Exercice numéro 6: Estimation au sens du MV et estimation bayésienne

Considérons le système $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$

1. Dans une approche d'estimation au sens du maximum de vraisemblance (MV) supposons que le bruit $\boldsymbol{\epsilon}$ soit supposé centré, blanc, gaussien et de variance ${\sigma_\epsilon}^2$ fixée. Montrez que l'estimation au sens du MV de $\boldsymbol{f}$, *i.e.*;
$$\widehat{\boldsymbol{f}}_{\text{MV}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{g}|\boldsymbol{f})\right\}$$
s'obtient en minimisant
$$J_1(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2.$$

2. Quelle est l'expression de cet estimateur si $\boldsymbol{\epsilon}$ est supposé suivre une loi gaussienne généralisée, *i.e.*;
$$\boldsymbol{\epsilon} \sim p(\boldsymbol{\epsilon}) \propto \exp\left[-\|\boldsymbol{\epsilon}\|^\alpha\right] = \exp\left[-\sum_m |\epsilon_m|^\alpha\right], \quad 1 < \alpha \leq 2$$

3. Dans une approche bayésienne pour résoudre ce problème, supposons que l'on puisse attribuer des lois gaussiennes aux deux vecteurs $\boldsymbol{\epsilon}$ et $\boldsymbol{f}$:
$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_\epsilon), \qquad \boldsymbol{f} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_f),$$
où $\boldsymbol{R}_\epsilon = \sigma_\epsilon^2 \boldsymbol{I}$ et $\boldsymbol{R}_f = \sigma_f^2 \boldsymbol{P}_0 = \sigma_f^2 (\boldsymbol{D}'\boldsymbol{D})^{-1}$ sont les matrices de covariance de $\boldsymbol{\epsilon}$ et de $\boldsymbol{f}$.
Écrivez l'expression des lois $p(\boldsymbol{f}), p(\boldsymbol{\epsilon}), p(\boldsymbol{g}, \boldsymbol{f})$ et $p(\boldsymbol{g}|\boldsymbol{f})$.

4. Montrez que la loi *a posteriori* $p(\boldsymbol{f}|\boldsymbol{g})$ est une loi gaussienne de la forme
$$p(\boldsymbol{f}|\boldsymbol{g}) \propto \exp\left[-\frac{1}{2}\left[[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]'\boldsymbol{R}_\epsilon^{-1}[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}] + \boldsymbol{f}'\boldsymbol{R}_f^{-1}\boldsymbol{f}\right]\right]$$

5. Si on note $\widehat{\boldsymbol{f}}_3$ la solution qui maximise $p(\boldsymbol{f}|\boldsymbol{g})$ (l'estimation au sens du maximum *a posteriori* MAP), montrez qu'elle s'obtient par
$$\widehat{\boldsymbol{f}}_3 = \arg\min_{\boldsymbol{f}} \left\{J_3(\boldsymbol{f}) = [\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]'\boldsymbol{R}_\epsilon^{-1}[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}] + \boldsymbol{f}'\boldsymbol{R}_f^{-1}\boldsymbol{f}\right\}$$

6. Si $\boldsymbol{R}_\epsilon = \sigma_\epsilon^2 \boldsymbol{I}$ et $\boldsymbol{R}_f = \sigma_f^2 \boldsymbol{P}_0$, Montrez que $\widehat{\boldsymbol{f}}_3$ s'obtient par
$$\widehat{\boldsymbol{f}}_3 = \left[\boldsymbol{H}'\boldsymbol{H} + \lambda \boldsymbol{P}_0^{-1}\right]^{-1} \boldsymbol{H}'\boldsymbol{g} \quad \text{avec} \quad \lambda = \frac{\sigma_\epsilon^2}{\sigma_f^2}$$
Que peut–on conclure en comparant les solutions $\widehat{\boldsymbol{f}}_2$ et $\widehat{\boldsymbol{f}}_3$?

