A variational approach to path estimation and parameter inference of hidden diffusion processes

arXiv:1508.00506

Tobias Sutter[§], Arnab Ganguly[⋆] and Heinz Koeppl[‡]

[§]Automatic Control Laboratory, ETH Zurich, Switzerland
[⋆]Department of Mathematics, Louisiana State University, USA
[‡]Department of Electrical Engineering, TU Darmstadt, Germany

January 11, 2016

ETH zürich

LSU
LOUISIANA STATE UNIVERSITY

TECHNISCHE UNIVERSITÄT DARMSTADT

Model setup

- signal process

$$dX_t = f(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0, \quad 0 \le t \le T$$

- observation process

$$Y_t = \int_0^t h(X_s)ds + B_t$$

- assumptions
  - f and σ are globally Lipschitz
  - h is twice continuously differentiable
- smoothing density

$$\mathbb{E}\big[\phi(X_t)|\mathbb{F}_T^Y\big] = \int \phi(x)\mathbb{P}_S(x,t)\ dx \quad \text{a.s.}$$

  - $\mathbb{F}_T^Y := \sigma(Y_t \ : \ t \le T)$
  - ϕ bounded measurable function

Smoothing density

- Zakai equation

$$\begin{cases} \mathrm{d}p(x,t) = \mathscr{A}^* p(x,t)\mathrm{d}t + p(x,t)h(x)^\top \mathrm{d}Y_t \\ p(x,0) = p_0(x). \end{cases}$$

- Pardoux equation

$$\begin{cases} \mathrm{d}v(x,t) = -\mathscr{A} v(x,t)\mathrm{d}t - v(x,t)h(x)^\top \mathrm{d}Y_t \\ v(x,T) = 1. \end{cases}$$

$\Rightarrow$ smoothing density $\quad \mathbb{P}_S(x,t) = \frac{p(x,t)v(x,t)}{\int p(x,t)v(x,t)\mathrm{d}x}$

- posterior process

$$d\bar{X}_t^T = g(\bar{X}_t^T, t)dt + \sigma(\bar{X}_t^T)d\bar{W}_t, \quad \bar{X}_0^T = x_0$$

$$g(x,t) := f(x) + \sigma(x)\sigma(x)^\top \nabla \log v(x,t)$$

- $\mathbb{P}\left[X \in A | \mathbb{F}_T^Y\right] = \mathbb{P}\left[\bar{X}^T \in A\right] =: \Pi_{\text{post}}(A, Y_{[0,T]})$

A variational approach to path estimation

approximation $\boxed{\min_{Q \in \mathbb{Q}} \mathrm{D}\big(Q \| \Pi_{\text{post}}(\cdot, Y_{[0,T]})\big)}$

1. how to choose $\mathbb{Q} \subset$ probability measures on $C([0,T])$?
   - too large $\rightarrow$ computationally demanding optimization problem
   - too small $\rightarrow$ bad approximation quality
2. how to evaluate $\mathrm{D}\big(Q \| \Pi_{\text{post}}(\cdot, Y_{[0,T]})\big)$?
   - $\Pi_{\text{post}}(\cdot, Y_{[0,T]})$ is unknown

$\Rightarrow$ reformulation as an optimal control problem

A variational approach to path estimation (cont'd)

$$\boxed{\min_{Q\in\mathbb{Q}} \mathrm{D}\big(Q \| \Pi_{\text{post}}(\cdot, Y_{[0,T]})\big)} \qquad (\bigstar)$$

- prior law $\quad \Pi_{\text{prior}}(A) := \mathbb{P}\left[X \in A\right]$
- $H_T(X,y) := -h(X_T)y_T + \int_0^T y_s \mathrm{d}h(X_s) + \frac{1}{2}\int_0^T \|h(X_s)\|^2 \mathrm{d}s$
- negative log-likelihood
  $I(H_T(\cdot,y)) := -\log\left(\int \exp\left(-H_T(\cdot,y)\right)\mathrm{d}\Pi_{\text{prior}}\right)$

Lemma ([Mitter & Newton'03])

$$\mathrm{D}\big(Q\|\Pi_{\text{post}}(\cdot,y)\big) = -I(H_T(\cdot,y)) + \mathrm{D}\big(Q\|\Pi_{\text{prior}}\big) + \mathbb{E}_Q\big[H_T(\cdot,y)\big]$$

- $(\bigstar)$ is equivalent to

$$\boxed{\min_{Q\in\mathbb{Q}} \mathrm{D}\big(Q\|\Pi_{\text{prior}}\big) + \mathbb{E}_Q\big[H_T(\cdot,y)\big]}$$

A variational approach to path estimation (cont'd)

$$\boxed{\min_{Q \in \mathbb{Q}} \ \mathrm{D}\big(Q \| \Pi_{\text{prior}}\big) + \mathbb{E}_Q\big[H_T(\cdot, y)\big]}$$

Problem ($\square$)

Minimize $\mathrm{D}\big(Q \| \Pi_{\text{prior}}\big) + \mathbb{E}_Q\big[H_T(\cdot, y)\big]$ subject to

(i) Q is a probability distribution induced by an SDE of the form

$$\mathrm{d}Z_t = u(Z_t, t)\mathrm{d}t + \sigma(Z_t)\mathrm{d}W_t, \quad Z_0 = x_0, \quad 0 \le t \le T;$$

(ii) The marginals of Q at time t, i.e., the distribution of Z_t, belong to a chosen family of distributions.

