ENC Paris-Saclay
L3 SAPHIRE
—

Année 2018-2019
Examen
—

Examen - Estimation Paramétrique

L'usage des calculatrices est interdit. Seule une feuille de notes manuscrites A4 est autorisée. Les parties sont indépendantes. Il est conseillé de lire attentivement le sujet avant de commencer. Les réponses devront être soigneusement argumentées. ***Les exercices 1 et 2, les exercices 3 et 4, et l'exercice 5 seront à rédiger sur 3 feuilles séparées.***

1 Gradient, Newton, et Gauss-Newton

On considère la fonction de Rosenbrock

$$f(\mathbf{x}) = (1 - x_1)^2 + \left(x_2 - x_1^2\right)^2. \tag{1}$$

1. Montrer que le point $\mathbf{x} = (1, 1)^{\mathrm{T}}$ est une argument du minimum global de $f(\mathbf{x})$.

On utilise une méthode de recherche itérative du minimum de $f(\mathbf{x})$. Pour cela, on part du point $\mathbf{x}_0 = (-1, 0)^{\mathrm{T}}$.

2. Méthode du gradient
   (a) Donner la direction de descente $\mathbf{d}_{\mathrm{g}}$ donnée par cette méthode.
   (b) Evaluer $f(\mathbf{x}_1)$ avec $\mathbf{x}_1 = \mathbf{x}_1 + \mathbf{d}_{\mathrm{g}}$.
3. Méthode de Newton.
   (a) Donner la direction de descente $\mathbf{d}_{\mathrm{n}}$ donnée par cette méthode.
   (b) Evaluer $f(\mathbf{x}_1)$ avec $\mathbf{x}_1 = \mathbf{x}_1 + \mathbf{d}_{\mathrm{n}}$.
4. Méthode de Gauss-Newton.
   (a) Donner la direction de descente $\mathbf{d}_{\mathrm{gn}}$ donnée par cette méthode.
   (b) Evaluer $f(\mathbf{x}_1)$ avec $\mathbf{x}_1 = \mathbf{x}_1 + \mathbf{d}_{\mathrm{gn}}$.
5. Conclusion.

2 Estimation au sens des moindres carrés

On considère le système d'équations surdéterminé suivant issu de l'observation d'un système linéaire et invariant dans le temps

$$\begin{cases} 3a & = 5 \\ 4a & = -1 \\ -3a & = 1 \\ 2a & = 3 \end{cases}$$

Donnez l'estimée de a au sens des moindres carrés. Détaillez votre réponse.

3 Choix d'un guichet

Vous entrez dans un bureau de poste comportant deux guichets. On vous a appris qu'au guichet 1, le temps de service (le temps mis par l'agent pour servir un client) est décrit par une variable aléatoire X_1, distribuée suivant une loi exponentielle

$$p_{X_1}(t) = a \exp(-2t)\, u(t),$$

où $u(t)$ est l'échelon unité. Le temps de service au guichet 2 est décrit par une variable aléatoire X_2 distribuée suivant une loi uniforme sur l'intervalle $[0, 3]$. Vous ne trouvez pas les numéros des guichets et choisissez donc l'un des deux guichets au hasard. Soit G la variable aléatoire représentant le numéro du guichet. Soit Y la variable aléatoire décrivant le temps de service au guichet que vous avez choisi. Il s'agit de déterminer à partir de Y si vous avez été au guichet 1 ou au guichet 2.

1. Calculez la valeur de a.
2. Déduisez de l'énoncé la probabilité *a priori* $\Pr(G = 1)$ de choisir le guichet 1. Même question pour $\Pr(G = 2)$ concernant le guichet 2.
3. Sachant que le temps mis par l'agent pour vous servir est $Y = y$, donnez en fonction de y l'expression de la probabilité *a posteriori* que le guichet que vous avez choisi soit le guichet 1. Donnez en fonction de y l'expression de la probabilité *a posteriori* que le guichet que vous avez choisi soit le guichet 2. Représentez ces deux probabilités *a posteriori* en fonction de y.
4. Pour quelles valeurs de y décidez-vous que vous avez choisi le guichet 1 ? Quelle est, en fonction de y, la probabilité d'erreur associée à ce choix ?
5. Pour quelles valeurs de y décidez-vous que vous avez choisi le guichet 2 ? Quelle est, en fonction de y, la probabilité d'erreur associée à ce choix ?
6. Quelle est en moyenne la probabilité d'erreur ?

