Stability of the pathwise filter equation - a variational approach

Imperial College, February 24, 2005

Wilhelm Stannat
Fakultät für Mathematik
Universität Bielefeld
stannat@mathematik.uni-bielefeld.de

www.mathematik.uni-bielefeld.de/~stannat

1. A nonlinear problem from filtering theory

(S) $$dX(t) = B(X(t))\,dt + \underbrace{C\,dW(t)}_{\textit{systematic error}} \qquad \text{on } \mathbb{R}^d$$

(O) $$dY(t) = GX(t)\,dt + \underbrace{d\tilde{W}(t)}_{\textit{measurement error}}, \quad Y(0) = 0 \qquad \text{on } \mathbb{R}^p$$

Problem Estimate $X(t)$ given $Y(s)$, $s \in [0, t]$, up to time t

For simplicity

$(W(t))_{t\geq 0}$, $(\tilde{W}(t))_{t\geq 0}$ independent Brownian motions

i.e., for $0 \leq t_0 < t_1 \ldots < t_n$ the increments

$$W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1})$$

- **independent**
- **normally distributed**

$$P\left[W(t_i) - W(t_{i-1}) \in dx\right] = \underbrace{\frac{1}{(2\pi(t_i - t_{i-1}))^{\frac{d}{2}}} \exp\left(-\frac{\|x\|^2}{2(t_i - t_{i-1})}\right) \, dx}_{\mathrm{N}\,(0, (t_i - t_{i-1}) \cdot I)(dx)}$$

Classical (least square method) compute

$$\eta_t(A) := E\left[1_A\left(X(t)\right)|\mathcal{Y}(t)\right] \text{ for measurable } A \subset \mathbb{R}^d$$
$$= \text{ cond. distribution of } X(t) \text{ given } \mathcal{Y}(t) := \sigma\left\{Y(s)|s \in [0,t]\right\}$$

Note η_t depends on distr. μ_0 of the signal $X(0)$
(which is unknown!)

Hence interested in the stability of η_t w.r.t. μ_0

In the following write

$$E_{\mu_0}\left[1_A\left(X(t)\right)|\mathcal{Y}(t)\right]$$

to indicate dependence on μ_0

Heuristic

$X(t)$ ergodic (i.e., $\lim_{t\to\infty} E_{\mu_0}[f(X(t))] = \int f\, d\nu$ for some distr. ν)

$\Longrightarrow X(t)$ *forgets* μ_0 for large t

$\Longrightarrow \eta_t$ should also forget μ_0 for large t

$\Longrightarrow$ stability

corresponding work (for compact state space) by

- Kunita, Stettner
- Ocone, Pardoux (extensions to the linear case)
- da Prato, Malliavin, Fuhrmann (log. derivative of heat-kernel)
- Zeitouni, Atar (introducing new metric: Hilbert distance)

However ergodicity of the signal *not necessary* for stability

The pathwise filter equation

for simplicity $B = 0$, $C = c \cdot I$ and $G = I$, i.e.

(S) $$X(t) = X(0) + c\,W(t)$$

(O) $$dY(t) = X(t)\,dt + d\tilde{W}(t)$$

with no observation the distr. $X(t)$ satisfies

$$d\eta_t = \frac{c^2}{2}\Delta\eta_t\,dt \qquad \textit{(Fokker-Planck)}$$

given the observation $Y(\cdot)$ this changes to

$$d\eta_t = \frac{c^2}{2}\Delta\eta_t\,dt + x\eta_t\,dY(t) \qquad \textit{(Zakai)}$$

