Centre de Compétence Technique (CCT)
Centre National d'Études Spatiales (CNES)
Toulouse, 24 novembre 2003

Filtrage Particulaire
Introduction et Aspects Algorithmiques

François Le Gland + Fabien Campillo
IRISA (UMR 6074) / INRIA Rennes

{legland,campillo}@irisa.fr
http://www.irisa.fr/sigma2/

Plan

- introduction
- bootstrap filter (SIR), systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés (HMM) généraux
- filtre particulaire, SIR vs. SIS
- variantes algorithmiques
  - choix du noyau de mutation
  - stratégie de redistribution
  - marginalisation (Rao–Blackwellisation)
- contexte scientifique

méthodes particulaires

méthodes de *simulation* séquentielles, de type Monte Carlo

- des particules *explorent l'espace* d'état, en évoluant de manière indépendante comme le processus sous–jacent
- et *interagissent* sous l'effet d'un mécanisme de *sélection*, qui concentre automatiquement les particules (i.e. la *puissance de calcul*) dans les régions d'intérêt de l'espace d'état

très facile à mettre en œuvre : il suffit de simuler de manière indépendante des trajectoires / transitions du processus sous–jacent

analogie avec les algorithmes génétiques, les systèmes de particules en interaction, etc.

filtrage particulaire

méthode(s) particulaire(s) pour le filtrage

filtrage : estimation bayésienne récursive d'un état *caché* X_k

(par exemple, position, et vitesse d'un mobile)

à partir d'une suite d'*observations* $Y_{0:k} = (Y_0 \cdots Y_k)$

(par exemple, mesures d'angle, de distance, etc.)

filtrage particulaire : approximation numérique du filtre optimal par la distribution de probabilité empirique associée à un système pondéré de N particules, i.e.

$$\mathbb{P}[X_k \in dx \mid Y_{0:k}] \approx \sum_{i=1}^{N} w_k^i \; \delta_{\xi_k^i}$$

caractérisé par

- *positions* $(\xi_k^1 \cdots \xi_k^N)$
- et *poids* $(w_k^1 \cdots w_k^N)$ des particules

transition de μ_{k-1}^N à μ_k^N : pour chaque particule

- *mutation* ou *exploration* de l'espace d'état
  imitation d'une trajectoire / transition typique de l'état caché
- *sélection*
  cohérence (quantifiée par la fonction de *vraisemblance*) entre un point de l'espace d'état et l'observation courante

avantages du filtrage particulaire

- facilité de mise en œuvre
- possibilité de prendre en compte des modèles complexes (hybrides, avec contraintes, etc.)
- indépendance de la dimension

historique

- filtre pondéré séquentiel (Handschin et Mayne, 1970's)
- filtre particulaire (Del Moral, Rigal et Salut, 1992)
- bootstrap filter (Gordon, Salmond et Smith, 1993)
- Monte Carlo filter (Kitagawa, 1996)

nombreux domaines d'applications

- localisation, navigation et poursuite
- poursuite multi–cible
- vision
- robotique
- traitement du signal audio
- communications numériques

contributions provenant de plusieurs communautés scientifiques

- poursuite
- vision (algorithme CONDENSATION)
- statistique bayésienne, méthodes de Monte Carlo
- probabilités appliquées (systèmes de particules en interaction)

Plan

- introduction
- bootstrap filter (SIR), systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés (HMM) généraux
- filtre particulaire, SIR vs. SIS
- variantes algorithmiques
  - choix du noyau de mutation
  - stratégie de redistribution
  - marginalisation (Rao–Blackwellisation)
- contexte scientifique

systèmes non–linéaires / non–gaussiens

▶ modèle *a priori* pour évolution de l'état *caché*

$$X_k = f_k(X_{k-1}) + W_k \qquad \text{avec} \qquad W_k \sim p_k(w)\, dw$$

d'où la *probabilité de transition*

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = p_k(x' - f_k(x))\, dx'$$

il suffit de savoir *simuler*

$$X_0 \sim \mu_0(dx) \qquad \text{et} \qquad W_k \sim p_k(w)\, dw$$

plus généralement

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = Q_k(x, dx')$$

▶ relation entre *observation* et état *caché*

$$Y_k = h_k(X_k) + V_k \qquad \text{avec} \qquad V_k \sim q_k(v)\, dv$$

d'où la *probabilité d'émission*

$$\mathbb{P}[Y_k \in dy' \mid X_k = x'] = q_k(y' - h_k(x'))\, dy'$$

il suffit de *connaître* / savoir *calculer* la fonction de *vraisemblance*

$$\Psi_k(x') = q_k(Y_k - h_k(x'))$$

cohérence entre un état *possible* et l'observation *réelle*, e.g.

$$\Psi_k(x') = \exp\{-\tfrac{1}{2}\, |Y_k - h_k(x')|^2\}$$

plus généralement

$$\mathbb{P}[Y_k \in dy' \mid X_k = x'] = g_k(x', y')\, \lambda_k(dy')$$

$$\Psi_k(x') = g_k(x', Y_k)$$

bootstrap filter (SIR)

approximation numérique du filtre optimal

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}]$$

par un système pondéré de N particules, i.e.