Pour vous aider dans vos applications numériques, on vous rappelle que $\ln(6) = 1.79$ et que $\exp(-6) = 0.025$.

4 Entrées D-optimales (choix des entrées minimisant l'incertitude d'estimation)

On considère le modèle à temps discret

$$y_{\mathrm{m}}(k, \mathbf{p}) = p_1 u(k) + p_2 u(k-1), \tag{2}$$

où $u(k)$ est l'entrée à l'instant k et où $\mathbf{p} = (p_1, p_2)^T$.

1. Ce modèle est dit à réponse impulsionnelle finie. Expliquer pourquoi.
2. On fait l'hypothèse que les mesures effectuées sur le système satisfont

$$y(k) = y_{\mathrm{m}}(k, \mathbf{p}^*) + \varepsilon(k),\, k = 1, ..., N, \tag{3}$$

où les $\varepsilon(k)$ sont des réalisations de variables aléatoires indépendantes et identiquement distribuées suivant une loi normale de moyenne nulle et de variance connue σ^2. Détailler le calcul de l'estimée $\widehat{\mathbf{p}}$ de $\mathbf{p}$ au sens du maximum de vraisemblance (ou des moindres carrés, si vous ne savez pas faire). On supposera $u(k)$ connu pour $k = 0, \ldots, N$ et nul pour $k < 0$. On montrera en particulier que la vraisemblance s'écrit

$$\pi(\mathbf{y}|\mathbf{p}) = \left(2\pi\sigma^2\right)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{y}-\mathbf{R}\mathbf{p}\right\|^2\right) \tag{4}$$

où $\mathbf{R}$ est une matrice dont il faudra préciser le contenu.

La matrice d'information de Fisher

$$\mathbf{M}\left(\mathbf{p}, u\left(0\right), \ldots, u\left(N\right)\right) = \mathrm{E}\left(\frac{1}{N}\frac{\partial \log \pi\left(\mathbf{y}|\mathbf{p}\right)}{\partial \mathbf{p}}\frac{\partial \log \pi\left(\mathbf{y}|\mathbf{p}\right)}{\partial \mathbf{p}^T}\middle| \mathbf{p}\right), \tag{5}$$

où $\mathrm{E}\left(\cdot\right)$ est l'espérance, permet d'avoir une idée de l'incertitude d'estimation des paramètres. En effet, on peut montrer que pour $N \to \infty$, $N\left(\mathbf{p} - \mathbf{p}^*\right)$ est distribué suivant une loi Gaussienne de moyenne nulle et de matrice de covariance l'inverse $\mathbf{M}^{-1}\left(\mathbf{p}, u\left(0\right), \ldots, u\left(N\right)\right)$ de la matrice d'information de Fisher. Plus la matrice d'information de Fisher sera « grande », et meilleure sera la précision d'estimation de $\mathbf{p}^*$ réalisée par $\widehat{\mathbf{p}}$. Dans le cas du modèle donné par (2) et avec un modèle de mesure donné par (3), (5) devient

$$\mathbf{M}\left(\mathbf{p}, u\left(0\right), \ldots, u\left(N\right)\right) = \frac{1}{\sigma^2}\mathbf{R}^T\mathbf{R}. \tag{6}$$

