Mean-field approach to Bayesian estimation of Markovian signals

Wilhelm Stannat

Institut für Mathematik, TU Berlin

Annecy, June 27, 2022

Part I:

Mean-field type approximation of the posterior

The Bayesian approach - general framework

estimating a signal $X \in \mathbb{R}^d$ via noisy data $Y = G(X, \zeta)$

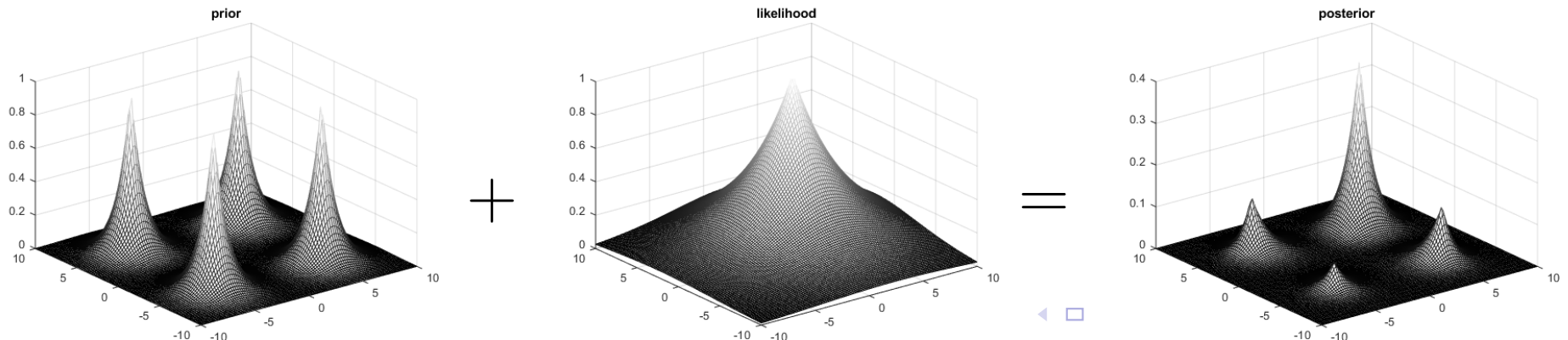
- ζ - measurement noise
- Y - observation, e.g. $\in \mathbb{R}^p$, $L^2(D)$, ...

given/known in applications

- model for X, e.g. in terms of (S)ODE, (S)PDE, ... yields **prior** P_X
- distribution of meas. noise $P(Y \in dy \mid X = x) = e^{-\ell(y,x)} dy$

yields **posterior** distribution (Bayes theorem)

$$P_{X|Y=y}(dx) = \frac{e^{-\ell(y,x)} P_X(dx)}{\int e^{-\ell(y,\bar{x})} P_X(d\bar{x})} \propto e^{-\ell(y,x)} P_X(dx)$$



Monte-Carlo approximation of $P_{X|Y=y}$

given $X_1, \ldots, X_M \in \mathbb{R}^d$ (*particles*), e.g. indep. with common distr. P_X, then

$$\pi^M := \frac{1}{M} \sum_{i=1}^{M} \delta_{X_i} \approx P_X$$

since the SLLN implies

$$\int f \, d\pi^M = \frac{1}{M} \sum_{i=1}^{M} f(X_i) \to E\left(f(X)\right) = \int f \, dP_X$$

introducing weights $w_i = \exp(-l(y, X_i))$ yields

$$\tilde{\pi}^M := \sum_{i=1}^{M} \frac{w_i}{\sum_{j=1}^{M} w_j} \delta_{X_i} \approx P_{X|Y=y}$$

since

$$\int f \, d\tilde{\pi}^M = \frac{\sum_{i=1}^{M} w_i f(X_i)}{\sum_{i=1}^{M} w_i} \to \frac{\int f \, e^{-\ell(y,\cdot)} \, dP_X}{\int e^{-\ell(y,\cdot)} \, dP_X}$$

drawback w_i might degenerate for most particles, requires resampling in iterations (causing additional approximation errors)

