

Une théorie champ moyen du filtrage non linéaire

P. Del Moral

INRIA Centre Bordeaux-Sud Ouest

Colloquium du lab. MAP5, Univ. Paris Descartes, Mars 2008

Outline

- 1 Une introduction au filtrage stochastique
- 2 Les équations du filtrage non linéaire
- 3 L'heuristique du filtrage particulaire
- 4 Modèles stochastiques
- 5 Quelques références

En qq mots :

- 2 processus aléatoires : X Signal \oplus Y Observation.
- Objectif \rightsquigarrow Calculer les lois conditionnelles de X sachant Y

$$\text{Loi}(X | Y)$$

Vocabulaire "équivalent" :

- *Assimilation de données (météo).*
- Chaines de Markov cachées (statistique fréquentiste).
- *Calcul de lois a posteriori (a priori=Loi(X))* (statistique bayésienne).
- *Filtrage (proba. appliquées).*

Modèles de Filtrage de processus stochastiques

- Signal=Processus stochastique

ingénierie/physique/biologique/économique :

- Cibles non coopérative (militaire : missile, char, avion,...).
- Physiques (fluides : tornades, cyclones, modèles océano, pression/température/coeff de diffusivité,...).
- Financiers (actifs boursiers, portefeuilles, volatilités,...).
- Signaux (parole, codages/transmission informations,...)

Dynamiques et aléas :

- Equations d'évolution physiques (Ex. $\sum_i u_i \vec{F}_i = \vec{A}$)
- Perturbations et aléas :
 - Erreurs de modèles \oplus Perturbations extérieures.
 - Contrôles et paramètres inconnus.
 \rightsquigarrow lois a priori \oplus (inconnues= réalisation de v.a.)

Modèles stochastiques de signaux (temps : $(n \in \mathbb{N}) \vee (t \in \mathbb{R}_+)$)

- **Equation d'évolution** : (\in espace d'états E)

$$X_n = F_n(X_{n-1}, W_n) \quad \text{ou} \quad dX_t = A_t(X_t) dt + B_t(X_t) dW_t$$

- **Interprétations et exemples** :

$E = \mathbb{R}, \mathbb{R}^d, \mathbb{R}^{\mathbb{Z}^2}$, espaces de trajectoires, excursions

$W_n =$ Erreurs de modèles, perturbations extérieures

Contrôles et paramètres inconnus.

Note :

$(E \rightsquigarrow E_n) \longrightarrow$ mod. traject. $X_n = (X'_0, \dots, X'_n)$ non exclus

Modèles de Filtrage de processus stochastiques

- Observations = Processus stochastique :

ingénierie/physique/biologique/économique :

- Ingénierie : Radar, Sonar, GPS, ...
- Physiques (capteurs de pression/température/...).
- Financiers (actifs boursiers, portefeuilles, volatilités,...).
- Statistique (données réelles : médecine, pharmaceutiques, politique, éco-finances,...).

Dynamiques et aléas :

- Observations partielles : mélanges, coordonnées partielles.
- Perturbations et aléas :
 - erreurs de mesures (bruits thermiques).
 - perturbations extérieures.
 - erreurs de modèle.

Modèles stoch. de capteurs/observations $((n \in \mathbb{N}) \vee (t \in \mathbb{R}_+))$

- **Equation d'observation:** (\in espace d'états E')

$$Y_n = H_n(X_n, V_n) \quad \text{ou} \quad dY_t = H_t(X_t) dt + J_t(X_t) dV_t$$

- **Interprétations et exemples :**

$E = \mathbb{R}, \mathbb{R}^d, \mathbb{R}^{\mathbb{Z}^2}$, espaces de trajectoires, excursions

$V_n =$ Erreurs de mesures/modèles/extérieures

Objectifs

- **Prédiction** : $m < n \rightsquigarrow \text{Loi}(X_n \mid (Y_0, \dots, Y_m))$.
- **Filtrage** : $m = n \rightsquigarrow \text{Loi}(X_n \mid (Y_0, \dots, Y_n))$.
- **Lissage** : $m < n \rightsquigarrow \text{Loi}(X_m \mid (Y_0, \dots, Y_n))$.