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i}$$

caractérisé par

- *positions* $(\xi_k^1 \cdots \xi_k^N)$
- et *poids* $(w_k^1 \cdots w_k^N)$ des particules

transition de μ_{k-1}^N à μ_k^N : pour chaque particule

- *mutation* ou *exploration* de l'espace d'état
  imitation d'une trajectoire / transition typique de l'état caché

$$X_k = f_k(X_{k-1}) + W_k \qquad \text{avec} \qquad W_k \sim p_k(w)\, dw$$

- *sélection*
  cohérence, quantifiée par la fonction de *vraisemblance*

$$\Psi_k(x') = q_k(Y_k - h_k(x'))$$

entre un point de l'espace d'état et l'observation courante

- réééchantillonnage (sélection) : tirage aléatoire avec remise

$$\widehat{\xi}^i_{k-1} \sim \mu^N_{k-1} = \sum_{i=1}^N w^i_{k-1} \, \delta_{\xi^i_{k-1}}$$

e.g.

$$\tau^i_{k-1} \sim (w^1_{k-1} \cdots w^N_{k-1}) \qquad \text{et} \qquad \widehat{\xi}^i_{k-1} = \xi^{\tau^i_{k-1}}_{k-1}$$

indépendamment pour tout $i = 1 \cdots N$

- prédiction (mutation) :

$$\xi^i_k = f_k(\widehat{\xi}^i_{k-1}) + W^i_k \qquad \text{avec} \qquad W^i_k \sim p_k(dw)$$

indépendamment pour tout $i = 1 \cdots N$

- correction (pondération) :

$$w^i_k = q_k(Y_k - h_k(\xi^i_k)) \; / \; [\, \sum_{j=1}^N q_k(Y_k - h_k(\xi^j_k)) \,]$$

pour tout $i = 1 \cdots N$

effet du rééchantillonnage

- les particules de plus fort poids (vraisemblance) sont présentes en plus grand nombre à la génération suivante
- les particules de trop faible poids sont absentes à la génération suivante

si on n'effectue pas de rééchantillonnage

- prédiction (mutation) :

$$\xi^i_k = f_k(\xi^i_{k-1}) + W^i_k \qquad \text{avec} \qquad W^i_k \sim p_k(dw)$$

indépendamment pour tout $i = 1 \cdots N$

- correction (pondération) :

$$w^i_k = w^i_{k-1} \, q_k(Y_k - h_k(\xi^i_k)) \; / \; [\, \sum_{j=1}^N w^j_{k-1} \, q_k(Y_k - h_k(\xi^j_k)) \,]$$

pour tout $i = 1 \cdots N$

démo : importance de la redistribution

points faibles de l'algorithme sans rééchantillonnage

- les poids s'accumulent de façon multiplicative le long de chaque trajectoire simulée
- il n'y a pas d'interaction entre les différentes trajectoires simulées

dégénerescence des poids : trop d'importance accordée aux observations passées

- filtre à mémoire limitée
- filtre à oubli exponentiel
- choix d'une distribution d'importance (SIS)
- redistribution (SIR)

Plan

- introduction
- bootstrap filter (SIR), systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés (HMM) généraux
- filtre particulaire, SIR vs. SIS
- variantes algorithmiques
  - choix du noyau de mutation
  - stratégie de redistribution
  - marginalisation (Rao–Blackwellisation)
- contexte scientifique

modèles de Markov cachés (HMM) généraux

état caché $\{X_k\}$ chaîne de Markov à valeurs dans un espace

- fini $F = \{1 \cdots |F|\}$
- continu $E = \mathbb{R}^m$
- hybride continu / discret $E = \mathbb{R}^m \times \{1 \cdots |F|\}$
- avec contraintes
- dépendant du temps
- trajectoriel (croissant avec le temps)

$$\Xi_k = (X_0 \cdots X_k)$$

chaîne de Markov à valeurs dans $E_k = \underbrace{E \times \cdots \times E}_{(k+1) \text{ fois}}$