The posterior as a push-forward measure

Ansatz realize $P_{X|Y=y}$ as image measure $P_{X|Y=y} = \Phi_{\#} P_X$

$$\int f(x)\, P_{X|Y=y}(dx) = \int f(\Phi(x))\; P_X(dx)$$

if $P_X(dx) = \pi(x)\, dx$, Φ may be obtained as solutions of the Monge-Ampère equation

$$\det\,(D\Phi(x)) = Z \exp\left(\ell(y, \Phi(x))\right) \frac{\pi(x)}{\pi(\Phi(x))},$$

with normalizing constant

$$Z = \int \exp\left(-\ell(y, x)\right) \pi(x)\, dx$$

- fully nonlinear
- non-unique: $\Phi \circ \Psi$ (resp. $\Psi \circ \Phi$) again solution for any Ψ differentiable and measure-preserving P_X (resp. $P_{X|Y=y}$)
- requires additional constraints for uniqueness

Linearized Monge-Ampère equation

$\ell \ll 1$ implies $\Phi(x) \approx x + K(x)$ with $K \ll 1$

neglecting higher order terms in ℓ, K and Jacobian DK gives

$$\begin{aligned} \det\,(D\Phi(x)) &\approx 1 + \text{trace}\ (DK(x)) = 1 + \nabla \cdot K(x) \\ \exp\left(-\ell(y, \Phi(x))\right) &\approx (1 - \ell(y, \Phi(x))) \approx (1 - \ell(y, x)) \\ \eta(\Phi(x)) &\approx \eta(x) + \nabla\eta(x) \cdot K(x) \end{aligned}$$

inserting approximations gives

$$\nabla \cdot (\pi K) = \pi \nabla \cdot K + \nabla \pi \cdot K = -\pi\big(\ell(y, \cdot) - \int \ell(y, \cdot)\pi\, dx\big) \tag{1}$$

MC-approximation: $X_1, \ldots, X_M$ i.i.d. with distr. P_X

$$\frac{1}{M}\sum_{m=1}^{M} \delta_{X_m} \approx P_X \quad \Rightarrow \quad \frac{1}{M}\sum_{m=1}^{M} \delta_{X_m + K(X_m)} \approx P_{X|Y=y}$$

and $K(x) = K(x, \pi)$ is called **Kalman gain**

Gradient type solutions - Poisson equation

$$\nabla \cdot (\pi K(\cdot, \pi)) = -\pi(\ell - \pi[\ell]), \quad \pi[\ell] = \int \ell(y, x)\, \pi(x)\, dx$$

unique up to π-measure preserving transform. $\mathcal{J}(x, \pi)$, i.e. $\nabla \cdot (\pi\, \mathcal{J}) = 0$

Ansatz $K(\cdot, \pi) = \nabla \phi$ leads to weighted Poisson eq

$$\nabla \cdot (\pi \nabla \phi) = -\pi(\ell - \pi[\ell]) \tag{2}$$

(`Young, Mehta, Meyn, et al (2013)`), with formal solution

$$\phi = \int_0^\infty e^{u\left(\Delta + \frac{\nabla \pi}{\pi} \cdot \nabla\right)} (\ell - \pi[\ell])\, du$$

solving (2) diffusion map approximation

$$\begin{aligned} \phi &= e^{\varepsilon\left(\Delta + \frac{\nabla \pi}{\pi} \cdot \nabla\right)} \phi + \int_0^\varepsilon e^{u\left(\Delta + \frac{\nabla \pi}{\pi} \cdot \nabla\right)} (\ell - \pi[\ell])\, du \\ &\approx e^{\varepsilon\left(\Delta + \frac{\nabla \pi}{\pi} \cdot \nabla\right)} \phi + \varepsilon(\ell - \pi[\ell]) \approx T_\varepsilon \phi_\varepsilon + \varepsilon(\ell - \pi[\ell]) \end{aligned}$$

yields MCMC approximation (`[Pathiraja, S., FoDS '21]`)

$$\phi_\varepsilon = \varepsilon \sum_{k=0}^{\infty} T_\varepsilon^k \left(\ell - \pi_\varepsilon[\ell]\right), \quad T_\varepsilon(x, A) = \frac{P_{2\varepsilon}^{BM}\left(1_A \tilde{\pi}\right)}{P_{2\varepsilon}^{BM}\left(\tilde{\pi}\right)}, \quad \tilde{\pi} = \frac{\pi}{\sqrt{P_{2\varepsilon}^{BM} \pi}}$$

