

réunion de l'ARC ADOQA  
Assimilation de Données pour la Qualité de l'Air  
Grenoble, 17 et 18 octobre 2006

## Filtre de Kalman d'ensemble vs. Filtres particuliers

François Le Gland  
IRISA / INRIA Rennes

legland@irisa.fr  
<http://www.irisa.fr/aspi/>

Valérie Monbet + Vu Duc Tran  
université de Bretagne Sud (UBS), Vannes

## Plan

- systèmes non-linéaires / non gaussiens
- filtre de Kalman d'ensemble (EnKF)
- filtre bayésien : formulation conjointe vs. équation récurrente
- distribution d'importance optimale
- filtre particulaire SIS (sans ré-échantillonnage)
- filtre particulaire SIR (avec ré-échantillonnage)
- exemple : modèle de Lorenz
- théorèmes limites

---

 systèmes non-linéaires / non gaussiens

état caché, en dimension  $m$

$$X_k = f_k(X_{k-1}, W_k), \quad W_k \sim p_k^W(dw) \quad \text{ou bien} \quad p(x_k | x_{k-1})$$

observations partielles et bruitées, en dimension  $1 \ll d \ll m$

$$Y_k = h_k(X_k) + V_k, \quad V_k \sim q_k^V(v) dv \quad \text{ou bien} \quad p(y_k | x_k)$$

cas particulier rencontré en assimilation de données

$$X_k = f_k(X_{k-1}) + W_k, \quad W_k \sim \mathcal{N}(0, Q_k^W)$$

et

$$Y_k = H_k X_k + V_k, \quad V_k \sim \mathcal{N}(0, Q_k^V)$$

estimer l'état caché  $X_k$  au vu des observations passées  $Y_{0:k} = (Y_0, \dots, Y_k)$

estimation bayésienne : calcul (approché) de  $p_k(x_k) = p(x_k | y_{0:k})$

---

## filtre de Kalman d'ensemble (EnKF)

ensemble de  $i = 1, \dots, N$  éléments

► étape de prédiction (forecast) : indépendamment pour tout  $i = 1, \dots, N$

$$X_k^{f,i} = f_k(X_{k-1}^{a,i}) + W_k^i, \quad W_k^i \sim \mathcal{N}(0, Q_k^W)$$

moyenne et matrice de covariance empiriques

$$\hat{X}_k^f = \frac{1}{N} \sum_{i=1}^N X_k^{f,i} \quad \text{et} \quad \hat{P}_k^f = \frac{1}{N-1} \sum_{i=1}^N (X_k^{f,i} - \hat{X}_k^f) (X_k^{f,i} - \hat{X}_k^f)^*$$

avantage : estimation empirique vs. équation de Lyapunov

► étape de correction (analysis) : indépendamment pour tout  $i = 1, \dots, N$

$$V_k^i \sim \mathcal{N}(0, Q_k^V)$$

moyenne et matrice de covariance empiriques

$$\hat{V}_k = \frac{1}{N} \sum_{i=1}^N V_k^i \quad \text{et} \quad \hat{Q}_k^V = \frac{1}{N-1} \sum_{i=1}^N (V_k^i - \hat{V}_k) (V_k^i - \hat{V}_k)^*$$

mise-à-jour

$$X_k^{a,i} = X_k^{f,i} + \hat{P}_k^f H_k^* [H_k \hat{P}_k^f H_k^* + \hat{Q}_k^V]^{-1} (Y_k - H_k X_k^{f,i} - V_k^i)$$

bilan

- chaque élément d'ensemble contribue, à travers la matrice de covariance empirique, à un gain de Kalman unique
- les observations interviennent en déplaçant chaque élément d'ensemble selon l'étape de correction du filtre de Kalman, avec un gain de Kalman unique

question : que calcule-t-on vraiment, par exemple que représente la distribution empirique

$$\frac{1}{N} \sum_{i=1}^N \delta_{X_k^{a,i}}$$

comportement limite quand la taille  $N$  de l'ensemble tend vers l'infini

---

**filtre bayésien : formulation conjointe**

loi jointe du couple (état caché + observation) vu comme chaîne de Markov

$$p(x_{0:n}, y_{0:n}) = \prod_{k=1}^n p(x_k, y_k \mid x_{k-1}, y_{k-1}) p(x_0, y_0)$$