observations *conditionnellement indépendantes*

$$\begin{tikzcd}
\arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\
& Y_{k-1} & Y_k & Y_{k+1} &
\end{tikzcd}$$

exemple :

$$Y_k = h_k(X_k) + V_k$$

noyau de transition et probabilités d'émission

$$\begin{aligned}
\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] &= Q_k(x, dx') \\
\mathbb{P}[Y_k \in dy' \mid X_k = x'] &= g_k(x', y')\, \lambda_k(dy')
\end{aligned}$$

fonction de vraisemblance

$$\Psi_k(x') = g_k(x', Y_k)$$

observations *conditionnellement markoviennes*

$$\begin{tikzcd}
{} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\
{} \arrow[r] & Y_{k-1} \arrow[r] & Y_k \arrow[r] & Y_{k+1} \arrow[r] & {}
\end{tikzcd}$$

exemple : modèle AR à paramètre markovien

$$Y_k = h_k(X_k, Y_{k-1}) + V_k$$

noyaux de transition

$$\begin{aligned}
\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] \;&=\; Q_k(x, dx') \\
\mathbb{P}[Y_k \in dy' \mid Y_{k-1} = y, X_k = x'] \;&=\; g_k(x', y, y')\, \lambda_k^F(y, dy')
\end{aligned}$$

fonction de vraisemblance

$$\Psi_k(x') = g_k(x', Y_{k-1}, Y_k)$$

plus généralement : observations et états cachés *conjointement markoviens*

hypothèse sur le noyau de transition conjoint

$$\begin{aligned}
&\mathbb{P}[X_k \in dx', Y_k \in dy' \mid X_{k-1} = x, Y_{k-1} = y] \\
&\qquad\qquad = R_k(x, y, y', dx')\, \lambda_k^F(y, dy')
\end{aligned}$$

hypothèse équivalente

$$\mathbb{P}[Y_k \in dy' \mid X_{k-1} = x, Y_{k-1} = y] = R_k(x, y, y', E)\, \lambda_k^F(y, dy')$$

séparément, l'état caché n'est pas nécessairement markovien

noyau global (abus de notation)

$$R_k(x, dx') = R_k(x, Y_{k-1}, Y_k, dx')$$

inclus le cas des observations

- conditionnellement indépendantes
- ou conditionnellement markoviennes

comme cas particulier, avec le noyau global

$$R_k(x, dx') = Q_k(x, dx') \; \Psi_k(x')$$

combine *prédiction* et *correction* simultanément

plus généralement : factorisation *mutation / sélection* du noyau global

$$R_k(x, dx') = W_k(x, x') \; P_k(x, dx')$$

avec le choix arbitraire

- d'un noyau de mutation markovien $P_k(x, dx')$
- d'une fonction de poids $W_k(x, x')$ pour réaliser la sélection

exemple : en toute généralité

$$R_k(x, dx') = \underbrace{R_k(x, E)}_{\widehat{\Psi}_k(x)} \; \underbrace{\frac{R_k(x, dx')}{R_k(x, E)}}_{\widehat{Q}_k(x, dx')}$$

mise en garde : il faut

- savoir *simuler* selon le noyau de mutation $P_k(x, dx')$
- *connaître* / savoir *calculer* la fonction de sélection $W_k(x, x')$

bénéfice espéré : on peut choisir un noyau de mutation dépendant de la nouvelle observation de façon à forcer les particules à explorer les régions où la distribution de probabilité *a posteriori* est concentrée estimation bayésienne récursive :

équation récursive pour le filtre optimal

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}]$$

dans le cas général

$$\mu_{k-1} \longrightarrow \mu_k = \frac{\mu_{k-1}\ R_k}{(\mu_{k-1}\ R_k)(E)} = \bar{R}_k(\mu_{k-1})$$

par définition

$$\begin{aligned} \mu_{k-1}\ R_k(dx') &= \int_E \mu_{k-1}(dx)\ R_k(x, dx') \\ &= \{ \int_E \mu_{k-1}(dx)\ Q_k(x, dx')\ \}\ \Psi_k(x') \end{aligned}$$

dans le cas plus simple des HMM, des processus AR à paramètre markovien

$$\mu_{k-1} \xrightarrow{\text{prédiction}} \mu_{k|k-1} = \mu_{k-1}\ Q_k \xrightarrow{\text{correction}} \mu_k = \Psi_k \cdot \mu_{k|k-1} = \frac{\Psi_k\ \mu_{k|k-1}}{\langle \mu_{k|k-1}, \Psi_k \rangle}$$