(unnormalized version)

$$\mu_t^Y := e^{-Y(t)\cdot x}\eta_t$$

Ito-calculus gives

$$d\mu_t^Y = \underbrace{e^{-Y(t)\cdot x} d\eta_t - e^{-Y(t)\cdot x} x\eta_t\, dY(t)}_{\text{usual product rule}} \quad - \quad \underbrace{\frac{1}{2}\|x\|^2 e^{-Y(t)\cdot x}\eta_t\, dt}_{\text{correction term}}$$

$$= \quad e^{-Y(t)\cdot x}\frac{c^2}{2}\Delta\eta_t\, dt \quad - \quad \frac{1}{2}\|x\|^2 e^{-Y(t)\cdot x}\eta_t\, dt$$

$$= \left(\frac{c^2}{2}\Delta\mu_t^Y + c^2\, Y(t)\cdot\nabla\mu_t^Y + \frac{c^2}{2}\|Y(t)\|^2\mu_t^Y\right) dt - \frac{1}{2}\|x\|^2\mu_t^Y\, dt$$

$$= \left(\hat{A}_t^Y \mu_t^Y + \sigma_t^Y \mu_t^Y\right) dt$$

with

- $A_t^Y f(x) := \frac{c^2}{2}\Delta f(x) - c^2\, Y(t)\cdot\nabla f(x)$
- $\sigma_t^Y(x) := \frac{c^2}{2}\|Y(t)\|^2 - \frac{1}{2}\|x\|^2$

Hence

$$E_{\mu_0}\left[1_A\left(X(t)\right)|\mathcal{Y}(t)\right] = \frac{\int 1_A\left(x\right)e^{Y(t)\cdot x}\,\mu_t^Y\left(dx\right)}{\int e^{Y(t)\cdot x}\,\mu_t^Y\left(dx\right)} \quad a.s.$$

where

$$\text{(1)} \qquad \frac{d}{dt}\mu_t^Y = \hat{A}_t^Y\mu_t^Y + \left(\sigma_t^Y - \int_{R^d}\sigma_t^Y\,d\mu_t^Y\right)\mu_t^Y$$

Interested in stability of μ_t^Y w.r.t. μ_0

Stochastic relaxation (e.g. simulated annealing)

Determine global maximum of

$$\sigma : \mathbb{R}^d \quad \longrightarrow \quad]-\infty, K]$$

using the solution of

$$\frac{d}{dt}\mu_t = \frac{c^2}{2}\Delta\mu_t + \beta(t)\left(\sigma - \int_{\mathbb{R}^d} \sigma \, d\mu_t\right) \mu_t \tag{2}$$

where

$$\beta : [0,\infty) \to [0,\infty) \qquad \nearrow \qquad \text{(\textit{inverse temperature})}$$

under suitable conditions on σ and β μ_t concentrates more and more on

$$M = \{x \in \mathbb{R}^d : \sigma(x) = \sup \sigma\}$$

Genetic algorithms (GA)

Example: GA with mutation $\frac{c^2}{2}\Delta$ and fitness function σ

$$\frac{d}{dt}\mu_t = \frac{c^2}{2}\Delta\mu_t + \left(\sigma - \int_{\mathbb{R}^d} \sigma \, d\mu_t\right) \mu_t \tag{3}$$

describes approx. the empirical distribution of types within a population with mutation $\frac{c^2}{2}\Delta$ and selection given by σ

Remark

- Δ can be replaced by any generator of a Markov process
- $\mathbb{R}^d$ can be replaced by any type space

2. Long-time behaviour of (3)

changes to the population come from two opposing sources:

- selection - encourages conformity

  in fact: $c = 0$ implies

$$\int \sigma \, d\mu_t \qquad \nearrow \qquad \textit{(mean fitness)}$$

  since $\mu_t = \frac{e^{t\sigma}\mu_0}{\int e^{t\sigma}\,d\mu_0}$, so that

$$\frac{d}{dt}\int \sigma \, d\mu_t = \frac{\int \sigma^2 e^{t\sigma}\,d\mu_0}{\int e^{t\sigma}\,d\mu_0} - \left(\frac{\int \sigma e^{t\sigma}\,d\mu_0}{\int e^{t\sigma}\,d\mu_0}\right)^2 = \mathrm{Var}_{\mu_t}(\sigma)$$

  hence μ_t concentrates more and more on

$$M(\mu_0) = \{x \in \mathbb{R}^d : \sigma(x) = \mu_0 - ess\, sup\, \sigma\}$$

- mutation - improves genetic diversity

The variational approach

[St, Probab. Theory Relat. Fields 04]