Part II:

Bayesian inference of time-structured data

State estimation of sde's

$$(S) \qquad \mathrm{d}X_t = f(X_t)\mathrm{d}t + Q(X_t)^{\frac{1}{2}}\mathrm{d}W_t$$

$$(O) \qquad \mathrm{d}Y_t = g(X_t)\mathrm{d}t + R^{\frac{1}{2}}\mathrm{d}V_t$$

Bayesian Ansatz in principle known:

- law of $X_{0:t}$
- conditional law of $Y \mid X$ - Brownian motion with drift, i.e.

$$Y_{0:t} \mid X_{0:t} \sim \exp\left(\int_0^t g(X_s)R^{-1}dY_s - \frac{1}{2}\int_0^t \|R^{-\frac{1}{2}}g(X_s)\|^2\, ds\right) P_0(dY_{0:t})$$

  where $P_0 =$ Wiener meas. (Girsanov-Theorem)

yields - according to Bayes rule -

$$X_{0:t} \mid Y_{0:t} \sim \frac{\exp\left(\int_0^t g(X_s)R^{-1}dY_s - \frac{1}{2}\int_0^t \|R^{-\frac{1}{2}}g(X_s)\|^2\, ds\right) P(dX_{0:t})}{\int \exp\left(\int_0^t g(X_s)R^{-1}dY_s - \frac{1}{2}\int_0^t \|R^{-\frac{1}{2}}g(X_s)\|^2\, ds\right) P(dX_{0:t})}$$

Kushner-Stratonovich equation (SPDE)

$$\begin{aligned}\pi_t(dx) := \pi^{Y_{0:t}}(dx) &= P\left(X_t \in dx \mid Y_{0:t}\right) \\ &= \text{ cond. distr. of } X_t \text{ given } Y_s, s \in [0,t]\end{aligned}$$

solves the measure-valued spde

$$d\pi_t(\varphi) = \underbrace{\pi_t[L\varphi]}_{\text{forecast}} + \underbrace{\mathrm{cov}_{\pi_t}(g,\varphi)R^{-1}dI_t}_{\text{innovation}}$$

where

- $Lu(x) = \frac{1}{2}\mathrm{tr}\left(Q(x)u''(x)\right) + f(x)\nabla u(x)$ gen. of (S) hence $\pi_t[L\varphi]\,dt$ is of Fokker-Planck type
- $dI_t = dY_t - \pi_t[g]\,dt$ measures deviation of data from forecast

if $\pi_t \ll dx$, its density $\pi(t,x)$ is a solution of the spde

$$\begin{aligned}\partial_t \pi(t,x) = \frac{1}{2}\sum_{i,j} \partial^2_{x_i x_j}\left(Q_{ij}(x)\pi(t,x)\right) - \sum_i \partial_{x_i}\left(f_i(x)\pi(t,x)\right) \\ + \left(g^T(x) - \int g^T(y)\pi(t,y)\,dy\right)\pi(t,x)R^{-1}\frac{dI_t}{dt}\end{aligned}$$

Part III:

The feedback particle filter

Feedback Particle Filter

approximate π_t in terms of $\pi_t^M = \frac{1}{M}\sum_{i=1}^M \delta_{X_t^i}$ with

$$dX_t^i = f(X_t^i)dt + Q^{\frac{1}{2}}(X_t^i)dW_t^i \quad + \underbrace{K\left(X_t^i, \pi_t^M\right) \circ \left(dY_t - \pi_t^M[g]\,dt\right)}_{\text{interaction term}}$$

and $K(\cdot, \pi)$ solves

$$\nabla \cdot (\pi\, K(\cdot, \pi)) = -\pi\,(g - \pi[g])\,R^{-1}, \quad \pi[g] = \int g(x)\,\pi(x)\,dx$$

can expect for large M convergence of X_t^i to ind. solutions of

$$d\bar{X}_t = f(\bar{X}_t)\,dt + Q^{\frac{1}{2}}(\bar{X}_t)\,dW_t + K(\bar{X}_t, \pi_t) \circ (dY_t - \pi_t[g]\,dt)$$