densité de transition

$$\begin{aligned} p(x_k, y_k \mid x_{k-1}, y_{k-1}) &= \frac{p(x_k, x_{k-1}, y_k, y_{k-1})}{p(x_{k-1}, y_{k-1})} \\ &= \frac{p(y_k, y_{k-1} \mid x_k, x_{k-1}) p(x_k, x_{k-1})}{p(x_{k-1}, y_{k-1})} \\ &= \frac{p(y_k \mid x_k) p(y_{k-1} \mid x_{k-1}) p(x_k \mid x_{k-1}) p(x_{k-1})}{p(y_{k-1} \mid x_{k-1}) p(x_{k-1})} \\ &= p(y_k \mid x_k) p(x_k \mid x_{k-1}) = p(x_k, y_k \mid x_{k-1}) \end{aligned}$$

► avec la factorisation

$$p(x_k, y_k | x_{k-1}) = p(y_k | x_k) p(x_k | x_{k-1}) \quad \text{et} \quad p(x_0, y_0) = p(y_0 | x_0) p(x_0)$$

et après marginalisation

$$p(x_{0:n} | y_{0:n}) \propto \underbrace{\prod_{k=0}^n p(y_k | x_k)}_{g(x_{0:n})} \underbrace{\prod_{k=1}^n p(x_k | x_{k-1}) p(x_0)}_{q(x_{0:n})}$$

► avec la factorisation (optimale) alternative

$$p(x_k, y_k | x_{k-1}) = p(y_k | x_{k-1}) p(x_k | x_{k-1}, y_k) \quad \text{et} \quad p(x_0, y_0) \propto p(x_0 | y_0)$$

et après marginalisation

$$p(x_{0:n} | y_{0:n}) \propto \underbrace{\prod_{k=0}^n p(y_k | x_{k-1})}_{g(x_{0:n})} \underbrace{\prod_{k=1}^n p(x_k | x_{k-1}, y_k) p(x_0 | y_0)}_{q(x_{0:n})}$$

---

## filtre bayésien : équation récurrente

rappel

$$p(x_{0:n}, y_{0:n}) = \prod_{k=1}^n p(x_k, y_k \mid x_{k-1}) p(x_0, y_0)$$

forme récursive

$$p(x_{0:k}, y_{0:k}) = p(x_k, y_k \mid x_{k-1}) p(x_{0:k-1}, y_{0:k-1})$$

et après marginalisation (intégration par rapport à  $x_{0:k-1}$ )

$$p(x_k, y_{0:k}) = \int p(x_k, y_k \mid x_{k-1}) p(x_{k-1}, y_{0:k-1}) dx_{k-1}$$

soit

$$p(x_k \mid y_{0:k}) \propto \int p(x_k, y_k \mid x_{k-1}) p(x_{k-1} \mid y_{0:k-1}) dx_{k-1}$$

ré-écriture, avec la notation  $p_k(x_k) = p(x_k | y_{0:k-1})$  pour le filtre bayésien

$$p_k(x_k) \propto \int p(x_k, y_k | x_{k-1}) p_{k-1}(x_{k-1}) dx_{k-1}$$

► avec la factorisation standard

$$p_k(x_k) \propto p(y_k | x_k) \int p(x_k | x_{k-1}) p_{k-1}(x_{k-1}) dx_{k-1}$$

et

$$p_0(x_0) \propto p(y_0 | x_0) p(x_0)$$

► avec la factorisation optimale

$$p_k(x_k) \propto \int p(y_k | x_{k-1}) p(x_k | x_{k-1}, y_k) p_{k-1}(x_{k-1}) dx_{k-1}$$

et

$$p_0(x_0) \propto p(x_0 | y_0)$$

---

## distribution d'importance optimale

problème

- simuler selon la densité  $p(x_k | x_{k-1}, y_k)$  ou la densité  $p(x_0 | y_0)$
- évaluer la fonction  $p(y_k | x_{k-1})$

possible dans le cas particulier de systèmes avec

- condition initiale gaussienne, et bruits gaussiens additifs
- fonction d'observation linéaire

c'est-à-dire

$$X_k = f_k(X_{k-1}) + W_k, \quad W_k \sim \mathcal{N}(0, Q_k^W)$$

$$Y_k = H_k X_k + V_k, \quad V_k \sim \mathcal{N}(0, Q_k^V)$$

avec la condition initiale  $X_0 \sim \mathcal{N}(\bar{X}_0, Q_0^X)$

le vecteur aléatoire  $(X_0, Y_0)$  est gaussien, de moyenne et de matrice de covariance

$$\begin{pmatrix} \bar{X}_0 \\ H_0 \bar{X}_0 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} Q_0^X & Q_0^X H_0^* \\ H_0 Q_0^X & H_0 Q_0^X H_0^* + Q_0^V \end{pmatrix}$$

