,y)\,, \tag{1}$$

où λ est une variable inconnue appelée le multiplicateur de Lagrange.

Sous réserve que toutes les fonctions f et g sont continûment différentiables, on trouve les points stationnaires du problème comme solution du système obtenu en annulant les dérivées partielles de L par rapport à toutes les variables x, y et λ. Soit :

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} = 0\,, \tag{2}$$

$$\frac{\partial L}{\partial y} = \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} = 0\,, \tag{3}$$

$$\frac{\partial L}{\partial \lambda} = g(x,y) = 0\,. \tag{4}$$

Pour savoir si les points stationnaires correspondent à un maximum, un minimum, un col ou autre chose, il faut examiner les conditions suivantes.

On a un maximum si :

$$\left[\frac{\partial^2 L}{\partial x^2}\right]\left[\frac{\partial^2 L}{\partial y^2}\right] - \left[\frac{\partial^2 L}{\partial x \partial y}\right]^2 > 0\,,$$

$$\left[\frac{\partial^2 L}{\partial x^2}\right] < 0 \quad \text{ou} \quad \left[\frac{\partial^2 L}{\partial y^2}\right] < 0\,.$$

On a un minimum si :

$$\left[\frac{\partial^2 L}{\partial x^2}\right]\left[\frac{\partial^2 L}{\partial y^2}\right] - \left[\frac{\partial^2 L}{\partial x \partial y}\right]^2 > 0\,,$$
$$\left[\frac{\partial^2 L}{\partial x^2}\right] > 0 \quad \text{ou} \quad \left[\frac{\partial^2 L}{\partial y^2}\right] > 0\,.$$

Dans le cas où :

$$\left[\frac{\partial^2 L}{\partial x^2}\right]\left[\frac{\partial^2 L}{\partial y^2}\right] - \left[\frac{\partial^2 L}{\partial x \partial y}\right]^2 \leq 0\,,$$

il faut examiner plus en détail le voisinage de la solution (x, y) afin de déterminer si le point stationnaire correspond à un col ou toute autre forme particulière.

On passe maintenant à un petit exercice d'application.

① Soient la fonction objectif :

$$f(x, y) = 5x^2 + 6y^2 - xy$$

et la contrainte :

$$g(x, y) = x + 2y - 24 = 0\,.$$

Trouver le point stationnaire de f sous contrainte g.

② Préciser s'il s'agit d'un maximum, d'un minimum ou d'autre chose.

Telle qu'elle est présentée ci-dessus la méthode s'avère quelquefois fausse car elle ne tient pas compte de possibles dégénérescences. (Elle a cependant été présentée sous cette forme incomplète pendant un siècle avant que l'on s'aperçoive de cette difficulté.)

③ Pour confirmer ce point, considérons dans $\mathbb{R}^2$ la fonction objective suivante :

$$f(x, y) = x\,,$$

sujette à la contrainte :

$$x^2 + y^2 = 0\,.$$

Quelle est la solution du problème ? N'utilisez pas la méthode de Lagrange !

④ Tentez d'utiliser la méthode de Lagrange et montrez où elle est en défaut.

Pour tenir compte de possibles dégénérescences, il faut introduire un multiplicateur supplémentaire λ_0, poser $L = \lambda_0 f + \lambda g$ et chercher la solution du système :

$$\frac{\partial L}{\partial x} = \lambda_0 \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} = 0\,, \tag{5}$$
$$\frac{\partial L}{\partial y} = \lambda_0 \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} = 0\,, \tag{6}$$
$$g(x, y) = 0\,. \tag{7}$$

Qui est un système de trois équations à quatre inconnues et où les multiplicateurs λ_0 et λ ne sont pas simultanément nuls. Si on sait que $\lambda_0 \neq 0$, on peut alors supposer que $\lambda_0 = 1$ car on peut diviser le système par λ_0 sans changer la solution pour x et y. Si $\lambda_0 = 0$, le système ne fait pas intervenir f ce qui montre que la contrainte n'a rien à voir avec la fonction objective, le système est dégénéré. De façon pratique, il suffit de considérer les deux cas $\lambda_0 = 1$ ou $\lambda_0 = 0$. Le cas $\lambda_0 = 0$ est celui qui avait été négligé à l'origine.