Plan

- introduction
- bootstrap filter (SIR), systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés (HMM) généraux
- filtre particulaire, SIR vs. SIS
- variantes algorithmiques
  - choix du noyau de mutation
  - stratégie de redistribution
  - marginalisation (Rao–Blackwellisation)
- contexte scientifique

filtre particulaire avec interaction (SIR)

approximation particulaire pondérée

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i}$$

transporter exactement — à l'aide de l'opérateur d'évolution $\bar{R}_k(\cdot)$ du filtre optimal — l'approximation μ_{k-1}^N, est souvent impossible

idée : utiliser à la place une approximation particulaire pondérée de la distribution de probabilité $\bar{R}_k(\mu_{k-1}^N)$

$$\mu_{k-1} \longrightarrow \bar{R}_k(\mu_{k-1}) = \mu_k$$

$$\mu_{k-1}^N \longrightarrow \bar{R}_k(\mu_{k-1}^N) \approx \mu_k^N$$

transporter exactement l'approximation particulaire pondérée

$$\mu_{k-1} \approx \mu_{k-1}^N = \sum_{i=1}^N w_{k-1}^i \, \delta_{\xi_{k-1}^i}$$

à l'aide de l'opérateur d'évolution $\bar{R}_k(\cdot)$ du filtre optimal, donnerait à une constante de normalisation près

$$\begin{aligned} \bar{R}_k(\mu_{k-1}^N)(dx') &\propto \sum_{i=1}^N w_{k-1}^i \, R_k(\xi_{k-1}^i, dx') \\ &\propto \sum_{i=1}^N w_{k-1}^i \, W_k(\xi_{k-1}^i, x') \, P_k(\xi_{k-1}^i, dx') \qquad \text{où} \qquad \sum_{i=1}^N \pi_k^i = 1 \\ &\propto \sum_{i=1}^N \underbrace{\frac{w_{k-1}^i}{\pi_k^i} \, W_k(\xi_{k-1}^i, x')}_{r_{k-1}^i(x')} \, \underbrace{\pi_k^i \, P_{k-1}(\xi_{k-1}^i, dx')}_{m_{k-1}^i(dx')} \end{aligned}$$

marginale d'une distribution de probabilité sur l'espace produit $\{1 \cdots N\} \times E$

échantillonage pondéré : approximation Monte Carlo

$$\mu(dx) = \frac{r(x)\, m(dx)}{\displaystyle\int_E w(x)\; m(dx)} \propto r(x)\, m(dx)$$

densité $r(x)$ connue à une constante multiplicative près

$$I(\phi) = \int_E \phi(x)\; \mu(dx) = \frac{\displaystyle\int_E \phi(x)\; r(x)\; m(dx)}{\displaystyle\int_E r(x')\; m(dx')} \approx \frac{\displaystyle\sum_{i=1}^N \phi(\xi_i)\; r(\xi_i)}{\displaystyle\sum_{i=1}^N r(\xi_i)}$$

i.e.

$$\mu \approx \sum_{i=1}^N \frac{r(\xi_i)}{\displaystyle\sum_{j=1}^N r(\xi_j)}\; \delta_{\xi_i} \qquad \text{avec} \qquad \xi_i \sim m(dx)$$

indépendament pour tout $i = 1 \cdots N$

échantillonage pondéré sur l'espace produit $\{1 \cdots N\} \times E$

$$\sum_{i=1}^N \frac{r_{k-1}^{\tau_k^i}(\xi_k^i)}{\displaystyle\sum_{j=1}^N r_{k-1}^{\tau_k^j}(\xi_k^j)}\; \delta_{(\tau_k^i,\, \xi_k^i)}$$

avec

$$(\tau_k^i, \xi_k^i) \sim (m_{k-1}^1(dx') \cdots m_{k-1}^N(dx'))$$

puis marginalisation, d'où l'approximation particulaire pondérée

$$\bar{R}_k(\mu_{k-1}^N) = \frac{\mu_{k-1}^N\; R_k}{(\mu_{k-1}^N\; R_k)(E)} \approx \sum_{i=1}^N \frac{r_{k-1}^{\tau_k^i}(\xi_k^i)}{\displaystyle\sum_{j=1}^N r_{k-1}^{\tau_k^j}(\xi_k^j)}\; \delta_{\xi_k^i} = \mu_k^N$$

- rééchantillonnage (sélection) : tirage aléatoire avec remise

$$\widehat{\xi}_{k-1}^i \sim \sum_{i=1}^N \pi_k^i \, \delta_{\xi_{k-1}^i}$$

e.g.