► la densité  $p(x_0 | y_0)$  est donc gaussienne, de moyenne

$$\bar{X}_0 + Q_0^X H_0^* [H_0 Q_0^X H_0^* + Q_0^V]^{-1} (y_0 - H_0 \bar{X}_0)$$

et de matrice de covariance pré-calculable

$$Q_0^X - Q_0^X H_0^* [H_0 Q_0^X H_0^* + Q_0^V]^{-1} H_0 Q_0^X$$

grâce à la linéarité de la fonction d'observation

$$X_k = f_k(X_{k-1}) + W_k, \quad W_k \sim \mathcal{N}(0, Q_k^W)$$

$$Y_k = H_k f_k(X_{k-1}) + H_k W_k + V_k, \quad V_k \sim \mathcal{N}(0, Q_k^V)$$

conditionnellement à  $X_{k-1} = x$ , le vecteur aléatoire  $(X_k, Y_k)$  est gaussien, de moyenne et de matrice de covariance

$$\begin{pmatrix} f_k(x) \\ H_k f_k(x) \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} Q_k^W & Q_k^W H_k^* \\ H_k Q_k^W & H_k Q_k^W H_k^* + Q_k^V \end{pmatrix}$$

► en particulier, la densité  $p(y_k | x_{k-1})$  est gaussienne, de moyenne

$$H_k f_k(x_{k-1})$$

et de matrice de covariance inversible et pré-calculable

$$H_k Q_k^W H_k^* + Q_k^V$$

on notera la densité correspondante

$$q(y_k - H_k f_k(x_{k-1}), Q_k^W H_k^* + Q_k^V)$$

► la densité  $p(x_k | x_{k-1}, y_k)$  est également gaussienne, de moyenne

$$f_k(x_{k-1}) + Q_k^W H_k^* [H_k Q_k^W H_k^* + Q_k^V]^{-1} (y_k - H_k f_k(x_{k-1}))$$

et de matrice de covariance pré-calculable

$$Q_k^W - Q_k^W H_k^* [H_k Q_k^W H_k^* + Q_k^V]^{-1} H_k Q_k^W$$

---

 filtre particulaire SIS (sans ré-échantillonnage)

► avec la factorisation standard

$$p(x_{0:n} | y_{0:n}) \approx \sum_{i=1}^N \frac{g(\xi_{0:n}^i)}{\sum_{j=1}^N g(\xi_{0:n}^j)} \delta_{\xi_{0:n}^i}$$

avec

$$g(x_{0:n}) = \prod_{k=0}^n p(y_k | x_k)$$

et avec  $(\xi_{0:n}^1, \dots, \xi_{0:n}^N)$  un  $N$ -échantillon de densité

$$p(x_{0:n}) = \prod_{k=0}^n p(x_k | x_{k-1}) p(x_0)$$

forme récursive

- pour  $k = 0$ , indépendamment pour tout  $i = 1, \dots, N$   
on simule une v.a.  $\xi_i^0$  gaussienne, de moyenne  $\bar{X}_0$  et de matrice de covariance  $Q_0^X$ , et on pose

$$w_0^i = \frac{q(y_0 - H_0 \xi_0^i, Q_0^V)}{\sum_{j=1}^N q(y_0 - H_0 \xi_0^j, Q_0^V)}$$

- pour tout  $k = 1, \dots, n$ , indépendamment pour tout  $i = 1, \dots, N$   
on simule une v.a.  $\xi_k^i$  gaussienne, de moyenne  $f_k(\xi_{k-1}^i)$  et de matrice de covariance  $Q_k^W$ , et on met à jour

$$w_k^i = \frac{w_{k-1}^i q(y_k - H_k \xi_k^i, Q_k^V)}{\sum_{j=1}^N w_{k-1}^j q(y_k - H_k \xi_k^j, Q_k^V)}$$

