Filtrage Stochastique
HPC-BigDATA
2021-2022

Consignes

- Durée du test: 1h50.
- **Aucun Document autorisé**
- Le sujet est **recto-verso** et le barême est indiqué par sous-section

---

1 Interprétation des équations du filtre de Kalman (KF) (6 points)

Question 1. : *Énoncez les quatres équations du filtre de Kalman (KF) en séparant l'étape d'analyse et l'étape de prévision (vous expliciterez la valeur de la matrice de gain).*

Question 2. : *Quelle est l'interprétation probabiliste de ces équations ?*

On considère une expérience simulée d'assimilation de données. La dynamique est donnée par le transport d'une concentration par un écoulement

$$\partial_t C + u\partial_x C = 0, \tag{1}$$

avec $C(t,x)$ une concentration et $u(x)$ la vitesse de l'écoulement, stationnaire en temps mais hétérogène en espace. La condition initiale de l'incertitude est donnée par une matrice de covariance de variance 1 et de corrélation de fonctions Gaussienne (on utilise la même fonction gaussienne pour tous les points), notée $\mathbf{P}_0^f$.

On peut montrer que, pour la dynamique Eq.(1), l'équation aux dérivées partielles décrivant l'évolution de la variance d'erreur de prévision, $V^f(t,x)$, s'écrit

$$\partial_t V^f + u\partial_x V^f = 0, \tag{2}$$

avec $V^f(t,x) = \mathbb{E}\left[(\varepsilon^f)^2\right]$ où $\varepsilon^f = C^f - C^t$, C^f (resp. C^t) désignant le vrai champ de consentration prévu (res. vrai).

Question 3. : *À quelles équations du filtre de Kalman, les équations Eq.(1) et Eq.(2), correspondent-elles ? (donnez explicitement les équations ou part d'équations du KF correspondantes)*

Question 4. : *Interprétez l'équation Eq.(2). Peut-il y avoir création d'incertitude pour la dynamique Eq.(1) ?*

Question 5. : *Le résultat précédent est-il encore vrai avec l'équation de conservation $\partial_t C + \partial_x(uC) = 0$,pour laquelle l'évolution de la variance est donnée par $\partial_t V^f + u\partial_x V^f = -2V\partial_x u$. Si ce n'est pas le cas, identifiez quelles zones de l'écoulement correspondant aux sources et aux puits d'incertitudes (faire un dessin pour faciliter l'explication).*

2 Filtre particulaire (4 points)

Le filtre non-linéaire repose sur deux étapes: l'analyse décrite par la formule de Bayes

$$p(\mathcal{X}_q/\mathcal{Y}^o{}_{0:q}) \sim p(\mathcal{Y}^o{}_q/\mathcal{X}_q)p(\mathcal{X}_q/\mathcal{Y}^o{}_{0:q-1})$$

et la prévision des incertitudes. On considère la résolution du filtrage non-linéaire à l'aide d'un échantillonnage pour la distribution initiale $p(\mathcal{X}_0) = \frac{1}{N}\sum_k \delta(\mathcal{X} - \mathcal{X}_0^k)$, et des distributions ultérieures: il s'agit du formalisme du filtre particulaire. À un instant t_q, la distribution de prévision

$$p^f(\mathcal{X}_q) = p(\mathcal{X}_q/\mathcal{Y}^o{}_{0:q-1}) = \frac{1}{N}\sum_k \delta(\mathcal{X}_q - \mathcal{X}_q^k),$$

est mise à jour en repondérant les prévision x_q^k pour former la distrbution a posteriori $p^a(\mathcal{X}_q) = p(\mathcal{X}_q/\mathcal{Y}^o{}_{0:q}) = \sum_k w_q^k \delta(\mathcal{X}_q - \mathcal{X}_q^k)$, dont un rééchantillonage, $\left(\widetilde{\mathcal{X}}_q^k\right)$, permet de réécrire cette distribution sous la forme

$$p^a(\mathcal{X}_q) = p(\mathcal{X}_q/\mathcal{Y}^o{}_{0:q}) = \frac{1}{N}\sum_k \delta(\mathcal{X}_q - \widetilde{\mathcal{X}}_q^k).$$

On suppose que le système est déterministe d'équation $\frac{d\mathcal{X}}{dt} = \mathcal{M}(\mathcal{X})$. Le propagateur, entre deux instants t_q et t_q+1, est noté $\mathcal{M}_{t_{q+1}\leftarrow t_q}$