Remarques

- **Filtrage trajectoriel** :

$$\rightsquigarrow \text{Loi}((X_0, \dots, X_n) \mid (Y_0, \dots, Y_n))$$

- **Modèle trajectoriel = Filtrage "simple"** :

$$\mathcal{X}_n = (X_n, \dots, X_n) \in E_n \rightsquigarrow \text{Loi}(\mathcal{X}_n \mid (Y_0, \dots, Y_n))$$

Les équations du filtrage

2 Etapes d'apprentissage/d'adaptation naturelles :

$$\text{Loi}(X_n \mid (Y_0, \dots, Y_n)) \xrightarrow{\text{prédiction}} \text{Loi}(X_{n+1} \mid (Y_0, \dots, Y_n))$$

$$\text{Loi}(X_{n+1} \mid (Y_0, \dots, Y_n)) \xrightarrow{\text{correction}} \text{Loi}(X_{n+1} \mid ((Y_0, \dots, Y_n), Y_{n+1}))$$

QQ interprétations utiles :

- Prédiction = Examen de la stat. d'évol. $X_n \rightsquigarrow X_{n+1} (\perp Y)$
- Correction = Mise à jour des lois \sim novel obs. Y_{n+1}

\rightsquigarrow Potentiel de vraisemblance sur chaque site x :

$$G_{n+1}(x) = \text{"Proba de voir } Y_{n+1} \text{ si } X_{n+1} \text{ était en } x$$

L'heuristique du filtrage particulaire

Objectif

Dynamique de population de N "individus" /particules" t.q.

$$(\hat{\xi}_n^1, \dots, \hat{\xi}_n^N) \in E_n^N \rightsquigarrow \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\xi}_n^i} = \text{Loi}(X_n \mid (Y_0, \dots, Y_n))$$

2 Etapes d'apprentissage/d'adpatation naturelles :

$$(\hat{\xi}_n^1, \dots, \hat{\xi}_n^N) \in E_n^N \xrightarrow{\text{exploration}} (\xi_{n+1}^1, \dots, \xi_{n+1}^N) \in E_{n+1}^N$$

$$(\xi_{n+1}^1, \dots, \xi_{n+1}^N) \in E_{n+1}^N \xrightarrow{\text{adaptation}} (\hat{\xi}_{n+1}^1, \dots, \hat{\xi}_{n+1}^N) \in E_{n+1}^N$$

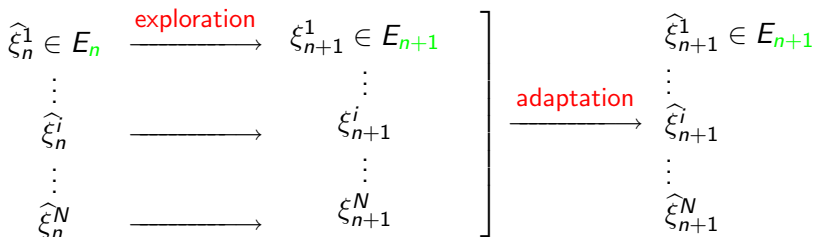
L'heuristique du filtrage particulaire

Interprétations \oplus Correspondances

(prédiction, correction) = (propositions, accept./rejet)
 = (mutation, selection)
 = (exploration, adaptation/branchements/ldots)

Méthode de simulation effective

- Prédiction \rightsquigarrow simulation de N transitions locales du signal
- Correction \rightsquigarrow processus de naissance et mort (taille fixe N)
 - On **tue** les individus sur des sites **peu vraisemblables**.
 - On **multiplie** les individus sur des sites **plus vraisemblables**.



