Particle filters and the Kushner-Stratonovich equation

Wilhelm Stannat
Fachbereich Mathematik
TU Darmstadt

stannat@mathematik.tu-darmstadt.de
www.mathematik.tu-darmstadt.de/~stannat

Taipei, March 26 2007

Contents

Introduction

Kushner-Stratonovich equation

Particle filters

Mutation/Selection algorithm

Asymptotic stability of the optimal filter

Stability in discrete time

Basic problem in signal processing

Estimate a noisy signal X ($\in S$)

$$Y = G(X, e) \tag{1}$$

- e - measurement error
- Y - observation ($\in \mathbb{R}^p$)

typical application positioning with GPS

- X - state of the system (eg position, velocity, acceleration)
- Y - measurement of the position of the vehicle through GPS

Further applications

tracking problems

From: Prof. Ji, RPI, New York

exploration

Prof. Fox, Univ. of Washington

Bayes Estimator

known in the examples:

- $P_X(dx)$ - **a-priori** distribution of X
- $P[Y \in dy \mid X = x]$ - the distribution of the meas. error

Ass $P[Y \in dy \mid X = x] = g(y, x)\, dy$

leads to the **a-posteriori** distribution via Bayes theorem

$$\begin{aligned}\eta^Y(dx) &= \frac{g(Y,x)P_X(dx)}{\int g(Y,\bar{x})\,P_X(d\bar{x})} \propto g(Y,x)P_X(dx)\\ &= \text{(regular) conditional distribution of X given Y}\\ &= P[X \in dx \mid Y = y]\end{aligned}$$

Bayes Estimator

$S \subset \mathbb{R}^d$

$$\hat{X} = \int x\eta^y(dx) \propto \int xg(Y,x)\,P_X(dx)$$
$$= \text{(regular) conditional expectation of } X \text{ given } Y$$
$$= E\,[X \mid Y = y]$$

minimizes the mean square error

$$\int \|x - \hat{X}\|^2\,d\eta^y(dx) = \min_{\alpha\in\mathbb{R}^d} \int \|x - \alpha\|^2\,d\eta^y(dx)$$

for general S

$$\int f(x)\,d\eta^Y(x) = E\,[f(X) \mid Y = y]\,, \qquad f \in \mathcal{B}_b(S)$$

Signals in continuous time

$$(S) \qquad dX_t = B(X_t)\,dt + C(X_t)\,dW_t \qquad \text{on } \mathbb{R}^d$$

- (W_t) Brownian motion
- generator of (X_t)

$$Af(x) = \frac{1}{2}\sum_{i,j=1}^{d}(CC^T)_{ij}(x)\partial_{ij}f(x) + \sum_{i=1}^{d} B_i(x)\partial_i f(x)$$

Additive measurement error

$$(O) \qquad Y_t = G(X_t) + e_t$$

- (e_t) Brownian motion, independent of $(X_t)_{t\geq 0}$

$$\begin{aligned}\eta_t &= \text{the a-posteriori distribution of (S) given } Y_{0:t} \\ &= \text{conditional distribution } P\left[X_t \in dx \mid Y_{0:t}\right]\end{aligned}$$

where $Y_{0:t} = (Y_s)_{0 \leq s \leq t}$

in particular

$$\int f \, d\eta_t = E\left[f(X_t) \mid Y_{0:t}\right]$$

Zakai equation

the a-priori distribution of the signal process is a solution of

$$d\eta_t = \hat{A}\eta_t \, dt \qquad \text{(Fokker-Planck)} \tag{2}$$

where $\hat{A}$ - dual operator of A

conditioned on the observation (Y_t) this changes to

$$d\tilde{\eta}_t = \hat{A}\tilde{\eta}_t \, dt + G^T \tilde{\eta}_t \, dY_t \qquad \text{(Zakai)} \tag{3}$$