Question 6. : *Étape d'analyse: Donnez l'expression de des w_q^k. Illustrez shématiquement les distributions p^f et p^a.*

Question 7. : *Étape de prévision: À partir de la dynamique des mesures vue dans le cadre déterministe, exprimez analytiquement la distribution $p^f(\mathcal{X}_{q+1})$ à partir des $\widetilde{\mathcal{X}}_q^k$.*

Dans le cadre linéaire Gaussien, le filtre de Kalman et le filtre particulaire sont équivalents d'un point de vue théorique. On considère la situation où la matrice de covariance d'erreur de prévision est la matrice identité, $\mathbf{P}^f = \mathbf{I}_n$ avec $n > 100$ la dimension du problème.

Question 8. : *Pourquoi la discrétisation de la formule de Bayes pose-t-elle problème en grande dimension ? Vous argumenterez à l'aide d'une interprétation géométrique précise, dont vous donnerez la démonstration, puis en comparant avec l'application du filtre de Kalman.*

3 Calcul stochastique

Dans la suite de l'énoncé, $(\Omega, \mathcal{F}, P)$ est un espace probabilisé et B un processus Brownien de filtration canonique $(\mathcal{F}_t)$, *i.e.* la suite croissante pour l'inclusion des sous-tribues $\mathcal{F}_t$ de $\mathcal{F}$ avec $\mathcal{F}_t = \sigma(\{B_s, s \leq t\})$ avec $t \in [0, T]$.

3.1 Formule d'Itô (5 points)

Soit un processus d'Itô $dX_t(\omega) = a_t(\omega)dt + b_t(\omega)dB_t(\omega)$, où a, b sont des processus adaptés au brownien B et d'énergie moyenne finie, *i.e.* $\mathbb{E}\left[\int_0^T a_t^2 dt\right] < \infty$ et $\mathbb{E}\left[\int_0^T b_t^2 dt\right] < \infty$, et $Y_t = f(t, X_t)$ où f est $C^{1,2}$

Question 9. : *Démontrez la formule d'Itô permettant d'exprimer la formulation différentielle dY_t de Y_t, puis exprimez la formulation intégrale de Y_t. Vous donnerez l'expression condensée utilisant l'opérateur $\bullet$ vu en cours, puis l'expression sous forme de processus d'Itô.*

Question 10. : *Résoudre l'équation différentielle stochastique $dX_t = \mu X_t dt + \sigma X_t dB_t$ avec μ et σ deux réels non nuls.*

3.2 Distribution limite pour un flot gradient stochastique (5 points)

On considère la dynamique donnée par

$$\frac{dx(t)}{dt} = -\nabla V[x(t)], \tag{3}$$

avec $V(x)$ un fonction potentielle, *i.e.* une fonction C^∞ telle que $\lim_{x\to\pm\infty} V(x) = +\infty$.

Question 11. : *Illustrez graphiquement le comportement des solutions de Eq. (3) en justifiant votre réponse à l'aide de l'étude de la dynamique de $v(t) = V[x(t)]$ où $x(t)$ est une solution de Eq. (3). En déduire les distributions limites possibles suivant l'incertitude initiale dans le cas du potentiel $V(x) = x^4 - 2x^2$.*

On considère maintenant la dynamique stochastique

$$dX_t = -\nabla V(X_t)dt + \sigma dB_t, \tag{4}$$

avec $\sigma > 0$ une constante (ni trop grande, ni trop petite).

Question 12. : *Exprimez la dynamique des observables f (fonctions bornées C^2) pour la dynamique Eq. (4).*

Question 13. : *En déduire la dynamique des mesures de probabilité qui sont à densité, p, par rapport à la mesure de Lebesgue.*

Question 14. : *Montrez qu'il existe une distribution limite $p^\infty(x) = \lim_{t\to\infty} p(t, x)$ que vous caractériserez à l'aide du potentiel V (on notera Z la constante de normalisation, i.e. $Z = \int_{-\infty}^{\infty} p^\infty(x)dx$).*

Question 15. : *Illustrez la situation dans le cas du potentiel $V(x) = x^4 - 2x^2$: vous donnerez une illustration d'une trajectoire typique $X_t(\omega)$ issue de la condition initiale $X_0(\omega) = 0$, puis le graphique de la distribution p^∞.*