- Exploration (\sim **simu. des aléas** des transitions locales)

(signal : $X_{n+1} = F(X_n, W_{n+1})$) $\rightsquigarrow \forall i \quad \xi_{n+1}^i = F(\widehat{\xi}_n^i, W_{n+1}^i)$

- Correction \rightsquigarrow *Exemple : branchements multinomiaux*

$(\widehat{\xi}_{n+1}^i)_{1 \leq i \leq N}$ "iid" de loi discrète : $\sum_{i=1}^N \frac{G_{n+1}(\xi_{n+1}^i)}{\sum_{j=1}^N G_{n+1}(\xi_{n+1}^j)} \delta_{\xi_{n+1}^i}$

Remarques importantes

Modèle trajectorien $\mathcal{X}_n = (X_0, \dots, X_n)$

- 1 particule=1 ligne ancestrale

$$\widehat{\xi}_n^i = (\widehat{\xi}_{0,n}^i, \widehat{\xi}_{1,n}^i, \dots, \widehat{\xi}_{n,n}^i)$$

$p \in \{0, \dots, n\} \rightarrow \widehat{\xi}_{p,n}^i$ ancêtre au niveau p de l'individu $\widehat{\xi}_{n,n}^i$

- les mesures d'occupation des arbres généalogiques :

$$\frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\widehat{\xi}_n^i} \simeq_{N \uparrow \infty} \text{Loi}((X_0, \dots, X_n) \mid (Y_0, \dots, Y_n))$$

Exemple :

$$\frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\widehat{\xi}_{0,n}^i} \simeq_{N \uparrow \infty} \text{Loi}(X_0 \mid (Y_0, \dots, Y_n))$$

Remarques importantes (suite)

Filtrage des bruits \rightsquigarrow Régulation optimale

$$\frac{1}{N} \sum_{1 \leq i \leq N} \delta_{(\widehat{W}_{0,n}^i, \widehat{W}_{1,n}^i, \dots, \widehat{W}_{n,n}^i)} \simeq_{N \uparrow \infty} \text{Loi}((W_0, \dots, W_n) \mid (Y_0, \dots, Y_n))$$

Algorithmes stoch."équivalents" :

- Algorithmes génétiques, processus de branchements.
- Algorithmes de "population Monte Carlo", DMC, QMC, ...
- Processus d'adaptation :
selection, pruning, reconfiguration, go with the winner, spawning, condensation, grouping, rejuvenations, harmony searches,...

En un certain sens :

La plupart des algorithmes stochastiques d'adaptation/d'apprentissage évolutionnaire sont des filtres optimaux (à population ∞)

Flot de mesures conditionnelles

- **Flot de mesures de probabilités aléatoires** : (ex.: prédicteurs)

$$n \in \mathbb{N} \mapsto \eta_n := \text{Loi}(X_n \mid (Y_0, \dots, Y_{n-1}))$$

- **Système dynamique non linéaire** : $\eta_n = \Phi_n(\eta_{n-1})$

$$\eta_n \xrightarrow{\text{Correction}} \hat{\eta}_n = \Psi_n(\eta_n) \xrightarrow{\text{Prédiction}} \eta_{n+1} = \hat{\eta}_n M_{n+1}$$

- **Transformation (non linéaire) de Boltzmann-Gibbs** :

$$\Psi_n(\eta_n)(dx) := \frac{1}{\eta_n(G_n)} G_n(x) \eta_n(dx) \quad (\text{"formule de Bayes"})$$

- **Transport linéaire markovien** :

$$(\hat{\eta}_n M_{n+1})(dy) := \int \hat{\eta}_n(dx) M_{n+1}(x, dy)$$

Chaînes de Markov non linéaires

- Chaînes de Markov non linéaires :

\exists Markov $K_{n,\eta}(x, dy)$ (non unique) t.q.