## bilan

- chaque particule se déplace indépendamment selon le modèle a priori uniquement
- les observations interviennent après coup, en affectant un poids plus ou moins grand à chaque particule, en fonction de sa vraisemblance

plutôt inefficace, en particulier si le modèle a priori est très sensible aux conditions initiales

## solutions

- déplacer chaque particule indépendamment selon une distribution d'importance qui prend en compte à la fois le modèle a priori et les observations (modification associée du poids d'importance affecté à chaque particule)
- utiliser les poids pour reproduire / terminer les différentes particules

► avec la factorisation optimale

$$p(x_{0:n} \mid y_{0:n}) \approx \sum_{i=1}^N \frac{g(\xi_{0:n}^i)}{\sum_{j=1}^N g(\xi_{0:n}^j)} \delta_{\xi_{0:n}^i}$$

avec

$$g(x_{0:n}) = \prod_{k=0}^n p(y_k \mid x_{k-1})$$

et avec  $(\xi_{0:n}^1, \dots, \xi_{0:n}^N)$  un  $N$ -échantillon de densité

$$q(x_{0:n}) = \prod_{k=0}^n p(x_k \mid x_{k-1}, y_k) p(x_0 \mid y_0)$$

forme récursive

- pour  $k = 0$ , indépendamment pour tout  $i = 1, \dots, N$ ,  
on simule une v.a.  $\xi_i^0$  gaussienne, de moyenne

$$\bar{X}_0 + Q_0^X H_0^* [H_0 Q_0^X H_0^* + Q_0^V]^{-1} (y_0 - H_0 \bar{X}_0)$$

et de matrice de covariance

$$Q_0^X - Q_0^X H_0^* [H_0 Q_0^X H_0^* + Q_0^V]^{-1} H_0 Q_0^X$$

- pour tout  $k = 1, \dots, n$ , indépendamment pour tout  $i = 1, \dots, N$   
on simule une v.a.  $\xi_k^i$  gaussienne, de moyenne

$$f_k(\xi_{k-1}^i) + Q_k^W H_k^* [H_k Q_k^W H_k^* + Q_k^V]^{-1} (y_k - H_k f_k(\xi_{k-1}^i))$$

et de matrice de covariance

$$Q_k^W - Q_k^W H_k^* [H_k Q_k^W H_k^* + Q_k^V]^{-1} H_k Q_k^W$$

et on met à jour

$$w_k^i = \frac{w_{k-1}^i q(y_k - H_k f_k(\xi_{k-1}^i), H_k Q_k^W H_k^* + Q_k^V)}{\sum_{j=1}^N w_{k-1}^j q(y_k - H_k f_k(\xi_{k-1}^j), H_k Q_k^W H_k^* + Q_k^V)}$$

## bilan

- chaque particule se déplace indépendamment comme une v.a. gaussienne
  - de moyenne égale au filtre de Kalman, initialisé à la position précédente de la particule
  - et de matrice de covariance égale à la matrice de covariance d'erreur du filtre de Kalman, initialisée à zéro
- les observations interviennent également après coup, en affectant un poids plus ou moins grand à chaque particule, mais en principe du même ordre de grandeur d'une particule à l'autre

---

## filtre particulaire SIR (avec ré-échantillonnage)

approximation particulaire sous la forme

$$p_k(x_k) = p(x_k | y_{0:k}) \approx \sum_{i=1}^N w_k^i \delta_{\xi_k^i}$$

injectée dans l'équation récurrente

$$p_k(x_k) \propto \int p(x_k, y_k | x_{k-1}) p_{k-1}(x_{k-1}) dx_{k-1}$$

donne

$$p_k(x_k) \propto \sum_{i=1}^N w_{k-1}^i p(x_k, y_k | \xi_{k-1}^i)$$

► avec la factorisation standard

$$p_k(x_k) \propto \underbrace{p(y_k | x_k)}_{g(x_k)} \underbrace{\sum_{i=1}^N w_{k-1}^i p(x_k | \xi_{k-1}^i)}_{q(x_k)}$$

d'où l'approximation

$$p_k(x_k) \approx \sum_{i=1}^N \frac{g(\xi_k^i)}{\sum_{j=1}^N g(\xi_k^j)} \delta_{\xi_k^i}$$