$$\tau_k^i \sim (\pi_k^1 \cdots \pi_k^N) \qquad \text{et} \qquad \widehat{\xi}_{k-1}^i = \xi_{k-1}^{\tau_k^i}$$

indépendamment pour tout $i = 1 \cdots N$

- mutation :

$$\xi_k^i \sim P_k(\widehat{\xi}_{k-1}^i, dx')$$

indépendamment pour tout $i = 1 \cdots N$

- pondération :

$$w_k^i \; = \frac{w_{k-1}^{\tau_k^i}}{\pi_k^{\tau_k^i}} \, W_k(\widehat{\xi}_{k-1}^i, \xi_k^i) \; / \; [\sum_{j=1}^N \frac{w_{k-1}^{\tau_k^j}}{\pi_k^{\tau_k^j}} \, W_k(\widehat{\xi}_{k-1}^j, \xi_k^j)\,]$$

pour tout $i = 1 \cdots N$

sans rééchantillonnage : échantillonnage pondéré séquentiel (SIS)

- mutation :

$$\xi_0^i \sim \pi_0(dx) \qquad \text{et} \qquad \xi_k^i \sim P_k(\xi_{k-1}^i, dx')$$

indépendamment pour tout $i = 1 \cdots N$

- pondération : càd

$$w_k^i = w_{k-1}^i \, W_k(\xi_{k-1}^i, \xi_k^i) \; / \; [\sum_{j=1}^N w_{k-1}^j \, W_k(\xi_{k-1}^j, \xi_k^j)\,]$$

pour tout $i = 1 \cdots N$

Plan

- introduction
- bootstrap filter (SIR), systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés (HMM) généraux
- filtre particulaire, SIR vs. SIS
- variantes algorithmiques
  - choix du noyau de mutation
  - stratégie de redistribution
  - marginalisation (Rao–Blackwellisation)
- contexte scientifique

quelques cas particuliers (choix des noyaux de mutation
et des probabilités de sélection)

▶ bootstrap filter

$$P_k(x, dx') = Q_k(x, dx') \qquad W_k(x, x') = \Psi_k(x') \qquad \pi_k^i = w_{k-1}^i$$

d'où le nouveau poids

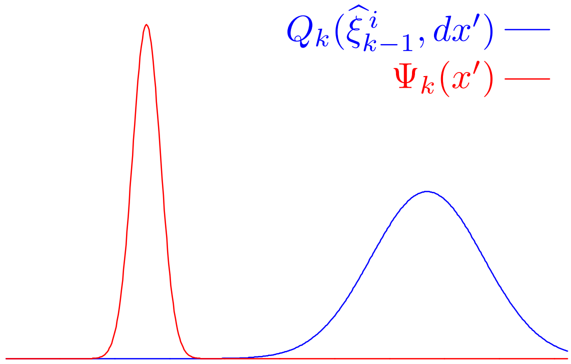
$$w_k^i = \Psi_k(\xi_k^i) \;/\; [\, \sum_{j=1}^N \Psi_k(\xi_k^j) \,]$$

pour tout $i = 1 \cdots N$

simuler des particules en aveugle peut entraîner une *dégénerescence des poids*, e.g. en cas de un mauvais accord entre distributions de probabilité a priori et fonction de vraisemblance, càd si

$$Q_k \; \Psi_k(\widehat{\xi}^i_{k-1}) = \int_E Q_k(\widehat{\xi}^i_{k-1}, dx') \; \Psi_k(x') \approx 0$$

pour la plupart des $i = 1 \cdots N$



▶ mutation guidée par les observations

$$P_k(x, dx') \;=\; \widehat{Q}_k(x, dx') = \frac{Q_k(x, dx') \; \Psi_k(x')}{Q_k \; \Psi_k(x)}$$

$$W_k(x, x') \;=\; \widehat{\Psi}_k(x) = Q_k \; \Psi_k(x)$$

dépend de l'observation future Y_k et ne dépend pas de x', donc peut être utilisé dans l'étape de sélection

$$\pi^i_k = \widehat{\Psi}_k(\xi^i_{k-1}) \; / \; [\; \sum_{j=1}^N \widehat{\Psi}_k(\xi^j_{k-1}) \;]$$

d'où le nouveau poids

$$w^i_k = w^{\tau^i_k}_{k-1} \; / \; [\; \sum_{j=1}^N w^{\tau^j_k}_{k-1} \;]$$

pour tout $i = 1 \cdots N$

bonne idée, mais il n'est en général pas facile

- de simuler des v.a. selon le noyau $\widehat{Q}_k(x, dx')$,
- ni de calculer la fonction de sélection $\widehat{\Psi}_k(x)$

sauf dans certains cas particuliers, e.g. bruits mélanges de v.a. gaussiennes et fonction d'observation linéaire