Math. problems

(1) Kalman gain K well-defined and sufficiently regular

(2) well-posedness of the mf-sde

(3) well-posedness and convergence of X_t^i to (ind. copies of) $\bar{X}_t$ (*mf-limit*)

Rem fair to say: (2) & (3) completely open

Well-posedness of the FPF - first steps

Ansatz for (1)

solution to the weighted Poisson equation

$$\nabla \cdot (\pi_t \nabla \phi_t) = (g - \pi_t[g])\pi_t$$

can formally be represented as

$$\phi_t = \int_0^\infty e^{u\left(\Delta + \frac{\nabla \pi_t}{\pi_t} \cdot \nabla\right)} (g - \pi_t[g])\, du$$

exp decay of $e^{u\left(\Delta + \frac{\nabla \pi_t}{\pi_t} \cdot \nabla\right)} (g - \pi[g])$ (in u) implied by Poincaré inequality

$$\int (\varphi - \pi_t[\varphi])^2 \, d\pi_t \leq \kappa_t \int |\nabla \varphi|^2 \, d\pi_t$$

our Ansatz 'Poincaré via log-concavity' $\pi_t \propto e^{-\mathcal{G}_t}$, $\mathcal{G}_t$ convex, e.g. $N(m, Q)$

Well-posedness of the FPF - first steps

Thm [Pathiraja, Reich, S., SICON '21] Let $\pi_t(x)$ be the posterior density. Suppose that

- $f = -\nabla F$ gradient type, $F'' \geq c_F \mathrm{Id}$, $c_F > 0$, $g(x) = Gx$ linear
- $V = -\Delta F + |\nabla F|^2 + |Gx|^2$ uniformly strictly convex, $V'' \geq c_V \mathrm{Id}$
- initial density $\pi_0 \propto \exp(-\mathcal{G}_0)$, $\mathcal{G}_0'' \geq c_G \mathrm{Id}$ with $c_G > c_F$

then

$$\int (\varphi - \pi_t[\varphi])^2 \, d\pi_t \leq \kappa_t \int |\nabla \varphi|^2 \, d\pi_t$$

with constant less than

$$\kappa_t = \left(c_F + \min\left(c_G - c_F, \sqrt{\frac{c_V}{2}} \right) \right)^{-1}.$$

main feature π_t log-concave for all t (very restrictive (!), e.g. excludes multi target tracking)

Part IV:

The Ensemble Kalman-Bucy filter:

the popular relative of the FPF

Linear signals: Kalman-Bucy filter (KBF)

$$(S) \qquad \mathrm{d}X_t = (AX_t + b)\,\mathrm{d}t + Q^{\frac{1}{2}}\mathrm{d}W_t$$

$$(O) \qquad \mathrm{d}Y_t = GX_t\mathrm{d}t + R^{\frac{1}{2}}\mathrm{d}V_t$$

main feature

$$\pi_0 = N(m_0, P_0) \text{ implies } \pi_t = N(m_t, P_t)$$

with

$$\mathrm{d}m_t = (Am_t + b)\,\mathrm{d}t + P_t G^{\mathrm{T}} R^{-1}(\mathrm{d}Y_t - Gm_t\mathrm{d}t)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t = AP_t + P_t A^{\mathrm{T}} + Q - P_t G^{\mathrm{T}} R^{-1} G P_t$$

Rem dim of (S) might be large, hence computation of P_t costly, requires low-dimensional approx.

Mean-field representation of the KBF

$$(S) \qquad \mathrm{d}X_t = (AX_t + b)\,\mathrm{d}t + Q^{\frac{1}{2}}\mathrm{d}W_t$$

$$(O) \qquad \mathrm{d}Y_t = GX_t\mathrm{d}t + R^{\frac{1}{2}}\mathrm{d}V_t$$

lin. Monge-Ampère eq reduces for $\pi = N(m, P)$ to

$$\langle P^{-1}(x-m), K\rangle - \nabla \cdot K = \langle R^{-1}G(x-m), dY_t\rangle - \frac{1}{2}\langle R^{-1}Gx, Gx\rangle + \text{const.}$$

with explicit solution

$$K(x,\pi) = PG^T R^{-1}\left(dY_t - \frac{1}{2}G(x+m)\,dt\right)$$

and McKean-Vlasov sde

$$d\bar{X}_t = \left(A\bar{X}_t + b\right)\,dt + Q^{\frac{1}{2}}dW_t + P_t G^T R^{-1}\left(dY_t - \frac{1}{2}G(\bar{X}_t + m_t)\,dt\right)$$