Kushner-Stratonovich equation

Ito calculus yields the SPDE for the normalized cond distr

$$d\eta_t = \hat{A}\eta_t \, dt + \left(G^T - \langle G^T, \eta_t \rangle \right) \eta_t \, dZ_t$$

where

$$dZ_t = dY_t - \langle G, \eta_t \rangle \, dt$$

(innovations process)

the Zakai equation can be solved directly

$$\int f \, d\tilde{\eta}_t = E\left[f(X_t)\exp\left(\int_0^t G(X_s)\, dY_s - \frac{1}{2}\int_0^t \|G(X_s)\|^2\, ds\right)\right]$$

so that

$$\int f \, d\eta_t = \frac{E\left[f(X_t)\exp\left(\int_0^t G(X_s)\, dY_s - \frac{1}{2}\int_0^t \|G(X_s)\|^2\, ds\right)\right]}{E\left[\exp\left(\int_0^t G(X_s)\, dY_s - \frac{1}{2}\int_0^t \|G(X_s)\|^2\, ds\right)\right]} \tag{4}$$

(Kallianpur-Striebel formula)

Robust version of (4)

partial integration

$$\int_0^t G(X_s)\,dY_s = Y_t G(X_t) - \int_0^t Y_s\,dG(X_s) \qquad a.s.$$

and Ito's formula gives a representation of (4) as ren. Feynman-Kac semigroup

$$\int f\,d\eta_t = \frac{E\left[f(\bar{X}_t)\exp\left(Y_t G(\bar{X}_t)\right)\exp\left(\int_0^t G(Y_s,\bar{X}_s)\,ds\right)\right]}{E\left[\exp\left(Y_t G(\bar{X}_t)\right)\exp\left(\int_0^t G(Y_s,\bar{X}_s)\,ds\right)\right]}$$

where

- $(\bar{X}_t)$ - (inhomogeneous) Markov process with generator $A_t = A - Y_t\langle (CC)^T DG, \cdot\rangle$
- $G(Y,\cdot) = \exp\left(Y\cdot G(\cdot)\right) A\left(\exp\left(-Y\cdot G(\cdot)\right)\right) - \frac{1}{2}\|G(\cdot)\|^2$

Particle filters

fix observation $Y_{0:t}$

Basic idea approximate the a-posteriori distribution by means of the empirical distribution of a **system of weighted particles**

$$\eta_t \sim \eta_t^N = \sum_{i=1}^{N} \frac{w_t(i)}{\sum_{j=1}^{N} w_t(j)} \delta_{X_t(i)} \qquad \in \mathcal{M}_1(S)$$

where $X_t(i)$ position of the ith particle

in particular

$$\int f \, d\eta_t \sim \int f \, d\eta_t^N = \frac{\sum_{i=1}^{N} w_t(i) f(X_t(i))}{\sum_{j=1}^{N} w_t(j)}$$

Monte Carlo particle filter

consider N independent copies of $\bar{X}_t$

$$X_t^{(1)}, \ldots, X_t^{(N)}$$

define the unnormalized weights

$$w_t(i) = \exp\left(Y_t G(X_t^{(i)}) + \int_0^t G(Y_s, X_s^{(i)})\, ds \right)$$

so that

$$\int f \, d\eta_t^N = \frac{\sum_{i=1}^N \exp\left(Y_t G(X_t^{(i)}) + \int_0^t G(Y_s, X_s^{(i)})\, ds \right) f(X_t(i))}{\sum_{i=1}^N \exp\left(Y_t G(X_t^{(i)}) + \int_0^t G(Y_s, X_s^{(i)})\, ds \right)}$$

Rem

- **LLN** $\eta_t^N \to \eta_t$ (mean square convergence)
- **CLT** $\sqrt{N}\left(\int f\, d\eta_t^N - \int f\, d\eta_t\right) \xrightarrow{w} N(0, V_t(f))$

for suitable f and

$$V_t(f) = Var_P\left(\frac{w_t}{\int w_t\, dP}\left(f(\bar{X}_t) - \eta_t(f)\right)\right)$$