▶ particules auxiliaires (Pitt et Shephard, 1999)

$$P_k(x, dx') = Q_k(x, dx') \qquad W_k(x, x') = \Psi_k(x') \qquad \pi_k^i = \Psi_k(\alpha_k^i)$$

où l'état α_k^i est *représentatif* de la distribution de probabilité $Q_k(\xi_{k-1}^i, dx')$

$$Q_k \ \Psi_k(\xi_{k-1}^i) = \int_E Q_k(\xi_{k-1}^i, dx') \ \Psi_k(x') \approx \Psi_k(\alpha_k^i)$$

d'où le nouveau poids

$$w_k^i = w_{k-1}^{\tau_k^i} \, \frac{\Psi_k(\xi_k^i)}{\Psi_k(\alpha_k^i)} \ / \ [\, \sum_{j=1}^N w_{k-1}^{\tau_k^j} \, \frac{\Psi_k(\xi_k^j)}{\Psi_k(\alpha_k^j)} \,]$$

pour tout $i = 1 \cdots N$

autres approches proposées pour réduire la dégénerescence des poids

- annealed importance sampling (Neal, 2001), bridging densities (Godsill et Clapp, 2001)
- correction progressive (Oudjane, 2000)
- échantillonneur SMC (Del Moral et Doucet, 2002)

multiplier / éliminer des particules peut entrainer une *dégénerescence des positions* (e.g. si le bruit d'état est faible)

pour préserver la diversité et la capacité d'exploration du système de particules

- étape de *roughening* (Gordon, Salmond et Smith, 1993), ou de régularisation (Hürzeler et Künsch, 1998, Musso et Oudjane, 2001)
- itérations MCMC (Berzuini et Gilks, 2001)

Plan

- introduction
- bootstrap filter (SIR), systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés (HMM) généraux
- filtre particulaire, SIR vs. SIS
- variantes algorithmiques
  - choix du noyau de mutation
  - stratégie de redistribution
  - marginalisation (Rao–Blackwellisation)
- contexte scientifique

redistribution

dans l'étape de sélection, générer les v.a.

$$\widehat{\xi}_k^i = \xi_{k-1}^{\tau_k^i} \qquad \text{où} \qquad \tau_k^i \sim (w_{k-1}^1 \cdots w_{k-1}^N)$$

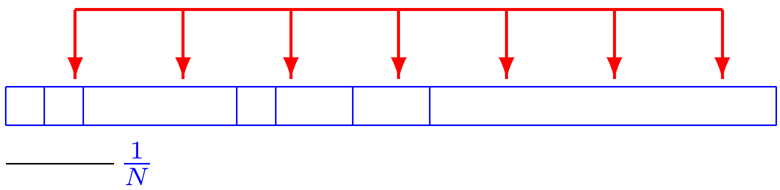
indépendamment pour tout $i = 1 \cdots n$, a pour effet de multiplier / éliminer les particules selon leur poids

si N_k^i désigne le nombre de descendants de la particule ξ_{k-1}^i à la génération suivante, alors $(N_k^1 \cdots N_k^N)$ suit une loi multinomiale

$$\mathbb{E}(N_k^i) = N\, w_{k-1}^i \qquad \text{et} \qquad \mathrm{var}(N_k^i) = N\, w_{k-1}^i\, (1 - w_{k-1}^i)$$

d'autres schémas de redistribution peuvent être utilisés, de façon à réduire la variance, quitte à perdre l'indépendance de l'échantillon généré

- échantillonnage résiduel (Liu et Chen, 1998)
- échantillonnage systématique (Carpenter, Clifford et Fearnhead, 1999, Kitagawa, 1996)



- branchement de Bernoulli (Crişan, Del Moral et Lyons, 1999)

les deux derniers schémas atteignent la variance minimale

$$\mathbb{E}(N_k^i) = N\, w_{k-1}^i \qquad \text{et} \qquad \mathrm{var}(N_k^i) = \bar{w}_{k-1}^i\, (1 - \bar{w}_{k-1}^i) < \tfrac{1}{4}$$

avec

$$\bar{w}_{k-1}^i = N\, w_{k-1}^i - \lfloor N\, w_{k-1}^i \rfloor < 1$$

redistribution adaptative :