Ensemble Kalman-Bucy filter (EnKBF)

Main idea replace

$$m_t \rightsquigarrow \bar{x}_t^M := \frac{1}{M}\sum_{i=1}^{M} X_t^i$$

$$P_t \rightsquigarrow P_t^M := \frac{1}{M-1}\sum_{i=1}^{M}(X_t^i - \bar{x}_t^M)(X_t^i - \bar{x}_t^M)^{\mathrm{T}}$$

where

$$\mathrm{d}X_t^i = f(X_t^i)\mathrm{d}t + Q(X_t^i)^{\frac{1}{2}}\mathrm{d}W_t^i + P_t^M G^T R^{-1}\left(\mathrm{d}Y_t - \frac{1}{2}G(X_t^i + \bar{x}_t^M)\mathrm{d}t\right)$$

for M ind. Brownian motions W^i, or its deterministic analogon

$$\mathrm{d}X_t^i = f(X_t^i)\mathrm{d}t + \frac{1}{2}Q(X_t^i)\left(P_t^M\right)^{-1}(X_t^i - \bar{x}_t^M)\mathrm{d}t + P_t^M G^T R^{-1}\left(\mathrm{d}Y_t - \frac{1}{2}G(X_t^i + \bar{x}_t^M)\mathrm{d}t\right)$$

avoids degeneracy problem of classical weighted particle filters, in part. avoids resampling

Well-posedness of strong solutions

coefficients of

$$\mathrm{d}X_t^i = f(X_t^i)\mathrm{d}t + Q(X_t^i)^{\frac{1}{2}}\mathrm{d}W_t^i + P_t^M G^T R^{-1}\left(\mathrm{d}Y_t - \frac{1}{2}G(X_t^i + \bar{x}_t^M)\mathrm{d}t\right)$$

have cubic growth, but $\mathrm{tr}(P_t^M)$ has semi-martingale decomposition

$$\mathrm{d}\,\mathrm{tr}(P_t^M) = \frac{2}{M-1}\sum_{i=1}^{M}\langle f(X_t^i) - f(\bar{x}_t), X_t^i - \bar{x}_t\rangle\,\mathrm{d}t$$
$$+\,\mathrm{tr}\left(\frac{1}{M}\sum_{i=1}^{M} Q(X_t^i) - P_t^M G^T R^{-1} G P_t^M\right)\mathrm{d}t + \mathrm{d}\,\mathrm{tr}(N_t)$$

with local martingale

$$\mathrm{d}N_t := \frac{1}{M-1}\sum_{i=1}^{M}\left(Q(X_t^i)^{\frac{1}{2}}\mathrm{d}\left(W_t^i - \bar{w}_t\right)\right)\left(X_t^i - \bar{x}_t\right)^T$$
$$+\left(X_t^i - \bar{x}_t\right)\left(Q(X_t^i)^{\frac{1}{2}}\mathrm{d}\left(W_t^i - \bar{w}_t\right)\right)^T dt$$

stochastic Gronwall lemma now implies for all $p < 1$

$$\mathbb{E}\left[\sup_{0\le t\le T\wedge\xi}\mathrm{tr}(P_t^M)^p\right] \le \mathcal{C}_p e^{2p\|f\|_{\mathrm{Lip}}T}\mathbb{E}\left[\left(\int_0^{T\wedge\xi}\frac{1}{M}\sum_{i=1}^{M}\mathrm{tr}(Q(X_t^i))\,dt\right)^p\right]$$

up to some life-time ξ

Well-posedness of strong solutions, ctd.