avec

$$g(x_k) = p(y_k | x_k)$$

et avec  $(\xi_k^1, \dots, \xi_k^N)$  un  $N$ -échantillon de densité (mélange fini)

$$q(x_k) = \sum_{i=1}^N w_{k-1}^i p(x_k | \xi_{k-1}^i)$$

## mise en œuvre pratique

- étape de sélection : indépendamment pour tout  $i = 1, \dots, N$ , on simule une v.a.  $I_k^i$  à valeurs dans les indices  $\{1, \dots, N\}$  selon les poids  $(w_{k-1}^1, \dots, w_{k-1}^N)$ , et on pose  $\hat{\xi}_{k-1}^i = \xi_{k-1}^{I_k^i}$
- étape de mutation : indépendamment pour tout  $i = 1, \dots, N$ , on simule une v.a.  $\xi_k^i$  gaussienne, de moyenne  $f_k(\hat{\xi}_{k-1}^i)$  et de matrice de covariance  $Q_k^W$
- étape de pondération : indépendamment pour tout  $i = 1, \dots, N$ , on pose

$$w_k^i = \frac{q(y_k - H_k \xi_k^i, Q_k^V)}{\sum_{j=1}^N q(y_k - H_k \xi_k^j, Q_k^V)}$$

## bilan

- chaque particule se déplace en partie indépendamment selon le modèle a priori, mais pas uniquement
- les observations interviennent à chaque étape en affectant un poids plus ou moins grand à chaque particule, en fonction de sa vraisemblance, et induisent un second mécanisme de déplacement des particules, par ré-échantillonnage

► avec la factorisation optimale

$$p_k(x_k) \propto \sum_{i=1}^N w_{k-1}^i p(y_k | \xi_{k-1}^i) p(x_k | \xi_{k-1}^i, y_k)$$

d'où l'approximation

$$p_k(x_k) \approx \frac{1}{N} \sum_{i=1}^N \delta_{\xi_k^i} \quad \text{c'est-à-dire que} \quad w_k^i = \frac{1}{N}$$

avec  $(\xi_k^1, \dots, \xi_k^N)$  un  $N$ -échantillon de densité (mélange fini)

$$\sum_{i=1}^N \frac{p(y_k | \xi_{k-1}^i)}{\sum_{j=1}^N p(y_k | \xi_{k-1}^j)} p(x_k | \xi_{k-1}^i, y_k)$$

mise en œuvre pratique

- étape de pondération : indépendamment pour tout  $i = 1, \dots, N$ , on pose

$$\pi_k^i = \frac{q(y_k - H_k f_k(\xi_{k-1}^i), H_k Q_k^W H_k^* + Q_k^V)}{\sum_{j=1}^N q(y_k - H_k f_k(\xi_{k-1}^j), H_k Q_k^W H_k^* + Q_k^V)}$$

- étape de sélection : indépendamment pour tout  $i = 1, \dots, N$ , on simule une v.a.  $I_k^i$  à valeurs dans les indices  $\{1, \dots, N\}$  selon les poids  $(\pi_k^1, \dots, \pi_k^N)$ , et on pose  $\hat{\xi}_{k-1}^i = \xi_{k-1}^{I_k^i}$
- étape de mutation : indépendamment pour tout  $i = 1, \dots, N$ , on simule une v.a.  $\xi_i$  gaussienne, de moyenne

$$f_k(\hat{\xi}_{k-1}^i) + Q_k^W H_k^* [H_k Q_k^W H_k^* + Q_k^V]^{-1} (y_k - H_k f_k(\hat{\xi}_{k-1}^i))$$

et de matrice de covariance

$$Q_k^W - Q_k^W H_k^* [H_k Q_k^W H_k^* + Q_k^V]^{-1} H_k Q_k^W$$

## bilan

- chaque particule se déplace indépendamment comme une v.a. gaussienne
  - de moyenne égale au filtre de Kalman, initialisé à la position précédente de la particule
  - et de matrice de covariance égale à la matrice de covariance d'erreur du filtre de Kalman, initialisée à zéro

mais pas uniquement

- les observations interviennent à chaque étape en affectant un poids plus ou moins grand à chaque particule, et induisent un second mécanisme de déplacement des particules, par ré-échantillonnage