Consider $\mu_0 = h\,dx$ with $h \in L^2(\mathbb{R}^d)$

replace the mean fitness by the energy

$$\lambda : \underbrace{\partial B_1(0)}_{\text{unit sphere in } L^2(\mathbb{R}^d)} \longrightarrow]-\infty+\infty]$$

$$h \mapsto \frac{c^2}{2} \int \|\nabla h\|^2 \, dx - \int \sigma h^2 \, dx$$

Then

- if h_* is a minimum of λ, $|h_*|$ is a minimum too, and

$$\mu_* := \frac{|h_*|\,dx}{\int |h_*|\,dx}$$

  is a stationary point

- existence and regularity of minimizing points with direct methods from variational calculus

- replacing (3) by the L^2-normalization

(4)
$$\begin{aligned} \frac{d}{dt}h_t = \frac{c^2}{2}\Delta h_t + \sigma\, h_t \quad &+ \quad \lambda(h_t)\, h_t \\ = \quad -\frac{1}{2}\lambda'(h_t) \quad &+ \quad \lambda(h_t)\, h_t \end{aligned}$$

  implies

$$\mu_t = \frac{h_t\,dx}{\int h_t\,dx} \qquad \text{and} \qquad \lambda(h_t) \qquad \searrow$$

Theorem 1 ([St, PTRF 04]) Let $h_* \geq 0$ be a minimum of λ. Then

$$\|h_t - h_*\|_{L^2(\mathbb{R}^d)} \leq e^{\left(\frac{1}{\int h_0 h_* \, dx}\right)} e^{-\frac{\kappa_*}{2}t} \|h_0 - h_*\|_{L^2(\mathbb{R}^d)}$$

Here $\kappa_* = \lambda_1 - \lambda_0$, where

$$\lambda_0 = \lambda(h_*) = \inf_{h \in \partial B_1(0)} \lambda(h) \qquad \text{and} \qquad \lambda_1 = \inf_{\substack{h \in \partial B_1(0) \\ \int h h_* \, dx = 0}} \lambda(h)$$

huge literature on h_* and estimates on κ_* (!)

(also for general mutation)

Corollary If $\nu_* = \frac{h_* \, dx}{\int h_* \, dx}$, then

$$\mu_t \longrightarrow \nu_* \qquad \text{vaguely}$$

i.e. $\lim_{t\to\infty} \int f \, d\mu_t = \int f \, d\nu_* \; \forall f \in C_0(\mathbb{R}^d)$

Remark if the mutation operator has a symmetrizing probability measure ν, then

$$\mu_t \longrightarrow \nu_* \quad \text{with exp. rate } \frac{\kappa_*}{2} \text{ w.r.t. } \|\cdot\|_{var}$$

i.e. $\left|\int f \, d\mu_t - \int f \, d\nu_*\right| \leq const \cdot e^{-\frac{\kappa_*}{2} t} \|f\|_\infty$

Remark immediate generalizations to

- **interactive selection**

$$\sigma : \mathbb{R}^d \times \mathcal{M}_1(\mathbb{R}^d) \to \mathbb{R}^d\,, \quad (x,\mu) \mapsto \sigma(x,\mu)$$

of interest in populations genetics

$$\sigma(x,\mu) = \int_{\mathbb{R}^{d(n-1)}} k(z_1,\ldots,z_{n-1},x)\,\mu(dz_1)\ldots\mu(dz_{n-1})$$

for some $k \in \mathcal{B}_{b,sym}(\mathbb{R}^{dn})_+$

(work by Fleming, Donnelly, Kurtz, Dawson, Greven ...)