$$\mathrm{d}X_t^i = f(X_t^i)\mathrm{d}t + Q(X_t^i)^{\frac{1}{2}}\mathrm{d}W_t^i + P_t^M G^T R^{-1}\left(\mathrm{d}Y_t - \frac{1}{2}G(X_t^i + \bar{x}_t^M)\mathrm{d}t\right)$$

similarly,

$$\mathrm{d}\,\bar{x}_t^M = (\frac{1}{M}\sum_{i=1}^{M} f(X_t^i) - P_t^M G^T R^{-1} G\bar{x}_t^M)\mathrm{d}t$$
$$+ P_t^M G^T R^{-1}\mathrm{d}Y_t + \mathrm{d}\bar{N}_t$$

with local martingale

$$\mathrm{d}\bar{N}_t := \frac{1}{M}\sum_{i=1}^{M} Q(X_t^i)^{\frac{1}{2}}\mathrm{d}W_t^i$$

another application of the stochastic Gronwall lemma implies

$$\mathbb{E}\left[\sup_{0\le t\le T\wedge\xi} \|\bar{x}_t^M\|^p\right] < \infty$$

so that $P[\xi \le T] = 0$ for all T

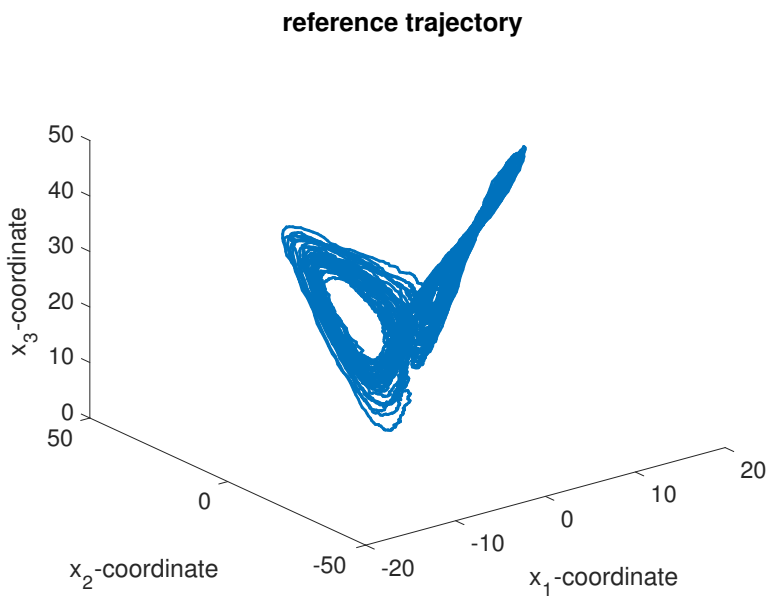
Numerical Illustration

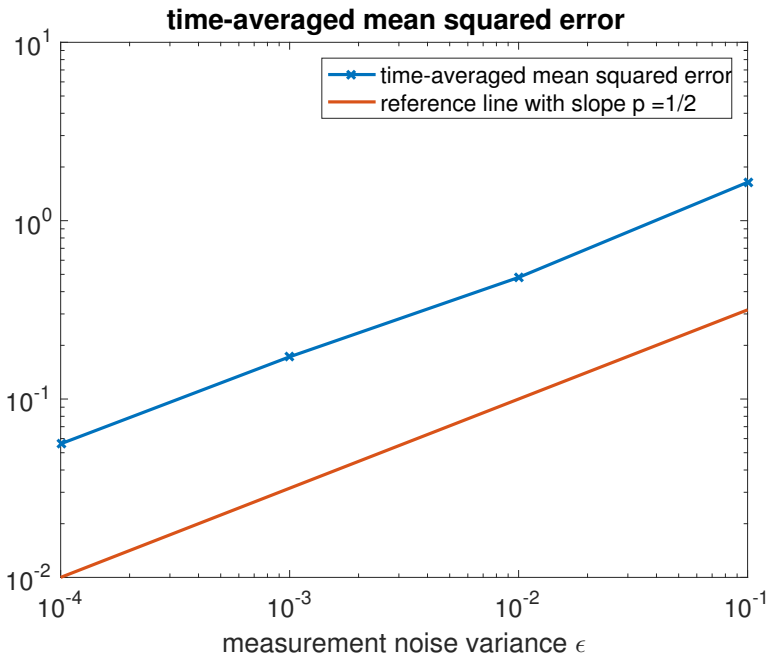
Lorenz-63 system with additive noise

$$(S) \qquad \mathrm{d}X_t = f(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

$$(O) \qquad \mathrm{d}Y_t = X_t\mathrm{d}t + \sqrt{\varepsilon}\mathrm{d}V_t$$

$$f(x) = \begin{pmatrix} 10(x_2 - x_1) \\ (28 - x_3)x_1 - x_2 \\ x_1 x_2 - \frac{8}{3}x_3 \end{pmatrix}$$