⑤ Montrez que cette nouvelle formulation de la méthode donne bien la solution du système dégénéré précédent.

($2 + 1 + 2 + 1 + 1 = 7$ points.)

A plusieurs dimensions et pour plusieurs contraintes la méthode de Lagrange nous dit que la solution est obtenue au point où le gradient de la fonction objective est dans l'enveloppe linéaire des gradients des contraintes. Plus complètement, pour tenir compte du cas dégénéré, elle dit que le gradient de la fonction objective et les gradients des contraintes sont linéairement dépendants. Ce que l'on peut noter :

$$\nabla L = \sum_{i=0}^{n} \lambda_i \nabla g_i = 0 \,,$$

où l'on a posé $g_0 = f$ la fonction objective et $g_i, i \geq 1$ les contraintes. Les λ_i ne sont pas tous nuls et il n'y a que deux cas à considérer : $\lambda_0 = 1$ ou $\lambda_0 = 0$. On trouve la solution du problème à l'aide de ce système et en satisfaisant les contraintes. Cette solution est, par exemple, un minimum si la matrice des dérivées secondes de L (le hessien) est définie positive.

Corrigé.

❶ On forme le lagrangien :

$$\begin{aligned} L(x, y, \lambda) &= f(x, y) + \lambda g(x, y) \,, \\ &= 5x^2 + 6y^2 - xy + \lambda(x + 2y - 24) \end{aligned}$$

et on annule les dérivées partielles :

$$\begin{aligned} \frac{\partial L}{\partial x} &= 10x - y + \lambda = 0 \,, \\ \frac{\partial L}{\partial y} &= -x + 12y + 2\lambda = 0 \,, \\ \frac{\partial L}{\partial \lambda} &= x + 2y - 24 = 0 \,. \end{aligned}$$

Ce système est un système linéaire de trois équations à trois inconnues. Il a pour solution $x = 6$, $y = 9$ et $\lambda = -51$. Le multiplicateur λ nous importe peu ici, il représente le rapport entre les modules du gradient suivant f et suivant g. La méthode elle-même nous dit que pour la solution du problème ces deux gradients sont colinéaires (au plus haut point d'une route de montagne la ligne de plus grande pente est perpendiculaire à la chaussée).

❷ On a

$$\begin{aligned} \left[\frac{\partial^2 L}{\partial x^2}\right] &= 10 \,, \\ \left[\frac{\partial^2 L}{\partial y^2}\right] &= 12 \,, \\ \left[\frac{\partial^2 L}{\partial x \partial y}\right] &= -1 \,. \end{aligned}$$

De sorte que :

$$\begin{aligned}
\left[\frac{\partial^2 L}{\partial x^2}\right]\left[\frac{\partial^2 L}{\partial y^2}\right] - \left[\frac{\partial^2 L}{\partial x \partial y}\right]^2 &= 10 \cdot 12 - (-1)^2\,, \\
&= 119 > 0\,, \\
\left[\frac{\partial^2 L}{\partial x^2}\right] &> 0\,, \\
\left[\frac{\partial^2 L}{\partial y^2}\right] &> 0\,.
\end{aligned}$$

Il s'agit d'un minimum.

❸ L'unique solution de $x^2 + y^2 = 0$ est $x = y = 0$. (La norme euclidienne d'un vecteur n'est nulle que si le vecteur est nul.)

❹ La méthode de Lagrange donne :

$$\begin{aligned}
1 + 2\lambda x &= 0\,, \\
2\lambda y &= 0,\,.
\end{aligned}$$

Le multiplicateur λ n'est pas nul car on aurait alors $1 = 0$. La première équation est alors en contradiction avec $x^2 + y^2 = 0$ qui implique $x = 0$ et donc aussi $1 = 0$. La méthode est en défaut.

❺ A présent on a :

$$\begin{aligned}
\lambda_0 + 2\lambda x &= 0\,, \\
2\lambda y &= 0,\,.
\end{aligned}$$

Le cas $\lambda_0 = 1$ conduit à la contradiction précédente alors que le cas $\lambda_0 = 0$ conduit à la solution $x = y = 0$ (si $\lambda_0 = 0$ alors $\lambda \neq 0$ car ils ne sont pas tous les deux nuls).