more general

$$\sigma(x,\mu) = \frac{\partial G_\sigma}{\partial \delta_x}(\mu) := \frac{dG_\sigma}{ds}(\mu + s\delta_x)_{|s=0}$$

for some potential

$$G_\sigma : \mathcal{M}_+(\mathbb{R}^d) \to \mathbb{R}$$

- **time dependent settings**

- **time discrete analoga** replace Δ by the one-step transition kernel P of a Markov chain

$$\mu_n \overset{selection}{\longrightarrow} \frac{e^{\sigma(\cdot,\mu_n)}\mu_n}{\int e^{\sigma(\cdot,\mu_n)}\,d\mu_n} =: \mu'_{n+1} \overset{mutation}{\longrightarrow} \hat{P}\mu'_{n+1} =: \mu_{n+1}$$

Simulation via particle approximation

the solution of (3) admits the representation

$$\mu_t(A) = \frac{E_{\mu_0}\left[1_A(X(t)) \exp\left(\int_0^t \sigma(X(s))\, ds\right)\right]}{E_{\mu_0}\left[\exp\left(\int_0^t \sigma(X(s))\, ds\right)\right]}$$

$$\uparrow$$

$$\text{ren. Feynman} - \text{Kac semigroup}$$

- **classical** Monte-Carlo
- **adaptive methods** (del Moral, Miclo 00) approximation via *n-particle Moran processes with mutation* $\frac{c^2}{2}\Delta$ *and selection* σ

$$Y_1^{(n)}(t), \ldots, Y_n^{(n)}(t) \;, t \geq 0 \,,$$

Generator

$$L^{(n)}f(x) \quad = \underbrace{\tfrac{c^2}{2}\Delta^{(n)}f(x)}_{\text{ind. mutation}} \quad + \quad \underbrace{\tfrac{1}{n}\textstyle\sum_{i\neq j}\sigma(x_i)\left(\Phi_{ij}^{(n)}f(x)-f(x)\right)}_{\text{selection}}$$

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

approximation justified by

LLN (del Moral/Miclo 00): σ bounded, $(Y_i^{(n)}(0))$ iid (μ_0), implies

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{Y_i^{(n)}(t)} \overset{w}{\to} \mu_t\,, t \geq 0$$

3. The variational approach to (1)

Assume that

A.1 $B(x) = \frac{\nabla\varphi}{\varphi}(x)$ for some bounded log-concave function φ

$\Longrightarrow \nu(dx) := \varphi^2\, dx$ is symm. measure for A

A.2 $W(x) := \|Gx\|^2 + \frac{\Delta\varphi}{\varphi}(x)$ uniformly strictly convex

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

Remark Let

$$Af(x) := \frac{1}{2}\Delta f(x) + B(x)\cdot\nabla f(x) \quad \text{and} \quad \sigma(x) := -\frac{1}{2}\|Gx\|^2$$

$\Longrightarrow A + \sigma$ has a unique ground state h_*

and spectral gap $\geq \kappa_*$ below λ_0 (Brascamp, Lieb 76)

Long-time behaviour of (1)

$$\frac{d}{dt}\mu_t^Y = \hat{A}_t^Y \mu_t^Y + \quad \sigma_t^Y - \quad {}_{R^d} \sigma_t^Y \, d\mu_t^Y \quad \mu_t^Y$$

- $A_t^Y f(x) := Af(x) - G^T Y(t) \cdot \nabla f(x)$
- $\sigma_t^Y(x) := \sigma(x) - G^T Y(t) \cdot B(x) + \frac{1}{2}\|G^T Y(t)\|^2$

Heuristic $Y \equiv 0$: then (1) reduces to

$$\frac{d}{dt}\mu_t = \hat{A}\mu_t + \left(\sigma - \quad \sigma \, d\mu_t\right) \mu_t$$

Theorem 1 implies

$$\lim_{t\to\infty} \mu_t(dx) = \nu_*(dx) := \frac{h_*\varphi^2 \, dx}{\int h_*\varphi^2 \, dx}$$

with exponential rate $\frac{\kappa_*}{2}$ (w.r.t. $\|\cdot\|_{L^2}$ and $\|\cdot\|_{var}$)