EnKBF with ensemble size $M = 4$

accuracy: [de Wiljes, Reich, S., SIADS '18]

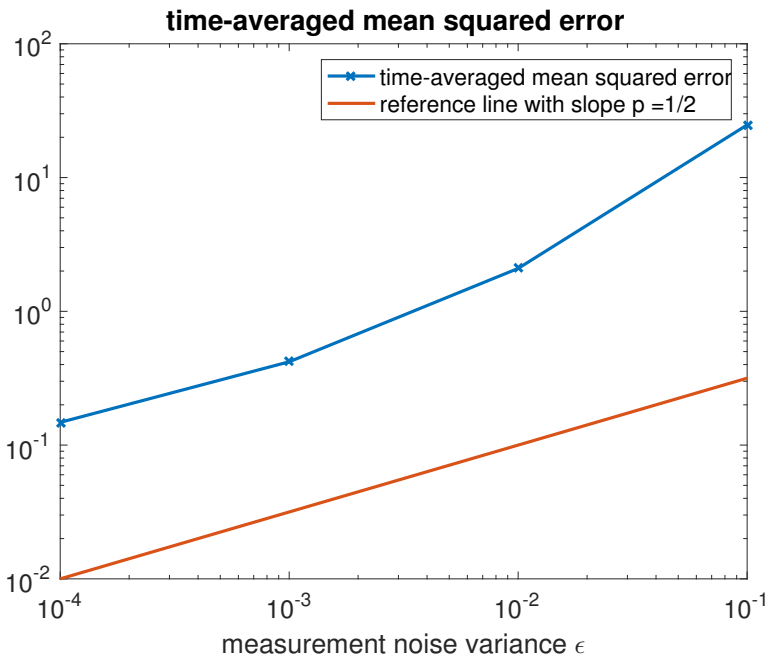
Numerical Illustration, $M = 2, 3$, P_t^M sing.

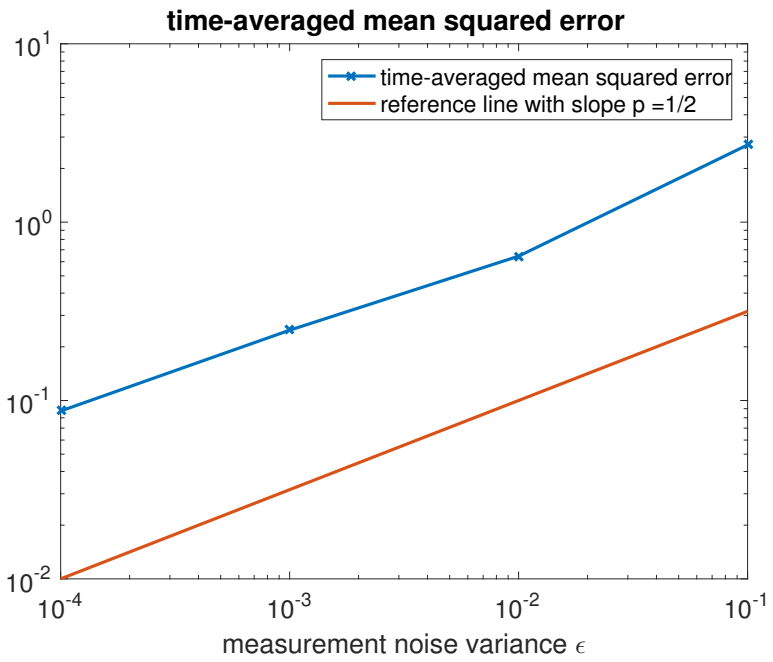
Lorenz-63 system with additive noise

$$(S) \qquad \mathrm{d}X_t = f(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

$$(O) \qquad \mathrm{d}Y_t = X_t\mathrm{d}t + \sqrt{\varepsilon}\mathrm{d}V_t$$

$$f(x) = \begin{pmatrix} 10(x_2 - x_1) \\ (28 - x_3)x_1 - x_2 \\ x_1 x_2 - \frac{8}{3}x_3 \end{pmatrix}$$