$\Rightarrow$ Problem ($\square$) can be recast as an optimal control problem

- constant diffusion term & Gaussian distribution [Archambeau & Opper'11]

SDE with prescribed marginal law

exponential family

$$\mathrm{EM} := \{p(\cdot,\Theta), \Theta \in \Lambda\}, \quad p(x,\Theta) := \exp\big(\langle \Theta, c(x)\rangle - \psi(\Theta)\big)$$

Let be given an exponential family EM, an initial density p_0 contained in EM and a diffusion term σ. Consider an SDE

$$\mathrm{d}X_t = u(X_t, t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = x_0 \quad (\star)$$

Problem (□□)

Given a curve $t \mapsto p(\cdot,\Theta_t)$ in EM, find a drift in u whose related SDE $(\star)$ has a solution with marginal density $p(\cdot,\Theta_t)$.

- example: normal density and constant diffusion term
  $\rightarrow$ linear drift $u(x,t) = A_t + B_t x$

SDE with prescribed marginal law (cont'd)

Theorem

Consider an SDE $(\star)$ with drift term

$$u_i(x,t) = \frac{1}{2}\sum_{j=1}^{n}\frac{\partial}{\partial x_j}a_{ij}(x) + \frac{1}{2}\sum_{j=1}^{n}a_{ij}(x)\left\langle \Theta_t, \frac{\partial c(x)}{\partial x_j}\right\rangle - \left\langle \dot{\Theta}_t, \int_{-\infty}^{x_i}\varphi_i((x_{-i},\xi_i),\Theta_t)\exp\left[\left\langle\Theta_t, c(x_{-i},\xi_i) - c(x)\right\rangle\right]\mathrm{d}\xi_i\right\rangle,$$

where the functions φ_i satisfy

$$\sum_{i=1}^{n}\left\langle \dot{\Theta}_t, \varphi_i\left((x_{-i},\xi_i),\Theta_t\right)\right\rangle\Big|_{\xi_i=x_i} = \left\langle \dot{\Theta}_t, c(x) - \nabla_\Theta\psi(\Theta_t)\right\rangle.$$

This solves Problem (□□).

- extension to mixture of exponential families possible (→ arXiv:1508.00506)
- one-dimensional non-mixture case [Brigo'00]

SDE with prescribed marginal law (Gaussian density)

- normal density $\Theta = (\eta, \theta) := \left(S^{-1}m, -\frac{1}{2}S^{-1}\right)$

Corollary (Gaussian density)

$$u(x,t) = \frac{1}{2}\mathrm{div}a(x) + \frac{1}{4}\theta_t^{-1}\dot{\theta}_t\theta_t^{-1}\eta_t - \frac{1}{2}\theta_t^{-1}\dot{\eta}_t - \frac{1}{2}\theta_t^{-1}\dot{\theta}_t x + a(x)\left(\frac{1}{2}\eta_t + \theta_t x\right)$$

- example: constant diffusion term
  $\to$ linear drift $u(x,t) = A_t + B_t x$
- ansatz: $u(x,t) = \frac{1}{2}\mathrm{div}a(x) + A_t + B_t x + a(x)\left(C_t + D_t x\right)$

$$\frac{\mathrm{d}C_t}{\mathrm{d}t} = -D_t A_t - B_t^\top C_t, \qquad \frac{\mathrm{d}D_t}{\mathrm{d}t} = -2D_t B_t \quad (\star\star)$$

Optimal control problem reformulation of Problem (□)

- const functional

$$
\begin{aligned}
&\mathrm{D}\big(Q \| \Pi_{\text{prior}}\big) + \mathbb{E}_Q\big[H_T(\cdot, y)\big] \\
&= \int_0^T \mathrm{E}_Q\Bigg[\frac{1}{2}\|u(X_t,t) - f(X_t)\|^2_{a(X_t)} + y_t\Bigg(u(X_t,t)^\top \nabla h(X_t) \\
&\quad + \frac{1}{2}\sigma(X_t)^\top \nabla^2 h(X_t)\sigma(X_t)\Bigg) + \frac{1}{2}\|h(X_t)\|^2\Bigg]\mathrm{d}t - y_T \mathbb{E}_Q\big[h(X_T)\big]
\end{aligned}
$$

- states: m_t, S_t, C_t, D_t
  ODEs for m_t, S_t from $(\star)$, ODEs for C_t, D_t from $(\star\star)$
- control input (decision variables): A_t, B_t
- motivation
  - necessary conditions via Pontryagin's maximum principle
  - semidefinite programming approach for certain problem classes
  - infinite dimensional linear programming approach