Theorem 2 ([St 04]) Let

- $Y \in C([0,\infty[;\mathbb{R}^p)$, $Y(0) = 0$
- $\mu_i \ll \nu_* = \frac{h_* \, \nu}{\int h_* \, d\nu}$ with densities bounded from below and from above

Then

$$\limsup_{t\to\infty} e^{\frac{\kappa_*}{2}t} \|\mu^Y_{1,t} - \mu^Y_{2,t}\|_{var} < \infty$$

Remark

- rate *independent* of Y
- precise quantitative bounds
- (1) can be stable even if (S) is non-ergodic - first proof in the nonlinear time-continuous case

Corollary Assume that (1) is unique for all $Y \in C([0,\infty[;\mathbb{R}^p)$, $Y(0) = 0$, then

$$\limsup_{t\to\infty} e^{\frac{\kappa_*}{2}t} \left|E_{\mu_1}\left[1_A\left(X(t)\right)|\mathcal{Y}(t)\right] - E_{\mu_2}\left[1_A\left(X(t)\right)|\mathcal{Y}(t)\right]\right| < \infty \quad a.s.$$

Remark

- simple sufficient criterion for stability

$$W(x) = \|Gx\|^2 + \frac{\Delta\varphi}{\varphi}(x)\,, \qquad B = \frac{\nabla\varphi}{\varphi}$$

$W'' \geq \kappa_*^2 \cdot I$ implies exp. stability with rate $\dfrac{\kappa_*}{2}$

- more precise and explicit results in the linear case (Kalman-Bucy filter)
  - also for non time-reversible signals
- simlar results for time-dependent signals and for the discrete time case

Stochastic partial differential equations (SPDE)

The Zakai-equation is a measure-valued SPDE of the type

$$(5) \qquad d\eta_t = \quad \frac{c^2}{2}\Delta\eta_t + F(\eta_t) \quad dt + C(\eta_t)\, dY(t)$$

Applications

- modelling of the evolution of spatially distributed populations (epidemiology, population genetics, ...)

  η_t = (density of) distr. of ind. within a given population

  Δ = migration/mutation

  C = branching/reproduction

  $Y(\cdot)$ = cylindrical Brownian motion on $L^2(\mathbb{R}^d)$

Example $C(\eta) = \sqrt{2\,\gamma\,\eta}$

$$\Longrightarrow \qquad u(t,x) := -\log E_{\delta_x}\left[e^{-\int f\, d\eta_t}\right]$$

solves

$$\partial_t u = \frac{c^2}{2}\Delta u - \gamma\, u^2 \quad , \quad u(0,\cdot) = f$$

Contributions

- long-time behaviour (existence & characterization of invariant measures, ergodicity, rates of convergence)
- regularity

Further applications

- Reaction-diffusion equations $F(\eta)(x) = f(\eta(x))$
- Hydrodynamics
  - Burgers equation (d=1) $F(\eta) = -\partial_x \left(\frac{\eta^2}{2}\right)$
  - Navier-Stokes equation $F(\eta) = -(\eta \cdot \nabla)\, \eta - \nabla p$, div $\eta = 0$

Contributions ($d = 2$, periodic bdy. conditions)

- Uniqueness of *statistical* solutions of a regularized Euler equation
- Uniqueness of a stochastic Navier-Stokes equation

References

- [St1] On the convergence of genetic algorithms - a variational approach, Probab. Theory Relat. Fields 129, 113-132 (2004).
- [St2] Stability of the pathwise filter equation on $\mathbb{R}^d$, Bielefeld (2004), submitted.
- [St3] Stability of the pathwise filter equation for a time-dependent signal on $\mathbb{R}^d$, Bielefeld 2004, to appear in Appl. Math. Optim.
- [St4] Stability of the optimal filter via pointwise gradient estimates, Bielefeld 2004, submitted.