- utilise l'algorithme SIS (sans redistribution)
- ne redistribue que si la distribution des poids $(w^1 \cdots w^N)$ est très déséquilibrée

critère de redistribution

- taille effective de l'échantillon (Kong, Liu et Wong, 1993)

$$\frac{1}{N} \sum_{i=1}^{N} (N\, w^i)^2 \geq c_{\mathrm{eff}} > 1$$

- entropie de l'échantillon (Pham, 2001)

$$\frac{1}{N} \sum_{i=1}^{N} (N\, w^i)\, \log(N\, w^i) \geq c_{\mathrm{ent}} > 0$$

Plan

- introduction
- bootstrap filter (SIR), systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés (HMM) généraux
- filtre particulaire, SIR vs. SIS
- variantes algorithmiques
  - choix du noyau de mutation
  - stratégie de redistribution
  - marginalisation (Rao–Blackwellisation)
- contexte scientifique

marginalisation (Rao–Blackwellisation)

réduction de dimension : permet de concentrer les particules dans un espace de dimension réduite (meilleure utilisation de la puissance de calcul)

condition d'application : existence d'une solution exacte au problème de filtrage

- conditionnellement aux observations
- et à certaines composantes du vecteur d'état

par exemple : système *conditionnellement* linéaire gaussien

$$\{r_k\} \text{ markovien}$$

$$\begin{aligned} X_{k+1} &= a_1(r_k)\, X_k + a_2(r_k) + W_k \\ Y_k &= c_1(r_k)\, X_k + V_k \end{aligned}$$

autre exemple

$$\begin{aligned} X_{k+1}^{\mathrm{L}} &= a_{11}(X_k^{\mathrm{NL}})\, X_k^{\mathrm{L}} + a_1(X_k^{\mathrm{NL}}) + W_k^{\mathrm{L}} \\ X_{k+1}^{\mathrm{NL}} &= a_{21}(X_k^{\mathrm{NL}})\, X_k^{\mathrm{L}} + a_2(X_k^{\mathrm{NL}}) + W_k^{\mathrm{NL}} \\ Y_k &= h(X_k^{\mathrm{NL}}) + V_k \end{aligned}$$

▶ si la composante $X_{0:k}^{\mathrm{NL}}$ était observée

$$\mathcal{L}(X_k^{\mathrm{L}} \mid X_{0:k}^{\mathrm{NL}}, Y_{0:k}) = \mathcal{L}(X_k^{\mathrm{L}} \mid X_{0:k}^{\mathrm{NL}})$$

loi gaussienne

▶ d'où la décomposition de la distribution de probabilité jointe

$$\mathcal{L}(X_k^{\mathrm{L}}, X_{0:k}^{\mathrm{NL}} \mid Y_{0:k}) = \mathcal{L}(X_k^{\mathrm{L}} \mid X_{0:k}^{\mathrm{NL}})\, \mathcal{L}(X_{0:k}^{\mathrm{NL}} \mid Y_{0:k})$$

et d'après la formule de Bayes

$$\mathcal{L}(X_{0:k}^{\mathrm{NL}} \mid Y_{0:k}) \propto \mathcal{L}(Y_{0:k} \mid X_{0:k}^{\mathrm{NL}})\, \mathcal{L}(X_{0:k}^{\mathrm{NL}})$$

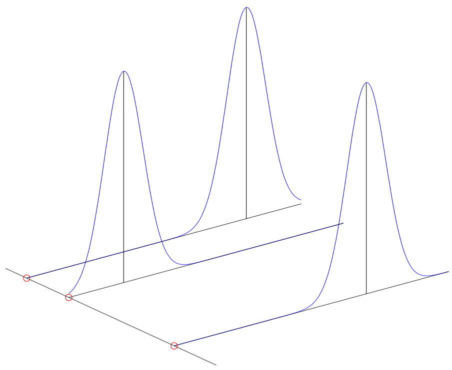
il suffit alors de considérer

- un système de particules évoluant dans le sous–espace correspondant à ces composantes du vecteur d'état
- et d'associer *à chaque particule* le filtre exact (par exemple filtre de Kalman) permetant d'estimer les autres composantes du vecteur d'état

la probabilité de mutation et le poids de sélection de chaque particule dépend du filtre exact qui lui est associé

(Chen et Liu, 2000, Doucet, Godsill et Andrieu, 2000, Andrieu et Doucet, 2002)

très utilisé dans les problèmes de navigation (Nordlund, 2002)



marginaliser

- implique des calculs plus complexes
- n'apporte pas d'amélioration significative dans l'espace de dimension réduite
- produit une bien meilleure approximation dans l'espace complémentaire (densité gaussienne vs. masse de Dirac : une précision équivalente nécessiterait bien plus qu'une unique masse de Dirac)