---

## conclusion sur les filtres particuliers

- SIS (sans ré-échantillonnage) avec factorisation standard  
très inefficace
- SIR (avec ré-échantillonnage) avec factorisation standard  
beaucoup plus efficace  
applicable aux systèmes linéaires avec bruits non gaussiens
- SIR ou SIS avec factorisation optimale  
très efficace aussi, ressemble au filtre de Kalman d'ensemble  
limité au cas particulier des systèmes avec
  - condition initiale gaussienne, et bruits gaussiens additifs
  - fonction d'observation linéaire

---

**exemple : modèle de Lorenz**

Vu Duc Tran, Valérie Monbet, François LeGland, *Filtre de Kalman d'ensemble et filtres particulières pour le modèle de Lorenz*, MajecSTIC, Lorient, novembre 2006.

$$\dot{x}_t = 10(y_t - x_t) + \dot{w}_t^x$$

$$\dot{y}_t = 28x_t - y_t + x_t z_t + \dot{w}_t^y$$

$$\dot{z}_t = x_t y_t - \frac{8}{3} z_t + \dot{w}_t^z$$

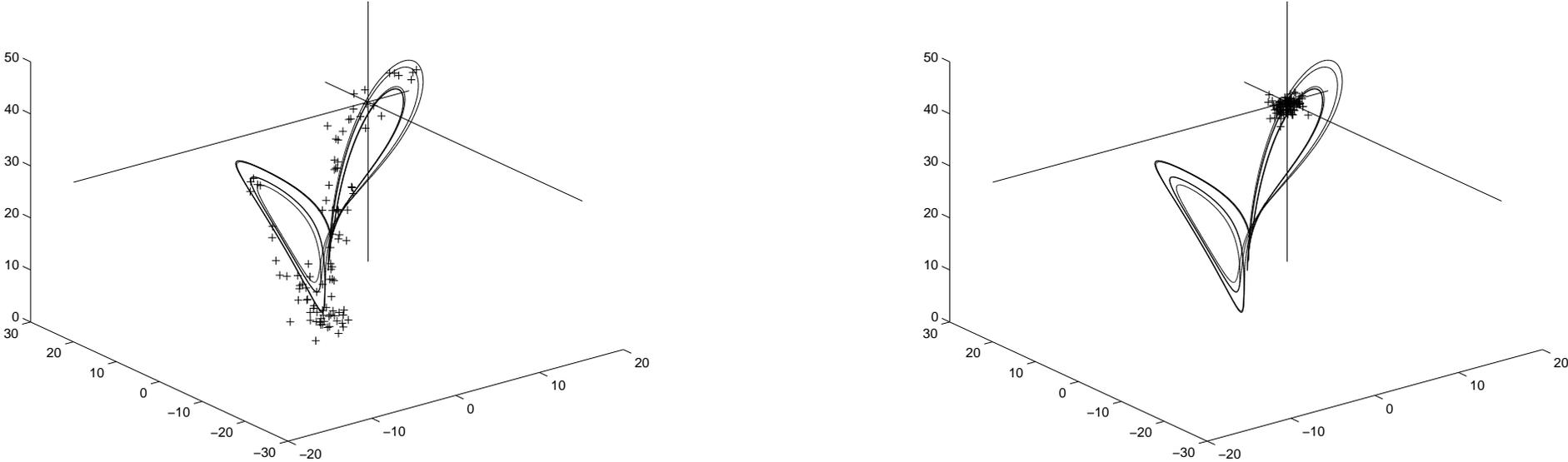


Figure 1: Filtre de Kalman d'ensemble : forecast (gauche) et analysis (droite)