Parameter inference

- signal process

$$dX_t^\kappa = f(X_t^\kappa, \kappa)dt + \sigma(X_t^\kappa, \kappa)dW_t, \quad X_0^\kappa = x_0, \quad 0 \le t \le T$$

- non-negativity of relative entropy

$$\underbrace{I(H_T^\kappa(\cdot, y))}_{\color{red}{\text{negative log-likelihood}}} \quad \le \underbrace{\mathrm{D}\big(Q \| \Pi_{\text{prior}}^\kappa\big) + \mathbb{E}_Q\big[H_T^\kappa(\cdot, y)\big]}_{\color{green}{=: \mathrm{F}(Q_i, \kappa) \text{ objective function}}}$$

EM-type algorithm

initialize	$i = 0, \ \kappa_i := \hat{\kappa}_0$
while	$i \le M$
Step 1:	compute Q_i by solving Problem ($\square$) with parameter κ_i
Step 2:	update parameter as $\kappa_{i+1} \in \arg\min_\kappa \mathrm{F}(Q_i, \kappa)$
Step 3:	set $i \to i + 1$

Example 1: Geometric Brownian motion

- signal process $\mathrm{d}X_t = \kappa X_t \mathrm{d}t + \lambda X_t \mathrm{d}W_t, \quad X_0 = x_0 \sim \log \mathrm{N}(\mu, \sigma)$
- discrete observations $Y_k = X_{t_k} + \rho_k$, $k = 1, \cdots, N$, $\rho_k \sim \mathrm{N}(1, R)$
- values: $\kappa = 1$, $\lambda = 0.1$, $R = 0.15$, $T = 0.2s$, $\mu = 0$, $\sigma = 0.25$, $N = 4$, $t_1 = T/4$, $t_2 = T/2$, $t_3 = 3T/4$ and $t_4 = T$

variational approximation

- drift function $u(x, t) = A_t + (\lambda^2 + B_t)x + \lambda^2 x^2 (C_t + D_t x)$

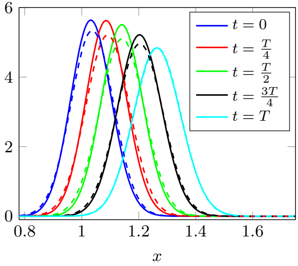

Figure: Smoothing density (PDE vs variational approx.)

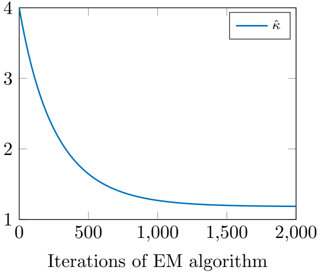

Figure: Parameter inference, $\hat{\kappa} = 1.18$

Example 2: Cox-Ingersoll-Ross

- signal process
  $dX_t = \kappa(b - X_t)dt + \lambda\sqrt{X_t}dW_t, \quad X_0 = x_0 \sim N(\mu, \sigma)$
- discrete observations $Y_k = X_{t_k} + \rho_k$, $k = 1, \cdots, N$, $\rho_k \sim N(1, R)$
- values: $\lambda = 0.2$, $\kappa = 1$, $b = 0.3$, $\mu = 1$, $\sigma = 0.1$, $R = 0.1$,
  $T = 0.3$s, $N = 2$, $t_1 = T/2$ and $t_2 = T$

variational approximation

- $u(x,t) = \frac{1}{2}\lambda^2 + A(t) + B(t)x + \lambda^2 x(C(t) + D(t)x)$

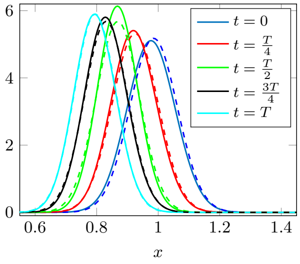

Figure: Smoothing density (PDE vs variational approx.)

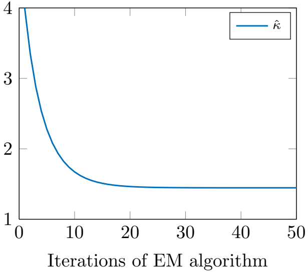

Figure: Parameter inference, $\hat{\kappa} = 1.44$

Conclusion and Outlook

- Conclusion
  - variational approximation to the smoothing density
  - requires solving an optimal control problem
  - parameter inference via maximum likelihood method
- Outlook
  - solve the underlying optimal control problem
    - necessary conditions via PMP (shooting method)
    - semidefinite programming approach for certain problem classes
    - infinite dimensional linear programming approach
  - guarantees for the EM-alogrithm

Acknowledgements and References

Special thanks to

- Debasish Chatterjee
- John Lygeros
- Peyman Mohajerin Esfahani

References

- Sutter, Ganguly & Koeppl, *A variational approach to path estimation and parameter inference of hidden diffusion processes*, arXiv-1508.00506, 2015

- Archambeau & Opper, *Approximate inference for continuous-time Markov processes*, Bayesian Time Series Models, Cambridge University Press, 2011, pp. 125-140