Moran particle approximation

for $N \geq 2$ consider N independent copies

$$X_t^{(1)}, \dots, X_t^{(N)}$$

Interaction

- every particle $X_t^{(i)}$ has exp. life time with rate $\frac{1}{N}$
- once a particle *dies* it is replaced by a copy of one of the remaining $N-1$ particles with probability prop. to $G(Y_t, X_t^{(j)})$

gives a new stochastic process

$$Y_t^{(1)}, \dots, Y_t^{(N)}$$

take $\hat{\eta}_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{Y_t^{(i)}}$ as an approximation of

$$\hat{\eta}_t \propto \exp\left(-Y_t G(\bar{X}_t)\right) \eta_t$$

Generator

$$L_t^N f(x) = \underbrace{A_t^N f(x)}_{\text{indep. mutation}} + \frac{1}{N} \sum_{i,j=1}^{N} G(Y_t, x_i)(\underbrace{\Phi_{ij}^{(N)} f(x)}_{\text{repl. operator}} - f(x))$$

where $\Phi_{ij}^{(N)} f(x_1, \ldots, x_N) = f(x_1, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_N)$

eg $\quad \Phi_{12}^{(2)} f(x_1, x_2) = f(x_1, x_1), \qquad \Phi_{21}^{(2)} f(x_1, x_2) = f(x_2, x_2)$

Theorem [Del Moral, Miclo 2000]

- A "regular" (eg $CC^T \in C_p^4(\mathbb{R}^d)$, $B \in C_p^3(\mathbb{R}^d)$)
- G bounded
- $(Y_0^{(i)})$ iid (μ)

then

- **LLN** $\hat{\eta}_t^N \longrightarrow \hat{\eta}_t$ (mean square convergence)
- **CLT** $\sqrt{N}\left(\int f\, d\hat{\eta}_t^N - \int f\, d\hat{\eta}_t\right) \xrightarrow{w} N(0, V_t(f))$
  for suitable f

$$V_t(f) = \mathrm{Var}_{\hat{\eta}_t}(f) + 2\int_0^t \int f_{s,t}^2\, G(Y_s, \cdot)\, d\hat{\eta}_s\, ds$$

where

$$f_{s,t} = \frac{E_{s,x}\left[f(\bar{X}_t)\exp\left(\int_s^t G(Y_u, \bar{X}_u)\, du\right)\right]}{\int E_{s,\bar{x}}\left[\exp\left(\int_s^t G(Y_u, \bar{X}_u)\, du\right)\right]\, d\hat{\eta}_s(\bar{x})}$$

Long-time behaviour of $\hat{\eta}_t$ versus $\hat{\eta}_t^N$

for fixed N

- Monte Carlo filter inherits the long-time behaviour of the signal process (X_t)
- in general **different** from the long-time behaviour of the optimal filter
- limits $N \to \infty$ and $t \to \infty$ cannot be interchanged
- question mainly open for the Moran particle filter

Mutation/Selection-algorithm

- $\mathbb{R}^d$ can be replaced by general S
- A_t can be replaced by general time-dep Markov generator
- $G(Y_t, x)$ can be replaced by

$$G_t : S \times M_1(S) \to S$$

fitness $G_t(x, \mu)$ of an individual of type x may depend on μ

Theorem 1 ([St '06]) Assume that

- S measurable space
- A_t is bounded
- G_t is bounded
- $|G_t(x,\mu) - G_t(x,\nu)| \leq \sum_{i=1}^m |\langle k_i,\mu\rangle - \langle k_i,\nu\rangle|$ $\qquad$ k_i bdd
- $(Y_0^{(i)})$ iid (μ)

then

$$E\left[\left(\int f\, d\hat{\eta}_t^N - \int f\, d\hat{\eta}_t\right)^2\right] \leq \frac{c_t}{N}\|f\|_\infty^2 \qquad \forall f \in \mathcal{B}_b(S)$$

where

$$\frac{d}{dt}\hat{\eta}_t = \hat{A}_t\hat{\eta}_t + \left(G_t - \int G_t\, d\hat{\eta}_t\right)\hat{\eta}_t\,, \qquad \hat{\eta}_0 = \mu$$

Rem

- no representation as ren. Feynman-Kac semigroup as in the noninteracting case
- rate of convergence independent of dimension
- c_t typically exponentially increasing with t

Long-time behaviour - the variational approach

- $A_t = A$ ν-symmetric
- $G_t = G$ noninteracting ($G(x, \mu) = G(x)$)