3. Exprimer les éléments de la matrice d'information de Fisher en fonction des valeurs successives de l'entrée $u(k), k = 0, ..., N$.
4. Sous la contrainte

$$|u(k)| \leqslant 1 \text{ pour tout } k \tag{7}$$

   trouver une suite d'entrées D-optimale, c'est-à-dire qui maximise le déterminant de la matrice d'information de Fisher (se contenter de maximiser les termes intervenant positivement et d'annuler ou de rendre négligeables ceux qui interviennent négativement).
5. Que se passerait-il si on négligeait l'existence de la contrainte (7) ?
6. Dans le cas particulier $N = 10$, comparer les performances du protocole D-optimal déterminé à la question 4 et du protocole correspondant à une entrée en échelon unité ($u(k) = 1$ pour tout $k \geqslant 0$) en termes
   (a) de déterminant de la matrice d'information de Fisher,
   (b) d'écart-type de l'erreur d'estimation de chacun des paramètres (racine carrée des éléments diagonaux de l'inverse de la matrice d'information de Fisher).

5 Localisation

Les objets connectés sont en plein développement et l'une des problématiques auxquelles les ingénieurs doivent faire face est leur localisation que ce soit en environnement extérieur ou à l'intérieur de bâtiments.

1. Indiquer brièvement pourquoi la localisation d'un objet connecté à l'aide d'un GPS n'est pas possible à l'intérieur d'un bâtiment.

Une solution pour localiser de tels objets connectés à l'intérieur d'un bâtiment est d'avoir recours à la mesure par différentes balises (WiFi, SigFox...) des instants de réception d'une onde électromagnétique émise par l'objet à localiser.

Dans la suite de cet exercice, nous considérons que l'objet est fixe, se trouve dans un plan auquel est attaché un repère $\mathcal{R}$; et que la position de l'objet dans ce repère est $\mathbf{x}_0 = \left(x_0, y_0\right)^T$. L'objet est entouré de N balises dont les positions dans $\mathcal{R}$ sont parfaitement connues et égales à $\mathbf{x}_i = \left(x_i, y_i\right)^T$, $i = 1, \ldots, N$. Chaque balise est équipée d'une horloge et nous supposerons que ces horloges sont parfaitement synchronisées. Par contre, elles ne sont pas nécessairement synchrones avec l'horloge de l'objet. Dans la suite, les temps seront indiqués dans la réferenciel temporel des balises. A l'instant $t = t_{0,0}$ (dans la base de temps des balises), l'objet émet une onde électromagnétique se déplaçant à la vitesse de la lumière dans le vide c. Cette onde est détectée aux instants $t_{0,i}$, $i = 1, \ldots, N$ par les balises. Ces mesures sont ensuite transmises à un organe de traitement central. En négligeant les réflections de l'onde et en supposant le bruit de mesure additif, les mesures peuvent s'écrire de la manière suivante

$$y_i^{\mathrm{TA}} = y_{\mathrm{m}}^{\mathrm{TA}}\left(t_{0,0}, \mathbf{x}_0, \mathbf{x}_i\right) + b_i^{\mathrm{TA}},\ i = 1, \ldots, N. \tag{8}$$

Nous supposerons les bruits de mesure b_i^{TA} gaussiens, de moyenne nulle, décorrélés, et de variances connues σ_i^2, $i = 1, \ldots, N$.

2. Donner l'expression de $y_{\mathrm{m}}^{\mathrm{TA}}\left(t_{0,0}, \mathbf{x}_0, \mathbf{x}_i\right)$.
3. En supposant que le seul vecteur de paramètres à estimer soit $\mathbf{x}_0$, le modèle $y_{\mathrm{m}}^{\mathrm{TA}}\left(t_{0,0}, \mathbf{x}_0, \mathbf{x}_i\right)$ est-il linéaire ou non-lineaire en les paramètres ?
4. Donner l'expression de la matrice de covariance du bruit de mesure $\mathbf{\Gamma}^{\mathrm{TA}} = \mathrm{E}\left(\left(\mathbf{b}^{\mathrm{TA}}\right)\left(\mathbf{b}^{\mathrm{TA}}\right)^T\right)$ avec $\mathbf{b}^{\mathrm{TA}} = \left(b_1^{\mathrm{TA}}, \ldots, b_N^{\mathrm{TA}}\right)$.