7. En développant le terme
$$\left[[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}]'\boldsymbol{R}_\epsilon^{-1}[\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}] + \boldsymbol{f}'\boldsymbol{R}_f^{-1}\boldsymbol{f}\right]$$
dans l'expression de $p(\boldsymbol{f}|\boldsymbol{g})$ montrez que l'on peut écrire
$$p(\boldsymbol{f}|\boldsymbol{g}) \propto \exp\left[-\frac{1}{2}[\boldsymbol{f} - \widehat{\boldsymbol{f}}_3]'\widehat{\boldsymbol{P}}^{-1}[\boldsymbol{f} - \widehat{\boldsymbol{f}}_3]\right]$$
Quelle est alors l'expression de la matrice de covariance *a posteriori* $\widehat{\boldsymbol{P}}$?
Que représentent les éléments diagonaux de cette matrice?

8. Écrivez l'expression des lois $p(f_n|\boldsymbol{g}), p(g_m|\boldsymbol{g})$ et $p(b_m|\boldsymbol{g})$. A quoi peuvent-elle servir ?

9. Supposons maintenant la suite $\{f_1, \cdots, f_N\}$ puissent être modélisée par une chaîne de Markov d'ordre un, c'est à dire :
$$p(f_n|f_1, \cdots, f_N) = p(f_n|f_{n-1})$$
Peut-on calculer $p(\boldsymbol{f})$? Et si on connaît de plus $p(f_1)$?
Que devient alors la solution au sens du MAP ?
Étudiez cette solution dans les deux cas suivants:
$$p(f_n|f_1, \cdots, f_{n-1}, f_{n+1}, \cdots, f_N) = p(f_n|f_{n-1}) = \mathcal{N}(f_{n-1}, \sigma_f^2), \quad \text{et} \quad p(f_1) = \mathcal{N}(0, \sigma_f^2)$$
et
$$p(f_n|f_1, \cdots, f_{n-1}, f_{n+1}, \cdots, f_N) = p(f_n|f_{n-1}) \propto \exp\left[-\alpha\phi(f_n - f_{n-1})\right], \quad \text{et} \quad p(f_1) \propto \exp\left[-\alpha\phi(f_1)\right]$$

1. Supposons maintenant la suite $\boldsymbol{f} = \{f_1, \cdots, f_N\}$ soit associée à une suite de variables binaires $\boldsymbol{q} = \{q_1, \cdots, q_N\}$ avec des relations suivantes :

$$p(f_n|f_1, \cdots, f_{n-1}, f_{n+1}, \cdots, f_N, q_n) = p(f_n|q_n, f_{n-1})$$

   Supposons d'abord $\boldsymbol{q}$ connu.
   Peut-on calculer $p(\boldsymbol{f}|\boldsymbol{q})$? Et si on connaît de plus $p(f_1|q_1)$?
   Que devient alors la solution au sens du MAP pour $\boldsymbol{f}$?
   Étudiez cette solution dans le cas suivant :

$$p(f_n|f_{n-1}, q_n) = \mathcal{N}((1-q_n)f_{n-1}, \sigma_f^2), \quad \text{et} \quad p(f_1|q_1 = 0) = p(f_1|q_1 = 1) = \mathcal{N}(0, \sigma_f^2)$$

2. Supposons maintenant que l'on souhaite estimer $\boldsymbol{q}$ et $\boldsymbol{f}$. Supposons d'abord que les Q_n sont iid avec $P(Q_n = 1) = \alpha$ et donc $P(Q_n = 0) = 1 - \alpha$.