$A = 0$ η_t optimizes the mean fitness

$$\int G \, d\eta_t$$

$A \neq 0$ Substitution for the mean fitness

$$\lambda(h) := -\int Ah \, h \, d\nu - \int Gh^2 \, d\nu \, , h \in S^1$$

where $\mu = h\nu$ with $h \in L^2(\nu)$

In fact consider the L^2-normalization

$$\frac{d}{dt}h_t = Ah_t + Gh_t - \lambda(h_t)h_t \, , h_0 = h \tag{5}$$

so that

$$\frac{h_t \, \nu}{\int h_t \, d\nu} = \hat{\eta}_t$$

Then

- $\lambda(h_t)$ decreases in time
- critical points of λ coincide with stationary points of $\hat{\eta}_t$

Interacting case

$\bar{G} : L^1(\nu) \to \mathbb{R}$ Frechet differentiable with differential G

$$h \mapsto G(\cdot, h^2)h \qquad \text{loc Lipschitz in } L^2(\nu)$$

$\Longrightarrow$ (5) has a unique classical solution for all $h \in D(-(-A)^{\frac{1}{2}}) \cap S^1$

Define

$$L_1 := \sup_{h,g \in S^1} \|G(\cdot, h^2) - G(\cdot, g^2)\|_\infty$$

$$L_2 := \sup_{h,g \in S^1_+} \frac{\int \left(G(\cdot, h^2) - G(\cdot, g^2)\right) h(h-g)\, d\nu}{\|h-g\|^2_{L^2(\nu)}}$$

Theorem 2 ([St '04]) Let $h_* \geq 0$ be a critical point of

$$F(h) := -\int Ah\,h\,d\nu - \bar{G}(h^2)\,,\, h \in S^1_+$$

and assume that the ground state transform

$$A_* f = \frac{1}{h_*}\,(A + G + \lambda(h_*))\,(h_* f) = \frac{1}{h_*}\,(A(h_* f) - Ah_* f)$$

has a mass gap of size κ_* in $L^2(h_*^2 \nu)$, then

$$\|h_t - h_*\|^2 \leq \frac{1}{\int hh_*\,d\nu} e^{(L_1 + L_2 - \kappa_*)t} \|h - h_*\|^2_{L^2(\nu)}$$

Rem

- in the noninteracting case $L_1 = L_2 = 0$ so that

$$\|h_t - h_*\|^2 \leq \frac{1}{\int hh_*\, d\nu} e^{-\kappa_* t} \|h - h_*\|^2_{L^2(\nu)}$$

- Th 2 implies: if $\kappa_* > L_1 + L_2$ then F has exactly one critical point which coincides with the minimizer of F
- well-known: if h_* is the unique ground state of $A + G$ then $\kappa_* = \lambda_1 - \lambda_0$ where

$$\lambda_0 = \lambda(h_*) = \inf_{h \in S^1} \lambda(h) \qquad \lambda_1 = \inf_{h \in S^1 : \int h\, h_*\, d\nu = 0} \lambda(h)$$

Asymptotic stability of the optimal filter

Problem

η_t depends on the initial distribution P_{X_0} of (S)

Unknown!

hence interested in stability of η_t wrt P_{X_0}

Heuristic

(X_t) ergodic (dh $\lim_{t\to\infty} E\left[f(X_t)\right] = \int f \, d\nu \; \forall f \in C_b(S)$)

$\Rightarrow X_t$ forgets initial distribution P_{X_0} for large t

$\Rightarrow \eta_t$ forgets P_{X_0} for large t too

$\Rightarrow \eta_t$ stable

corresponding work (for compact state space) by

- Kunita, Stettner
- Ocone, Pardoux (extensions to the linear case)
- da Prato, Malliavin, Fuhrmann (pde-methods)
- Zeitouni, Atar
- Del Moral, Miclo, Le Gland (via ultracontractivity)

However ergodicity not necessary for asymptotic stability

Stability - deterministic case

$$\text{(S)} \qquad dX_t = B(X_t)\,dt \qquad \text{on } \mathbb{R}^d$$

$$\text{(O)} \qquad Y_t = G(X_t) + e_t \qquad \text{on } \mathbb{R}^p$$

where $(e_t)_{t\geq 0}$ is a Brownian motion

$$\int f\,d\eta_t = \frac{\int f(x) w_t(Y_{0:t}, x)\,\mu(dx)}{\int w_t(Y_{0:t}, x)\,\mu(dx)}$$

$$\begin{aligned} w_t(Y_{0:t}, x) = \exp(&-\frac{1}{2}\int_0^t \|G(X_s(x)) - G(X_s(X_0))\|^2\,ds \\ &+ \frac{1}{2}\int_0^t \|G(X_s(X_0))\|^2\,ds + \int_0^t G(X_s(x))\,de_s) \end{aligned}$$