5.1 Localisation à l'aide des différences de temps d'arrivée

L'instant d'émission $t_{0,0}$ n'étant pas connu par les balises, une solution est de considérer une balise de référence $j \in \{1, \ldots, N\}$ et de calculer des différences entre les instants de détection des balises par rapport à l'instant de détection de la balise de référence. Ainsi, si la balise de référence est la balise 1, on obtient

$$y_i^{\mathrm{DTA}} = y_{i+1}^{\mathrm{TA}} - y_1^{\mathrm{TA}},\ i = 1, \ldots, N-1 \tag{9}$$

5. Montrer qu'on a alors
$$y_i^{\mathrm{DTA}} = y_{\mathrm{m}}^{\mathrm{DTA}}\left(\mathbf{x}_0, \mathbf{x}_i, \mathbf{x}_1\right) + b_i^{\mathrm{DTA}},\ i = 1, \ldots, N-1. \tag{10}$$
et donner l'expression de $y_{\mathrm{m}}^{\mathrm{DTA}}\left(\mathbf{x}_0, \mathbf{x}_i, \mathbf{x}_1\right)$ et de b_i^{DTA}.
6. Exprimer la matrice de covariance du bruit de mesure $\mathbf{\Gamma}^{\mathrm{DTA}} = \mathrm{E}\left(\left(\mathbf{b}^{\mathrm{DTA}}\right)\left(\mathbf{b}^{\mathrm{DTA}}\right)^T\right)$ avec $\mathbf{b}^{\mathrm{DTA}} = \left(b_1^{\mathrm{DTA}}, \ldots, b_N^{\mathrm{DTA}}\right)$. Les bruits de mesure dans ce cas sont-ils toujours décorrélés ?
7. Proposer une méthode de choix de la balise de référence.
8. Donner l'expression du critère à maximiser pour obtenir l'estimée au sens du maximum de vraisemblance $\widehat{\mathbf{x}}_0^{\mathrm{MV,DTA}}$ de $\mathbf{x}_0$.
9. Montrer que
$$\widehat{\mathbf{x}}_0^{\mathrm{MV,DTA}} = \arg\min_{\mathbf{x}} j^{\mathrm{DTA}}(\mathbf{x})$$
où
$$j^{\mathrm{DTA}}(\mathbf{x}) = \left(\mathbf{y}^{\mathrm{DTA}} - \mathbf{y}_{\mathrm{m}}^{\mathrm{DTA}}(\mathbf{x})\right)^T \left(\mathbf{\Gamma}^{\mathrm{DTA}}\right)^{-1} \left(\mathbf{y}^{\mathrm{DTA}} - \mathbf{y}_{\mathrm{m}}^{\mathrm{DTA}}(\mathbf{x})\right)$$
avec
$$\mathbf{y}^{\mathrm{DTA}} = \left(y_1^{\mathrm{DTA}}, \ldots, y_{N-1}^{\mathrm{DTA}}\right)^T$$
$$\mathbf{y}_{\mathrm{m}}^{\mathrm{DTA}}(\mathbf{x}) = \left(y_{\mathrm{m}}^{\mathrm{DTA}}\left(\mathbf{x}_0, \mathbf{x}_2, \mathbf{x}_1\right), \ldots, y_{\mathrm{m}}^{\mathrm{DTA}}\left(\mathbf{x}_0, \mathbf{x}_N, \mathbf{x}_1\right)\right)^T.$$
On rappelle qu'un vecteur $\mathbf{b}$ Gaussien de moyenne $\boldsymbol{\mu} \in \mathbb{R}^N$ et de matrice de covariance $\mathbf{\Gamma}$ a pour densité de probabilité
$$\pi_{\mathrm{B}}(\mathbf{b}) = (2\pi)^{-N/2} \left(\det(\mathbf{\Gamma})\right)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{b} - \boldsymbol{\mu})^T \mathbf{\Gamma}^{-1} (\mathbf{b} - \boldsymbol{\mu})\right). \tag{11}$$
10. Est-il possible d'obtenir facilement une expression explicite de $\widehat{\mathbf{x}}_0^{\mathrm{MV,DTA}}$?

