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

Le mardi 17 décembre 2002 de 10h à 12h amphi du LAM à Meudon.

Le contrôle est noté sur 20, le barème est indiqué à la fin du problème. Le total des points est de 22 points. Les questions sont relativement indépendantes. Le polycopié et les notes de cours sont les seuls documents autorisés.

Un programme d'ajustement.

Le but de ce problème est de vous mettre en position de pouvoir écrire un programme d'ajustement de raies spectrales dans un spectre observé. On traite ici un cas simplifié mais il possède tous les ingrédients d'un cas réel.

On suppose que le spectre observé est constitué de N mesures (pixels) et que la zone concernée peut être assimilée à un continu de hauteur inconnue et d'une raie d'émission de position, largeur et amplitude inconnues. Soient : θ_0 la hauteur du continu, θ_1 la position de la raie, θ_2 la largeur de la raie et θ_3 son amplitude. On suppose de plus que le profil de la raie spectrale est une gaussienne de telle sorte que le spectre théorique peut s'écrire :

$$f(\nu) = \theta_0 + \theta_3 \exp\left\{ -\left(\frac{\nu - \theta_1}{\theta_2} \right)^2 \right\},$$

où f désigne l'énergie spécifique moyenne reçue par unité de temps et par unité de fréquence (en $\mathrm{ergs/s/Hz}$). Le domaine spectral est $[\nu_1, \nu_N]$, il est petit par rapport aux fréquences de cet intervalle.

① On s'intéresse au nombre de photons moyen μ_i reçus dans l'intervalle spectral $[\nu_i, \nu_i + \Delta\nu]$. Ce nombre de photons lorsque $\Delta\nu$ est très petit devant les ν_i est donné par :

$$\mu_i = \frac{T}{h} \int_{\nu_i}^{\nu_i + \Delta\nu} \frac{f(u)}{u}\, du \approx \frac{T\Delta\nu}{h\nu_i} f(\nu_i)\,.$$

Puisque l'intervalle spectral $[\nu_1, \nu_N]$ est petit on remplacera ν_i par un ν_0 caractéristique de cet intervalle. (On ne pourrait pas faire ce genre d'approximation si l'analyse portait sur des données obtenues à très grand rapport signal-sur-bruit.) Finalement, dans le but de ne pas alourdir l'écriture, on supposera que les unités de mesure ont été choisies de telle sorte que $\frac{T\Delta\nu}{h\nu_0} = 1$.

On suppose que les photons proviennent d'un seul émetteur poissonien (en fait la superposition de plusieurs), quelle est la probabilité d'observer n_i photons dans l'intervalle $[\nu_i, \nu_i + \Delta\nu]$ pendant le temps T ?

② Les données expérimentales se présentent donc comme un vecteur de variables aléatoires $\{n_i, i = 1, N\}$, que l'on suppose indépendantes. L'expérimentateur désire ajuster ses données à l'aide d'une méthode des moindres-carrés pondérés, c'est-à-dire du moindre χ^2. On rappelle que le χ^2 est la somme des carrés des variables réduites. Donnez l'expression du χ^2 en fonction de n_i et μ_i.

③ Dans l'expression du χ^2, l'écart entre les données expérimentales et la moyenne attendue est pondérée par l'inverse de la variance des données. Pour ce qui nous intéresse, ce facteur complique un peu les choses car les variances sont inconnues. L'expérimentateur pense alors à approximer la variance des données par une expression qui dépend soit de n_i, soit de $\widehat{\mu}_i$, où les $\widehat{\mu}_i$ désignent une estimation des μ_i. Pouvez-vous donner ces expressions et discuter brièvement leurs avantages et leurs inconvénients ?

④ L'une des deux approximations étant admise, se trouve-t-on maintenant dans le cadre d'une estimation des $\theta_0, \ldots, \theta_3$ suivant la méthode des moindres-carrés pondérés ?

Pour simplifier encore plus avant le problème, l'expérimentateur choisit la pondération qui ne fait pas intervenir les paramètres inconnus. Sommes-nous alors en présence d'une méthode des moindres-carrés linéaire ?

⑤ On note $\chi^2(\theta)$ l'expression du χ^2 en fonction des paramètres inconnus (le symbole θ remplace les quatre paramètres inconnus). Une méthode itérative du genre de la méthode de Newton est envisagée afin de déterminer le lieu du minimum de χ^2 dans l'espace des θ. Cette méthode a besoin d'un point de départ afin d'amorcer les itérations, que proposez-vous comme première estimation des θ_k ? On demande de simplement poser les équations, pas de les résoudre.