2) Un problème d'ajustement avec contraintes

Un expérimentateur désire analyser un spectre y à l'aide d'un modèle μ. Disons qu'il s'agit du spectre d'un amas stellaire allant du bleu au rouge. Comme d'habitude, la partie bleue contient l'information sur le continu et la partie rouge celle sur les étoiles présentes dans l'amas. Cette affirmation doit être nuancée mais cela ne nous concerne pas pour l'instant.

Pour analyser son spectre, l'expérimentateur utilise un programme basé sur la méthode des moindres-carrés. Le problème c'est que l'énergie dans le continu est beaucoup plus grande que celle contenue dans les raies d'absorption, or ce sont les raies d'absorption qui sont révélatrices de la présence d'étoiles. Par conséquent, le programme aura tendance à bien ajuster le continu au détriment des raies et à sous-estimer la population d'étoiles. Plus spécifiquement il aura tendance à surestimer la population d'étoiles chaudes et à sous-estimer celle des étoiles froides.

Conscient de cette erreur systématique, l'expérimentateur pense à ajuster son spectre sous contrainte que l'erreur commise sur la partie rouge ne soit pas plus grande qu'une valeur jugée tolérable. Nous allons étudier le principe de la mise en œuvre de cette méthode. On ne demandera pas de montrer que les solutions trouvées sont bien des minimums.

① Les données sont sous forme d'un tableau (ou vecteur) y constitué de n nombres $(y_1, y_2, \dots, y_n)$ qui représentent des flux pour les longueurs d'ondes $(l_1, l_2, \dots, l_n)$. Ces données sont entachées d'erreurs $(\sigma_1, \sigma_2, \dots, \sigma_n)$. Le modèle dépend de k paramètres $(\theta_1, \theta_2, \dots, \theta_k)$ qui, par l'intermédiaire de l'ajustement, délivrent les nombres $(\mu_1, \mu_2, \dots, \mu_n)$. Ces nombres sont les flux estimés aux longueurs d'ondes considérées.

Pour une certaine valeur des k paramètres θ, l'expérimentateur juge de l'adéquation du modèle aux données en l_i à l'aide de la variable réduite :

$$z_i = \frac{y_i - \mu_i}{\sigma_i}\,.$$

Justifiez l'introduction de l'erreur σ_i dans le calcul de z_i ?

② L'expérimentateur juge de l'adéquation globale du modèle sur l'ensemble de toutes les longueurs d'ondes l_i en construisant un χ^2. Donnez l'expression de ce χ^2 ?

③ A quelle condition ce χ^2 suit-il effectivement une statistique du χ^2 ?

On sait qu'il est toujours possible de faire un changement de variables où, suivant les nouvelles variables, les erreurs σ_i sont toutes égales à un. C'est pourquoi nous allons supposer à partir de maintenant que les données y_i sont d'égales précision σ c'est-à-dire que $\forall i, \sigma_i = \sigma = 1$.

De plus on suppose que le modèle dépent linéairement des paramètres θ et que l'on a : $\mu = X\theta$. Dans cette expression μ est un vecteur colonne formé de n éléments, θ un vecteur colonne formé de k éléments et X une matrice rectangulaire formée de n lignes et k colonnes.

④ Le principe des moindres-carrés consiste à chercher les θ qui minimisent le carré de la norme euclidienne entre les vecteurs y et μ. Donnez l'expression de cette norme en fonction de y, X et θ. Pour être concis vous pouvez employer une notation matricielle. On notera toujours χ^2 le carré de cette norme.

⑤ Montrer que les θ qui minimisent χ^2 sont solutions d'un système linéaire (les équations normales).

⑥ On s'intéresse maintenant seulement à la partie rouge du spectre. Soient y_R la partie du spectre qui correspond au rouge et μ_R le modèle correspondant. Ces vecteurs colonnes sont formés de $m < n$ valeurs avec $m > k$.

Comment est construite la matrice X_R qui permet de calculer μ_R à partir des θ ?

Donnez l'équivalent des équations normales mais seulement pour la partie rouge. En d'autres termes donnez le système linéaire qui permet de determiner les θ assurant le meilleur ajustement de la seule partie rouge au sens des moindres-carrés.