$$\Phi_n(\eta_{n-1}) = \eta_{n-1} K_{n,\eta_{n-1}} = \text{Loi}(\bar{X}_n) = \eta_n$$

i.e. :

$$\mathbb{P}(\bar{X}_n \in dx_n \mid \bar{X}_{n-1}) = K_{n,\eta_{n-1}}(\bar{X}_{n-1}, dx_n)$$

Interpretation particulière de type champ moyen :

- Chaîne Markov $\xi_n = (\xi_n^1, \dots, \xi_n^N) \in E_n^N$ t.q.

$$\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i} \underset{N \uparrow \infty}{\simeq} \eta_n$$

- Transitions locales "approchées" ($\forall 1 \leq i \leq N$)

$$\xi_{n-1}^i \rightsquigarrow \xi_n^i \sim K_{n,\eta_{n-1}^N}(\xi_{n-1}^i, dx_n)$$

Systemes de particules en interaction

Diagramme schématique : $\xi_n \in E_n^N \rightsquigarrow \xi_{n+1} \in E_{n+1}^N$

$$\begin{array}{ccc}
 \xi_n^1 & \xrightarrow{K_{n+1, \eta_n^N}} & \xi_{n+1}^1 \\
 \vdots & & \vdots \\
 \xi_n^i & \longrightarrow & \xi_{n+1}^i \\
 \vdots & & \vdots \\
 \xi_n^N & \longrightarrow & \xi_{n+1}^N
 \end{array}$$

Idée intuitive :

$$\begin{aligned}
 \eta_n^N \simeq_{N \uparrow \infty} \eta_n &\implies K_{n+1, \eta_n^N} \simeq_{N \uparrow \infty} K_{n+1, \eta_n} \\
 &\implies \xi_n^i \text{ copies iid de } \bar{X}_n
 \end{aligned}$$

QQ Avantages

- Modèle de champ moyen = **linéarisation/perturbation stoch.** :

$$\eta_n^N = \Phi_n(\eta_{n-1}^N) + \frac{1}{\sqrt{N}} W_n^N$$

avec $W_n^N \simeq W_n$ Champs gaussiens centrés et \perp .

- $\eta_n = \Phi_n(\eta_{n-1})$ stable \Rightarrow non propagation des erreurs locales

\Rightarrow **control uniforme des erreurs globales / temps**

- "Pas besoin" d'étudier la cv à l'équilibre de modèles MCMC.
- Grille/Maillage stochastique adaptatif.
- Non linéarité du syst. \rightsquigarrow interactions bénéfiques.
- Algo. naturel et facile à simuler, etc.

Modèles de Feynman-Kac

(Correction, Prédiction) = (Selection, Exploration) = (G_n, M_n)

$$\eta_n^N \simeq_{N \uparrow \infty} \eta_n = \Phi_n(\eta_{n-1}) := \Psi_{G_{n-1}}(\eta_{n-1}) M_n$$

Solution : X_n Markov \sim transitions M_n

$$\eta_n(f_n) = \frac{\gamma_n(f_n)}{\gamma_n(1)} \quad \text{with} \quad \gamma_n(f_n) = \mathbb{E} \left(f_n(X_n) \prod_{0 \leq p < n} G_p(X_p) \right)$$

Formule multiplicative \rightsquigarrow est. non biaisé des Cts. de normalisation

$$\mathbb{E} \left(\prod_{0 \leq p < n} G_p(X_p) \right) = \prod_{0 \leq p < n} \eta_p(G_p) \simeq_{N \uparrow \infty} \prod_{0 \leq p < n} \eta_p^N(G_p)$$

↑↑↑ Domaines d'application

QQ exemples :

- Modèles trajectoriels $X_n = (X'_0, \dots, X'_n)$

- $G_n(X_n) = 1_A(X'_n)$

$$\eta = \text{Loi}((X'_0, \dots, X'_n) \mid \forall 0 \leq p < n \quad X'_p \in A)$$