Plan

- introduction
- bootstrap filter (SIR), systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés (HMM) généraux
- filtre particulaire, SIR vs. SIS
- variantes algorithmiques
  - choix du noyau de mutation
  - stratégie de redistribution
  - marginalisation (Rao–Blackwellisation)
- contexte scientifique

résultats asymptotiques

approximation du filtre optimal μ_k (plus généralement, flot de mesure de probabilités) par la distribution de probabilité empirique μ_k^N d'un système de N *particules en interaction*

résultats d'approximation, quand N croît vers l'infini

- estimation dans L^p

$$\sqrt{N} \;\left[\, \mathbb{E}|\mu_k(\phi) - \mu_k^N(\phi)|^p \, \right]^{1/p} \leq C$$

- convergence presque sure
- théorème central limite

$$\sqrt{N} \left[\mu_k(\phi) - \mu_k^N(\phi) \right] \Longrightarrow \mathcal{N}(0, V_k(\phi))$$

- propriétés de grandes déviations
- stabilité du filtre optimal, approximation uniforme en temps

nombreux travaux de Del Moral, Guionnet et Miclo

contributions à l'IRISA

- variantes algorithmiques : régularisation, correction progressive (annealed importance sampling)
- stabilité du filtre optimal (oubli de la condition initiale)
  $\longrightarrow$ approximation uniforme en temps
- approximation particulaire des mesures signées
  $\longrightarrow$ génération de *résidus* pour la surveillance et le diagnostic
  $\longrightarrow$ *identification récursive*
  $\longrightarrow$ calcul de *sensibilité* (greeks)
- *importance splitting* / *importance sampling* pour la simulation d'évènements rares (application au trafic aérien, projet HYBRIDGE (IST) avec NLR et CENA)
- poursuite multi–cible et poursuite d'objets multiples dans une suite d'images (projet Vista et Microsoft Research, Cambridge)

vie / animation scientifique

trois projets soutenus par le CNRS depuis 1997

- *méthodes particulaires et filtrage non–linéaire*
  (programme Modélisation et Simulation Numérique)
- *HMM et filtrage particulaire* (Math–STIC)
- *méthodes particulaires* (AS 67 du département STIC)

workshops organisés à Toulouse (déc 97), à Rennes (juin 98), à Cambridge (déc 99), à Paris (juin 01)

journées thématiques du GdR ISIS (déc 02), journée applications (déc 03)

collaboration avec . . .
LSP Toulouse (Pierre Del Moral)
ONERA (these de Nadia Oudjane, 97–00)
LTCI / ENST (Eric Moulines)

extension à . . .
Cambridge University (Arnaud Doucet)
Microsoft Research, Cambridge (Patrick Pérez)
Linköping University (Fredrik Gustafsson)

références (pour en savoir plus)

- *Sequential Monte Carlo Methods in Practice*, ouvrage collectif coordonné par Arnaud Doucet, Nando de Freitas, Neil Gordon, Springer, 2001
- *Monte Carlo Methods for Statistical Signal Processing*, numéro spécial de IEEE Transactions on Signal Processing, coordonné par Petar Djurić et Simon Godsill, février 2002
- *Branching and Interacting Particle Systems. Approximations of Feynman–Kac Formulas with Applications to Non–Linear Filtering*, Pierre Del Moral, Laurent Miclo (article de synthèse au Séminaire de Probabilités XXXIV, Springer, 2000)
- *Simulation et Algorithmes Stochastiques*, Nathalie Bartoli, Pierre Del Moral, Cépaduès, 2001
- *Filtrage Particulaire*, François Le Gland, 19ème Colloque GRETSI sur le Traitement du Signal et des Images, Paris 2003
- *Monte Carlo Strategies in Scientific Computing*, Jun Liu, Springer, 2001

Sequential Monte Carlo Methods (SMC), à Cambridge
www-sigproc.eng.cam.ac.uk/smc/

ISIS project: Sensor Fusion, par Fredrik Gustafsson à Linköping
www.control.isy.liu.se/isis/projects/sensorfusion.html

Monte Carlo Localization (MCL), state estimation in mobile robotics using particle filters, par Dieter Fox à University of Washington
www.cs.washington.edu/ai/Mobile_Robotics/mcl/

Filtrage Particulaire, par Fabien Campillo à l'IRISA
www.irisa.fr/sigma2/campillo/site-pf/