On note χ^2_R le carré de la norme euclidienne entre les vecteurs y_R et μ_R. On cherche maintenant à minimiser χ^2 (c'est-à-dire l'erreur sur tout le spectre) sous contrainte que χ^2_R (c'est-à-dire l'erreur sur la partie rouge du spectre seulement) ne soit pas supérieure à un seuil α^2 ($\alpha > 0$). Le seuil α^2 est choisi de telle sorte qu'il ne soit pas inférieur à l'erreur minimum d'ajustement calculée sur la seule partie rouge $\alpha^2 \geq \min_\theta \chi^2_R$, sinon le problème n'aurait pas de solution.

L'ensemble des θ pour lesquels $\chi^2_R \leq \alpha^2$ est le domaine admissible.

⑦ Alors de deux choses l'une :

(a) Soit la solution du problème non contraint appartient au domaine admissible et cette solution est aussi la solution du problème contraint (dans le cas linéaire la solution est unique).

(b) Soit il faut chercher la solution sur le bord du domaine admissible et chercher à résoudre le problème : minimiser χ^2 sous contrainte que $\chi^2_R = \alpha^2$.

A partir de maintenant, on se place dans le second cas (b). Quelle est alors la forme de la fonction lagrangienne ? On admettra que le problème n'est pas dégénéré. Si le problème était dégénéré cela voudrait dire que la solution du problème contraint ne dépendrait pas de la partie bleue ce qui serait physiquement inacceptable.

⑧ Donner le système d'équations, si possible sous forme matricielle, qui permet de résoudre le problème contraint.

⑨ On constate que, si le multiplicateur de Lagrange était connu, le système permettant de calculer les θ serait linéaire. L'idée consiste alors à résoudre ce système en considérant λ comme paramètre et de trouver $\theta = \theta(\lambda)$

Quelle est l'équation qui permet de déterminer λ ? (On ne demande pas de résoudre le système linéaire, mais d'utiliser sa solution $\theta(\lambda)$ afin de déterminer λ.)

⑩ Cette équation porte le nom d'*équation séculaire*, on la note ϕ. Trouver λ consiste à résoudre $\phi(\lambda) = 0$.

Sans faire de calculs compliqués, donner un argument qui montre que λ doit être positif.

Un calcul facile, mais un peu long, montre que ϕ est monotone décroissante pour $\lambda > 0$. Montrer alors que $\phi(0) > 0$ et qu'il existe un λ positif tel que $\phi(\lambda) \leq 0$ (ainsi λ peut être trouvé facilement à l'aide d'une méthode numérique).

Pour se fixer les idées, on peut s'aider du graphique ci-joint.

($1 + 1 + 2 + 1 + 1 + (1 + 1) + 1 + 3 + 1 + (2 + 1 + 1) = 17$ points.)

Corrigé.

❶ Il y a plusieurs réponses possibles, la plus simple est de dire que le facteur $1/\sigma_i$ rend compte du poids qu'il faut accorder aux données. Plus l'erreur est grande moins il faut tenir compte de la valeur y_i ce qui justifie une forme telle que $1/\sigma_i$.

❷
$$\chi^2 = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{\sigma_i} \right]^2$$

❸ Si les y_i sont indépendants et suivent une loi normale de moyenne μ_i et de variance σ_i^2.

❹ C'est une simple question de cours, on a :

$$\chi^2 = ||y - X\theta||_2^2 = (y - X\theta)^t (y - X\theta) \, .$$

La notation ^{t} désigne la transposition.

❺ En suivant là aussi le calcul du cours on obtient les équations normales :