⑥ Afin d'améliorer l'estimation, on développe le χ^2 autour de l'estimation actuelle $\theta^{(0)}$. Suivant le développement de Taylor, on a :

$$\chi^2(\theta^{(1)}) = \chi^2(\theta^{(0)}) + g^t(\theta^{(1)} - \theta^{(0)}) + \tfrac{1}{2}(\theta^{(1)} - \theta^{(0)})^t H(\theta^{(1)} - \theta^{(0)}) + \ldots$$

Ce qui peut encore s'écrire :

$$\Delta\chi^2 \approx g^t\Delta\theta + \tfrac{1}{2}\Delta\theta^t H\Delta\theta\,. \tag{1}$$

Dans ces expressions, par exemple g^t désigne la transposée du vecteur des gradients de χ^2 et H la matrice de ses dérivées secondes (le Hessien).

Donnez l'expression des composantes de vecteur g. Si vous peinez un peu dans les calculs, passez aux questions suivantes qui ne dépendent pas du résultat explicite demandé ici.

⑦ L'expérimentateur estime que l'approximation de χ^2 au second ordre est excellente, c'est-à-dire que χ^2 ne dépend des puissances des θ_k que jusqu'à l'ordre deux, les autres puissances pouvant être tenues pour négligeables. Sur la remarque que si tel est bien le cas alors H doit être constante, l'expérimentateur a alors l'idée d'écrire le développement de Taylor depuis le minimum supposé être en $\theta^{(1)}$, il obtient :

$$\chi^2(\theta^{(0)}) = \chi^2(\theta^{(1)}) + \text{terme linéaire} + \text{terme quadratique} + \ldots$$

Donnez l'expression du terme linéaire et du terme quadratique, on ne demande pas de calculer H.

⑧ En déduire le pas $\Delta\theta$ qu'il faut effectuer depuis $\theta^{(0)}$ pour atteindre le minimum.

⑨ Si le minimum n'est pas atteint en une itération, à cause de la présence de termes d'ordres supérieurs à deux (faibles certes mais non nuls), l'expérimentateur envisage alors de poursuivre les itérations à partir de $\theta^{(1)}$ et de construire la suite $\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(n)}, \dots$ qui doit bien finir par converger vers le minimum.

On remarque que cette méthode n'est efficace que si H peut toujours être considéré comme étant définie positive, sinon la limite des $\theta^{(n)}$ peut correspondre à un maximum, un col ou ne pas converger. L'expérimentateur envisage alors, en cas d'échec, de se tourner vers une autre méthode. Il pense linéariser le modèle autour de $\theta^{(0)}$ et de calculer la fonction χ^2 à l'aide de cette approximation. On note σ_i^{-2} la pondération des termes du χ^2 et on suppose que l'expérimentateur a choisi celle qui ne dépend pas des θ. Dans ces conditions, la fonction χ^2 est-elle quadratique ?

Si γ désigne le gradient du modèle par rapport aux θ, que vaut la matrice H ?

⑩ Finalement l'expérimentateur calcule les θ suivant la méthode des moindres-carrés ou, plus précisément, il calcule la correction $\Delta\theta$ à apporter à $\theta^{(0)}$ suivant cette méthode.

Donnez alors l'expression de $\Delta\theta$.

Comme dans la méthode précédente, on itère ce schéma si l'approximation linéaire n'est pas très bonne. Si $\theta^{(0)}$ est très éloigné du minimum du χ^2, la méthode peut ne pas converger mais est-on certain cette fois-ci que $\theta^{(1)}$ correspond au minimum de l'approximation locale et non pas à un maximum ou à un col ?

($1 + 1 + 2 + (1 + 1) + 2 + 4 + 2 + 2 + (1 + 2) + (2 + 1) = 22$ points.)

Corrigé.

❶ C'est une simple question de cours, seules les notations changent un peu :

$$\Pr\{n = n_i\} = \frac{\mu_i^{n_i}}{n_i!} e^{-\mu_i} .$$

❷ La forme générale du χ^2 est :

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \mu_i)^2}{\sigma_i^2} ,$$

où μ_i est la moyenne et σ_i^2 la variance de n_i. Pour une variable de Poisson la variance est égale à la moyenne d'où :

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \mu_i)^2}{\mu_i} .$$

❸ On peut songer à approximer μ_i par n_i ou par $\widehat{\mu}_i$.

Si on pose $\mu_i \approx n_i$ on court le risque de surévaluer la pondération de $n_i - \mu_i$ lorsque $n_i \approx 0$, c'est-à-dire de donner trop d'importance aux parties qui ne contiennent pas de signal.

Si on pose $\mu_i \approx \widehat{\mu}_i$, on a alors un χ^2 dont la dépendance en θ est fort complexe. Une recherche du minimum de χ^2 peut, par exemple, conduire à annuler le dénominateur et générer des instabilités numériques.

❹ Oui car le modèle est maintenant additif.
Non car le modèle n'est pas linéaire en θ.