- $G_n(X_n) = 1_{\notin \{X'_0, \dots, X'_{n-1}\}}(X'_n)$

$$\eta_n = \text{Loi}((X'_0, \dots, X'_n) \mid \forall p < q < n \quad X'_p \neq X'_q)$$

- Transitions $X_n = (X'_n \overset{M}{\rightsquigarrow} X'_{n+1}) \oplus G(x_1, x_2) = \frac{\pi(dx_2)K(x_2, dx_1)}{\pi(dx_1)M(x_1, dx_2)}$

$$\eta_0 = \delta_x \implies \eta_n = \text{Loi}_{\pi, K}((X'_n, X'_{n-1}, \dots, X'_0) \mid X'_n = x)$$

- X_n MCMC mesure inv. instantannées $e^{-\beta_n V} d\lambda$

$$G_n(x) = e^{-(\beta_n - \beta_{n-1})V(x)} \implies \eta_n \propto e^{-\beta_n V} d\lambda$$

Theorie "asymptotique" TCL, PGD, PDM, ... (n, N). QQ exemples :

- Cv de processus empiriques :

$$\sup_{N \geq 1} \sup_{n \geq 0} \sqrt{N} \mathbb{E}(\|\eta_n^N - \eta_n\|_{\mathcal{F}_n}^p) < \infty$$

- Inégalités de concentration uniformes :

$$\sup_{n \geq 0} \mathbb{P}(|\eta_n^N(f_n) - \eta_n(f_n)| > \epsilon) \leq c \exp - \frac{N\epsilon^2}{2\sigma^2}$$

- Propagations du chaos : $\mathbb{P}_{n,q}^N := \text{Loi}(\xi_n^{(1,N)}, \dots, \xi_n^{(q,N)})$

$$\mathbb{P}_{n,q}^N \simeq \eta_n^{\otimes q} + \frac{1}{N} \partial^1 \mathbb{P}_{n,q} + \dots + \frac{1}{N^k} \partial^k \mathbb{P}_{n,q} + \frac{1}{N^{k+1}} \partial^{k+1} \mathbb{P}_{n,q}^N$$

$$\text{avec } \sup_{N \geq 1} \|\partial^{k+1} \mathbb{P}_{n,q}^N\|_{\text{tv}} < \infty, \|\partial^1 \mathbb{P}_{n,q}\|_{\text{tv}} \leq c q^2.$$

Quelques références

Méthodes particulières en physique/ingénierie stochastique

- **Mean field and Feynman-Kac particle models :**
 - Feynman-Kac formulae. Genealogical and interacting particle systems, Springer (2004) \oplus Refs.
- \supset **Sequential Monte Carlo models :**
 - joint work with Doucet A., Jasra A. Sequential Monte Carlo Samplers. JRSS B (2006).
 - joint work with A. Doucet. On a class of genealogical and interacting Metropolis models. Sémin. de Proba. 37 (2003).

Nouvelle génération d'algorithmes MCMC en interaction **PUB**

- **i-MCMC algorithms :**

- joint work with A. Doucet. Interacting Markov Chain Monte Carlo Methods For Solving Nonlinear Measure-Valued Eq., HAL-INRIA RR-6435, (Feb. 2008).
- joint work with B. Bercu and A. Doucet. Fluctuations of Interacting Markov Chain Monte Carlo Models. HAL-INRIA RR-6438, (Feb. 2008).
- joint work with C. Andrieu, A. Jasra, A. Doucet. *Non-Linear Markov chain Monte Carlo via self-interacting approximations*. Tech. report, Dept of Math., Bristol Univ. (2007).
- joint work with A. Brockwell and A. Doucet. *Sequentially interacting Markov chain Monte Carlo*. Tech. report, Dept. of Statistics, Univ. of British Columbia (2007).

↔ 1 Bourse de thèse INRIA-CORDI
Centre INRIA Bordeaux-Sud Oest