$$X^t X\theta = X^t y \, .$$

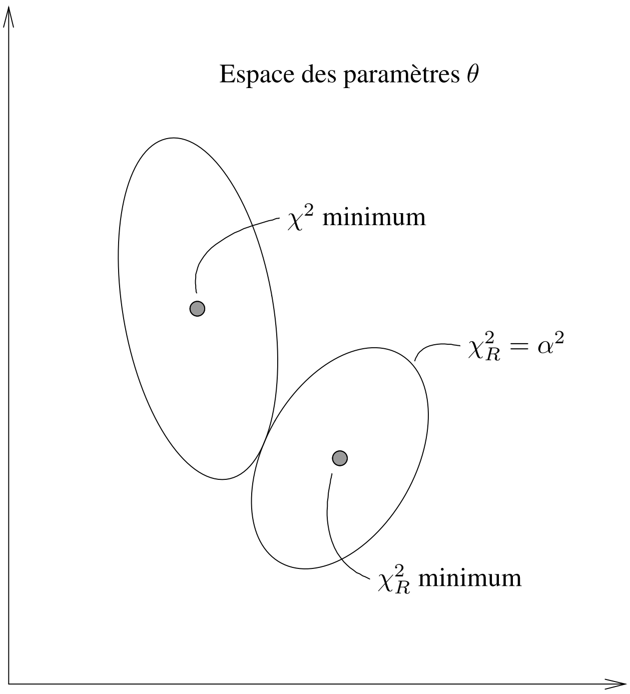

FIG. 1 – Cette figure illustre le problème dans le cas (b) où $\min \chi^2 > \alpha^2$. On se place dans l'espace qui nous intéresse c'est-à-dire celui des paramètres inconnus θ. On a porté un point représentant la solution non contrainte : « χ^2 minimum » et la solution qui minimise le χ^2 dans la seule partie rouge : « χ^2_R minimum ». On a représenté le domaine admissible autour de $\min \chi^2_R$, sa frontière est l'ellipse $\chi^2_R = \alpha^2$. Autour de $\min \chi^2$ on a tracé l'ellipse qui correspond au χ^2 de la solution contrainte. La solution contrainte est au point de contact de ces deux ellipses. Les ellipses sont inclinées car les formes quadratiques χ^2 et χ^2_R, en fonction des θ, présentent des termes croisés du genre $\theta_i\theta_j$.

❻ Les lignes de la matrice X_R sont celles de X qui correspondent aux indices i tels que l_i est dans la partie rouge qui nous intéresse. En pratique c'est un bloc inférieur de X en admettant que les l_i sont rangés par ordre croissant des longueurs d'ondes.

Pour la seule partie rouge les équations normales sont :

$$X_R^t X_R \theta = X_R^t y_R\,.$$

❼ $L = \chi^2 + \lambda(\chi^2_R - \alpha^2)$, soit sous forme matricielle :

$$L = (y - X\theta)^t(y - X\theta) + \lambda[(y_R - X_R\theta)^t(y_R - X_R\theta) - \alpha^2],\,.$$

❽ En suivant encore le calcul fait sur le cours qui établit les équations normales, on trouve :

$$(X^tX + \lambda X_R^t X_R)\theta = X^t y + \lambda X_R^t y_R\,.$$

Ce qui avec la contrainte $\chi^2_R = \alpha^2$ permet de résoudre le problème.

❾ C'est la contrainte : $\phi(\lambda) = \chi^2_R - \alpha^2 = 0$. Sous une forme plus explicite :

$$\phi(\lambda) = (y_R - X_R\theta(\lambda))^t(y_R - X_R\theta(\lambda)) - \alpha^2\,.$$

❿ Si λ est positif cela veut dire que pour la solution, les gradients sont colinéaires mais opposées (voir figure). Si ce n'était pas le cas, le gradient suivant χ^2 pointerait dans la même direction que le gradient suivant χ^2_R. Par ailleurs $-\nabla\chi^2$ indique la direction où χ^2 diminue qui serait aussi celle où χ^2_R diminue. Il y aurait alors une meilleure solution à l'intérieur du domaine admissible ce qui n'est pas possible puisque, par hypothèse, nous sommes sur la solution.

On a $\nabla L = \nabla\chi^2 + \lambda\nabla\chi^2_R = 0$. Si $\lambda = 0$ la solution a lieu pour $\nabla\chi^2 = 0$ c'est-à-dire pour le problème non contraint. Nous nous sommes placé dans le cas ou la solution du problème non contraint était en dehors du domaine admissible pour laquelle on a $\chi^2_R > \alpha^2$, donc $\phi(0) > 0$.

Lorsque λ augmente, le poids de χ^2 dans L diminue et la solution tend vers celle qui rend χ^2_R minimum, c'est-à-dire qu'elle se dirige vers l'intérieur du domaine admissible. A l'intérieur de ce domaine $\chi^2_R \leq \alpha^2$ et donc $\phi(\lambda) \leq 0$.