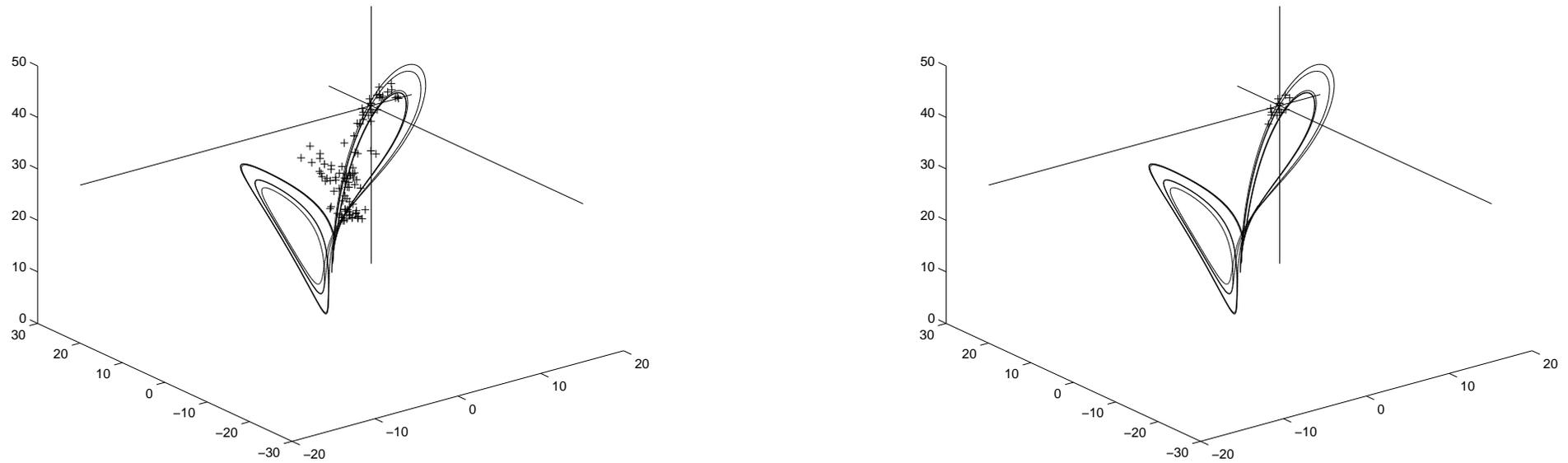


Figure 2: Filtre particulaire SIR (avec ré-échantillonnage) avec factorisation optimale : prédiction (gauche) et correction (droite)

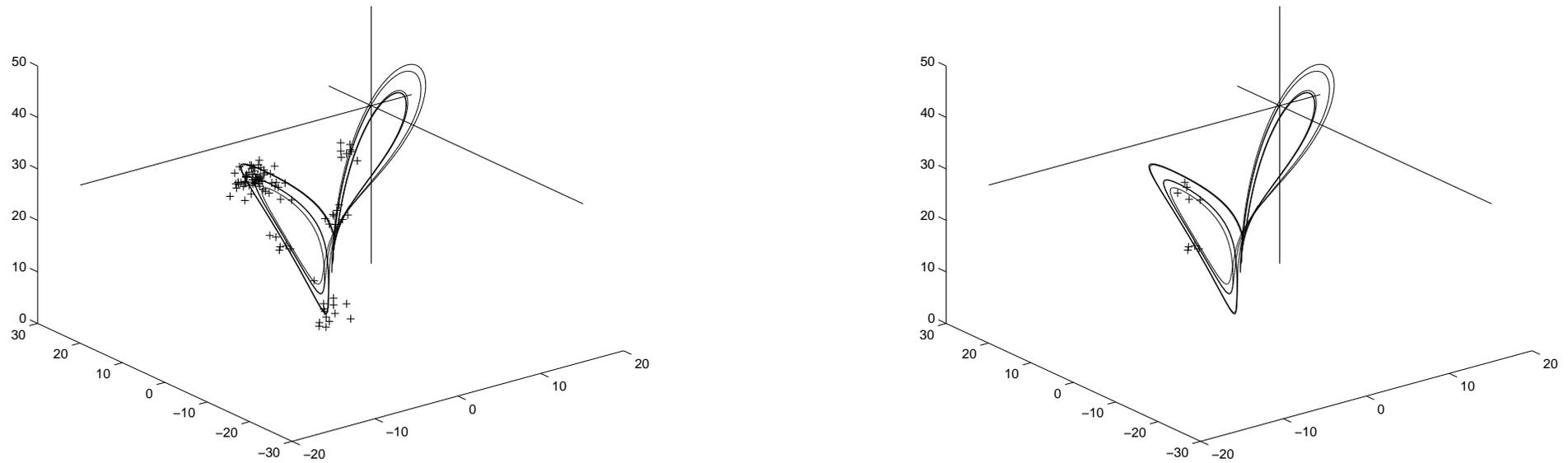


Figure 3: Filtre particulaire SIR (avec ré-échantillonnage) avec factorisation standard : prédiction (gauche) et correction (droite)