❺ On peut suggérer la méthode des moments, ce qui suppose de calculer les moments de la distribution des n_i jusqu'à l'ordre quatre, on obtiendrait :

$$
\begin{aligned}
\int f(\nu)\,d\nu &= \sum_{i=1}^{N} n_i\,, \\
\int \nu f(\nu)\,d\nu &= \sum_{i=1}^{N} \nu_i n_i\,, \\
\int \nu^2 f(\nu)\,d\nu &= \sum_{i=1}^{N} \nu_i^2 n_i\,, \\
\int \nu^3 f(\nu)\,d\nu &= \sum_{i=1}^{N} \nu_i^3 n_i\,.
\end{aligned}
$$

Ce qui conduit au système non-linéaire de quatre équations à quatre inconnues :

$$
\begin{aligned}
(\nu_N - \nu_1)\theta_0 + \sqrt{\pi}\theta_2\theta_3 &= \sum_{i=1}^{N} n_i \\
\frac{1}{2}(\nu_N^2 - \nu_1^2)\theta_0 + \sqrt{\pi}\theta_1\theta_2\theta_3 &= \sum_{i=1}^{N} \nu_i n_i\,, \\
\frac{1}{3}(\nu_N^3 - \nu_1^3)\theta_0 + \sqrt{\pi}\theta_2\theta_3(\tfrac{1}{2}\theta_2^2 + \theta_1^2) &= \sum_{i=1}^{N} \nu_i^2 n_i\,, \\
\frac{1}{4}(\nu_N^4 - \nu_1^4)\theta_0 + \sqrt{\pi}\theta_2\theta_3(\tfrac{3}{2}\theta_1\theta_2^2 + \theta_1^3) &= \sum_{i=1}^{N} \nu_i^3 n_i\,.
\end{aligned}
$$

Système qu'il n'est pas simple de résoudre analytiquement sauf si on utilise une autre méthode pour déterminer θ_0, les autres paramètres suivent alors. Pour estimer θ_0 on peut, par exemple, prendre la moyenne arithmétique de la moitié des points de plus faible amplitude si la raie est en émission ou de plus grande amplitude si elle est en absorption.

❻

$$
\begin{aligned}
\frac{\partial\chi^2}{\partial\theta_0} &= -2\sum_{i=1}^{N}\frac{n_i-\mu_i}{n_i}\,,\\
\frac{\partial\chi^2}{\partial\theta_1} &= -4\sum_{i=1}^{N}\frac{n_i-\mu_i}{n_i}\theta_3\exp\left\{-\left(\frac{\nu_i-\theta_1}{\theta_2}\right)^2\right\}\frac{\nu_i-\theta_1}{\theta_2^2}\,,\\
\frac{\partial\chi^2}{\partial\theta_2} &= -4\sum_{i=1}^{N}\frac{n_i-\mu_i}{n_i}\theta_3\exp\left\{-\left(\frac{\nu_i-\theta_1}{\theta_2}\right)^2\right\}\frac{(\nu_i-\theta_1)^2}{\theta_2^3}\,,\\
\frac{\partial\chi^2}{\partial\theta_3} &= -2\sum_{i=1}^{N}\frac{n_i-\mu_i}{n_i}\exp\left\{-\left(\frac{\nu_i-\theta_1}{\theta_2}\right)^2\right\}
\end{aligned}
$$

❼ Le terme linéaire est nul car au minimum le gradient est nul.
Le terme quadratique est identique car H est constante et $(-\Delta\theta)^t H(-\Delta\theta) = \Delta\theta^t H\Delta\theta$.

❽ On a les deux équations :

$$
\begin{aligned}
\Delta\chi^2 &= g^t\Delta\theta + \tfrac{1}{2}\Delta\theta^t H\Delta\theta\,,\\
-\Delta\chi^2 &= \qquad\quad + \tfrac{1}{2}\Delta\theta^t H\Delta\theta\,,
\end{aligned}
$$

d'où :

$$(g^t + \Delta\theta^t H)\Delta\theta = 0$$

et $g^t + \Delta\theta^t H = 0$ car $\Delta\theta$ est non nul. Finalement $\Delta\theta$ est solution du système linéaire : $H\Delta\theta = -g$, il s'agit bien de la méthode de Newton.

❾ Elle est quadratique comme carré de somme de termes au plus linéaires.
On a :

$$\chi^2 = \sum_{i=1}^{N}\frac{(n_i-\mu_i)^2}{\sigma_i^2}\,,$$

avec $\mu_i = \mu_i(\theta^{(0)}) + \gamma_i\Delta\theta_i$. Le Hessien est le coeffcient des termes quadratiques, c'est-à-dire :

$$H = \gamma^t V^{-1}\gamma\,,$$

ou V est la matrice des variances-covariances des observations. Dans notre cas il s'agit d'une matrice diagonale formée des σ_i^2 qui ont été approximés par les n_i.

❿ On pose $y_i = n_i - \mu_i(\theta^{(0)})$ et y représente le vecteur des y_i. Le méthode des moindres-carrés pondérés conduit à la correction :

$$\Delta\theta = (\gamma^t V^{-1}\gamma)^{-1}\gamma^t V^{-1} y\,.$$

Oui car la matrice $\gamma^t V^{-1}\gamma$ est, sauf cas dégénéré, définie positive.