3. Écrivez l'expression de $p(\boldsymbol{q}) = P(\boldsymbol{Q} = \boldsymbol{q})$. Écrivez aussi l'expression de $p(\boldsymbol{f}, \boldsymbol{q}|\boldsymbol{g})$. Peut-on envisager l'estimation jointe de $\boldsymbol{f}$ et de $\boldsymbol{q}$ par un algorithme itératif du genre

$$\left\{\begin{array}{lcl} \widehat{\boldsymbol{f}} & = & \arg\max\limits_{\boldsymbol{f}} \{p(\boldsymbol{f}, \widehat{\boldsymbol{q}}|\boldsymbol{g})\} \\ \widehat{\boldsymbol{q}} & = & \arg\max\limits_{\boldsymbol{q}} \left\{p(\widehat{\boldsymbol{f}}, \boldsymbol{q}|\boldsymbol{g})\right\} \end{array}\right.$$

4. Dans le cas gaussien, montrez que

$$\left\{\begin{array}{lcl} \widehat{\boldsymbol{f}} & = & \arg\max\limits_{\boldsymbol{f}} \{p(\boldsymbol{f}, \widehat{\boldsymbol{q}}|\boldsymbol{g})\} = \arg\min\limits_{\boldsymbol{f}} \left\{ J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda \sum\limits_j (1 - \widehat{q}_j)(f_j - f_{j-1})^2 \right\} \\ \widehat{q}_j & = & 1 \;\text{ si }\; |\widehat{f}_j - \widehat{f}_{j-1})| > T \end{array}\right.$$

5. Montrez que la solution obtenue par ces deux équations est équivalent à la solution qu'on peut obtenir en optimisant le critère non convexe suivant

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda \sum_j \phi(f_j - f_{j-1}) \right\}$$

   avec $\phi(t) = \max(T, t^2)$ qui est une quadratique tronquée.

6.

7. Écrivez l'expression de $p(\boldsymbol{g}|\boldsymbol{q})$ et $p(\boldsymbol{q}|\boldsymbol{g})$. Peut-on envisager l'estimation jointe de $\boldsymbol{f}$ et de $\boldsymbol{q}$ par un algorithme itératif du genre

$$\left\{\begin{array}{lcl} \widehat{\boldsymbol{f}} & = & \arg\max\limits_{\boldsymbol{f}} \{p(\boldsymbol{f}|\widehat{\boldsymbol{q}}, \boldsymbol{g})\} \\ \widehat{\boldsymbol{q}} & = & \arg\max\limits_{\boldsymbol{q}} \{p(\boldsymbol{q}|\boldsymbol{g})\} \end{array}\right.$$

8. Supposons qu'on connaît $\boldsymbol{q}$ sauf q_n. Notons par $\boldsymbol{q}_{-n} = \{q_1, \cdots, q_{n-1}, q_{n+1}, \cdots, q_n\}$.

   Écrivez les expressions de $P(q_n = 0|\boldsymbol{f}, \boldsymbol{q}_{-n})$, $P(q_n = 1|\boldsymbol{f}, \boldsymbol{q}_{-n})$ et $l_n(q_n|\boldsymbol{f}, \boldsymbol{q}_{-n}) = \ln \frac{P(q_n=1|\boldsymbol{f}, \boldsymbol{q}_{-n})}{P(q_n=0|\boldsymbol{f}, \boldsymbol{q}_{-n})}$

   Écrivez les expressions de $P(q_n = 0|\boldsymbol{g}, \boldsymbol{q}_{-n})$, $P(q_n = 1|\boldsymbol{g}, \boldsymbol{q}_{-n})$ et $l_n(q_n|\boldsymbol{g}, \boldsymbol{q}_{-n}) = \ln \frac{P(q_n=1|\boldsymbol{g}, \boldsymbol{q}_{-n})}{P(q_n=0|\boldsymbol{g}, \boldsymbol{q}_{-n})}$.

9. Étudiez les performances des détecteurs :
   $q_n = 1$ si $P(q_n = 1|\boldsymbol{f}, \boldsymbol{q}_{-n}) > P(q_n = 0|\boldsymbol{f}, \boldsymbol{q}_{-n})$
   et
   $q_n = 1$ si $P(q_n = 1|\boldsymbol{g}, \boldsymbol{q}_{-n}) > P(q_n = 0|\boldsymbol{g}, \boldsymbol{q}_{-n})$.