Stability - time-reversible case

$$\text{(S)} \qquad dX_t = \frac{\nabla\varphi}{\varphi}(X_t)\,dt + dW_t \qquad \text{on } \mathbb{R}^d$$

$$\text{(O)} \qquad dY_t = GX_t\,dt + de_t\,, \quad Y_0 = 0 \qquad \text{on } \mathbb{R}^p$$

where

- $(W_t)_{t\geq 0}$, $(e_t)_{t\geq 0}$ ind. Brownian motions
- $\varphi \sim \varphi_0$ with φ_0 bounded and log-concave
- $W(x) := \|Gx\|^2 + \frac{\Delta\varphi}{\varphi}(x)$ uniformly strictly convex

$$\exists \quad \kappa_* > 0 \text{ with } W'' \geq \kappa_*^2 \cdot I$$

Rem (Brascamp, Lieb 76) Let

$$\Longrightarrow \qquad Af(x) - \frac{1}{2}\|Gx\|^2 f(x)$$

has a unique ground state h_* and spectral gap $\geq \kappa_*$ below λ_0

Theorem 3 [St '04] Assume that

- $Y \in C([0,\infty[;\mathbb{R}^p),\ Y(0)=0$
- $(P_{X_0})_i \ll \nu_* = \frac{h_* \, \nu}{\int h_* \, d\nu}$ with bounded densities, uniformly bdd away from 0

then

$$\limsup_{t\to\infty} e^{\frac{\kappa_*}{2}t} \|\hat{\eta}^1_t - \hat{\eta}^2_t\|_{var} < \infty$$

Rem

- rate *independent* of Y
- precise quantitative estimates
- stability also for nonergodic signals

Stability in discrete time

$$
\begin{aligned}
(S) \qquad & X_t = B(X_{t-1}) + C(X_{t-1})\, W_t \quad \text{on } \mathbb{R}^d \\
(O) \qquad & Y_t = G(X_t) + \varepsilon\, \tilde{W}_t \quad \text{on } \mathbb{R}^d
\end{aligned}
$$

where

- (W_t) iid $N(0, I)$
- $(\tilde{W}_t)$ iid, $P_{\tilde{W}_t} = r\, dx$, r log-concave
- B bounded or Lipschitz
- G bijective, G and G^{-1} Lipschitz

$$\eta_t^{Y_{1:t}} = \text{ a-posteriori distribution of } X_t \text{ given } Y_{1:t} = (Y_s)_{1\leq s\leq t}$$

$$\Longrightarrow \quad \int f\, d\eta_t^{Y_{1:t}} = \frac{E\left[f(X_t)\prod_{s=0}^{t} g(Y_s, X_s)\right]}{E\left[\prod_{s=1}^{t} g(Y_s, X_s)\right]} = \frac{\int P(fg(Y_t,\cdot))\, d\eta_{t-1}^{Y_{1:t-1}}}{\int P(g(Y_t,\cdot))\, d\eta_{t-1}^{Y_{1:t-1}}}$$

consequently

$$\eta_t^{Y_{1:t}} = T(Y_t,\cdot) \circ T(Y_{t-1},\cdot) \circ \ldots \circ T(Y_1, P_{X_0})$$

with the (nonlinear) transfer operator

$$T(Y,\eta)(dx) \propto \int g(Y,x) P(\bar{x}, dx)\, d\eta(\bar{x})$$

Theorem 4 [St '06] There exists $\varepsilon_0 > 0$, s.t. for $|\varepsilon| < \varepsilon_0$

$$\limsup_{t \to \infty} \frac{1}{t} \log \| \eta_{t,1}^{Y_{1:t}} - \eta_{t,2}^{Y_{1:t}} \|_{var} < 0 \qquad a.s.$$

for initial conditions $(P_{X_0})_i$, $i = 1, 2$, with

$$\int \|B(x)\|^2 \, (P_{X_0})_i \, (dx) < \infty$$

Rem essential restriction: dim $Y_t =$ dim X_t