On propose une technique itérative du type gradient pour essayer d'obtenir $\widehat{\mathbf{x}}_0^{\mathrm{MV,DTA}}$. Pour cela, partant d'une estimée initiale $\mathbf{x}(0)$, on applique la relation de récurrence suivante

$$\mathbf{x}(k+1) = \mathbf{x}(k) - \lambda(k)\mathbf{g}\left(\mathbf{x}(k)\right) \tag{12}$$

où $\lambda(k)$ est un pas à adapter et $\mathbf{g}(\mathbf{x}(k))$ est le gradient du critère $j^{\mathrm{DTA}}(\mathbf{x})$ évalué en $\mathbf{x}(k)$.

11. Donner l'expression de $\mathbf{g}(\mathbf{x}(k))$ en fonction de $\mathbf{y}^{\mathrm{DTA}}$, $\mathbf{y}_{\mathrm{m}}^{\mathrm{DTA}}(\mathbf{x}(k))$, $\mathbf{\Gamma}^{\mathrm{DTA}}$ et des autres variables du problème.
12. Suggérer une initialisation possible pour $\mathbf{x}(0)$ qui permette une convergence plus rapide de l'algoritme itératif qu'avec une initialisation aléatoire.
13. Proposer une technique permettant de choisir $\lambda(k)$ à chaque itération.
14. Proposer un critère d'arrêt des itérations.

5.2 Localisation à l'aide des temps d'arrivée

Dans ce cas, l'instant d'émission $t_{0,0}$ doit être estimé conjointement avec la position $\mathbf{x}_0$ de l'objet. Il faut donc estimer le vecteur de paramètres $\mathbf{p} = \left(t_{0,0}; \mathbf{x}_0^T\right)^T$.

15. Montrer que dans ce cas, l'estimée au sens du maximum de vraisemblance de $\mathbf{p}$ nécessite de résoudre le problème suivant
$$\widehat{\mathbf{p}}^{\mathrm{MV,TA}} = \arg\min_{\mathbf{p}} j^{\mathrm{TA}}\left(\mathbf{p}\right)$$
où
$$j^{\mathrm{TA}}\left(\mathbf{p}\right) = \left(\mathbf{y}^{\mathrm{TA}} - \mathbf{y}_{\mathrm{m}}^{\mathrm{TA}}\left(\mathbf{p}\right)\right)^T \left(\mathbf{\Gamma}^{\mathrm{TA}}\right)^{-1} \left(\mathbf{y}^{\mathrm{TA}} - \mathbf{y}_{\mathrm{m}}^{\mathrm{TA}}\left(\mathbf{p}\right)\right)$$
avec
$$\begin{aligned} \mathbf{y}^{\mathrm{TA}} &= \left(y_1^{\mathrm{TA}}, \ldots, y_N^{\mathrm{TA}}\right)^T \\ \mathbf{y}_{\mathrm{m}}^{\mathrm{TA}}\left(\mathbf{p}\right) &= \left(y_{\mathrm{m}}^{\mathrm{TA}}\left(t_{0,0}, \mathbf{x}_0, \mathbf{x}_1\right), \ldots, y_{\mathrm{m}}^{\mathrm{TA}}\left(t_{0,0}, \mathbf{x}_0, \mathbf{x}_N\right)\right)^T. \end{aligned}$$
16. Un procédure itérative similaire à celle présentée au paragraphe précédent peut être mise en oeuvre. Comment initialiser la valeur $\mathbf{p}\left(0\right)$ du vecteur des paramètres ?
17. Des mesures aberrantes sont souvent présentes dans ce type d'application.
    (a) Quelle est l'origine de ces mesures aberrantes ?
    (b) Quel est l'impact de ces mesures aberrantes sur l'estimée obtenue ?
    (c) Quelle solution pouvez-vous mettre en oeuvre de manière à en limiter l'impact ?