Part V:

The MF-limit of the EnKBF

Mean field limit for $M \to \infty$

Thm [Lange, S., FoDS '21]

- $Q(x) \equiv Q$ (additive signal noise)
- X_0^i, $i = 1, \ldots, M$, i.i.d. with distribution $\bar{\pi}_0$ having finite second moments
- W_t^i ind. Brownian motions

Then

$$X_t^i = \bar{X}_t^i + r_t^i \tag{3}$$

- $\bar{X}_t^i$ are sol. to

$$\mathrm{d}\bar{X}_t^i = f(\bar{X}_t^i)\mathrm{d}t + Q^{\frac{1}{2}}\mathrm{d}W_t^i + \mathrm{Cov}(\bar{X}_t^i)R^{-1}\left(\mathrm{d}Y_t - \frac{1}{2}G(\bar{X}_t^i + E(\bar{X}_t^i))\mathrm{d}t\right)$$

- r_t^i satisfies a.s. w.r.t. to the distr. of Y

$$\sup_{t\in[0,T]} \frac{1}{M}\sum_{i=1}^{M}\left\|r_t^i\right\|^2 \to 0, \quad M \to \infty, \tag{4}$$

in probability for all $T \geq 0$

Rem extension to correlated meas. noise in `[Ertel, S., arXiv2205.14253]`

The spde driving $P_{\bar{X}_t}$

with the notation $\bar{m}_t = E\left(\bar{X}_t\right)$, $\bar{P}_t = \mathrm{Cov}_P(\bar{X}_t)$

Itô's formula implies that for

$$\mathrm{d}\bar{X}_t = f(\bar{X}_t)\mathrm{d}t + Q(\bar{X}_t)^{\frac{1}{2}}\mathrm{d}W_t + \bar{P}_t G^T R^{-1}\left(\mathrm{d}Y_t - \frac{1}{2}G(\bar{X}_t + \bar{m}_t)\mathrm{d}t\right)$$

the distr. $\bar{\pi}_t = \mathrm{Law}(\bar{X}_t)$ satisfies the nonl. Fokker-Planck spde

$$\begin{aligned}
\mathrm{d}\int \varphi \mathrm{d}\bar{\pi}_t &= \int L\varphi \mathrm{d}\bar{\pi}_t \mathrm{d}t + \frac{1}{2}\int \mathrm{tr}\left(\bar{P}_t G^T R^{-1} G \bar{P}_t \varphi''\right)\mathrm{d}\bar{\pi}_t \mathrm{d}t \\
&\quad + \int \nabla\varphi \mathrm{d}\bar{\pi}_t\, \bar{P}_t G^T R^{-1}\left(\mathrm{d}Y_t - G\bar{m}_t \mathrm{d}t\right) \\
&\quad - \frac{1}{2}\int \nabla\varphi\, \bar{P}_t G^T R^{-1} G\left(x - \bar{m}_t\right)\mathrm{d}\bar{\pi}_t \mathrm{d}t \\
&=: (I) + (II) + (III) + (IV)
\end{aligned}$$

where $L\varphi(x) = \frac{1}{2}\mathrm{tr}\left(Q(x)\varphi''\right)(x) + f(x)\cdot\nabla\varphi(x)$

Rem does not coincide with the KS-eq, unless $\bar{\pi}_t = \mathcal{N}\left(\bar{m}_t, \bar{P}_t\right)$

in this case partial integration in (II) yields $(II) = -(IV)$, and (III) reduces to

$$\int \varphi G(x - \bar{m}_t)^T \mathrm{d}\bar{\pi}_t R^{-1}\left(\mathrm{d}Y_t - G\bar{m}_t \mathrm{d}t\right)$$