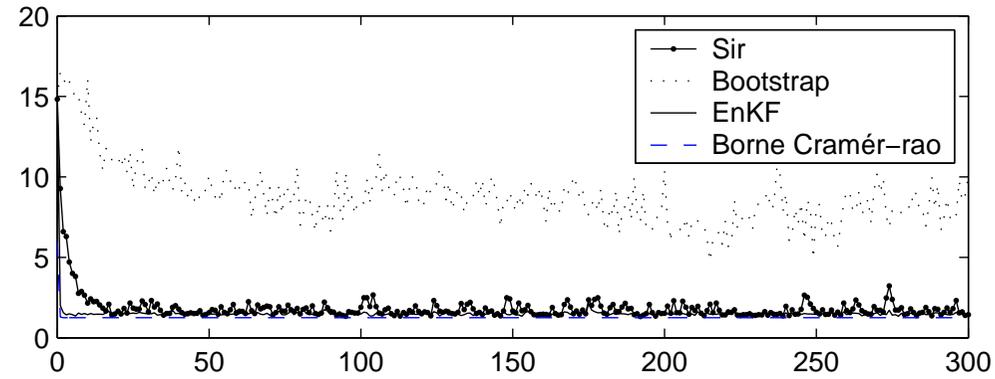


Figure 4: Comparaison des filtres : bruits faibles

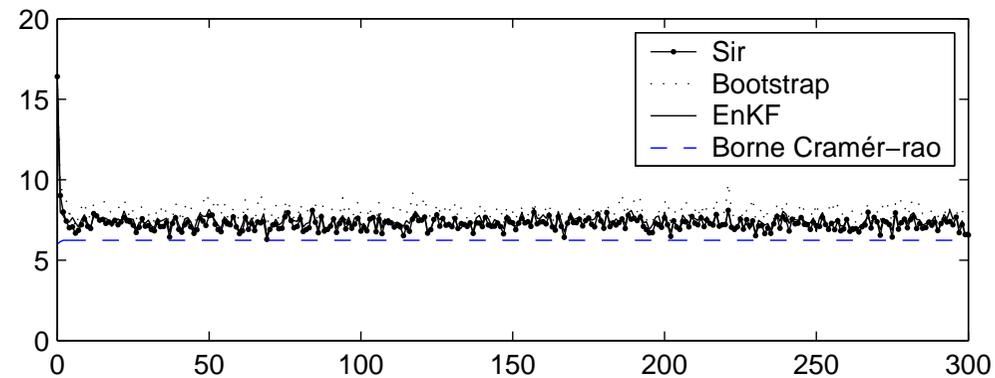


Figure 5: Comparaison des filtres : bruits élevés (facteur 5)

pour chacun des filtres particuliers (SIS / SIR, factorisation standard /optimale)

$$p(x_k | y_{0:k}) = p_k(x_k) \approx \sum_{i=1}^N w_k^i \delta_{\xi_k^i}$$

estimation d'erreur dans  $\mathbb{L}^p$  (non asymptotique)

$$\left\{ E \left| \int \phi(x_k) p_k(x_k) dx_k - \sum_{i=1}^N w_k^i \phi(\xi_k^i) \right|^p \right\}^{1/p} \leq \frac{c_n}{\sqrt{N}} \|\phi\|$$

où la constante  $c_n$  dépend de l'algorithme

théorème central limite (asymptotique)

$$\sqrt{N} \left( \int \phi(x_k) p_k(x_k) dx_k - \sum_{i=1}^N w_k^i \phi(\xi_k^i) \right) \Longrightarrow \mathcal{N}(0, V_n(\phi))$$

en distribution quand le nombre  $N$  de particules tend vers l'infini

la variance asymptotique  $V_n(\phi)$  dépend de l'algorithme et de la fonction test

problème ouvert pour le filtre de Kalman d'ensemble :  
comportement limite de la distribution empirique

$$\frac{1}{N} \sum_{i=1}^N \delta_{X_k^{a,i}} \approx ?$$

quand la taille  $N$  de l'ensemble tend vers l'infini, avant même de parler

- d'estimation d'erreur dans  $\mathbb{L}^p$
- ou de théorème central limite