Gain Function Approximation in the Feedback Particle Filter

Conference on Decision and Control
Las Vegas, December, 2016

Amirhossein Taghvaei
Joint work with P. G. Mehta

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

Dec 14, 2016

ILLINOIS

Gain Function Approximation in FPF
Problem formulation

Poisson equation:
$$-\frac{1}{\rho(x)}\nabla\cdot(\rho(x)\nabla\phi(x)) = h(x)-\hat{h}$$
$$\int_{\mathbb{R}^d}\phi(x)\rho(x)\,\mathrm{d}x = 0$$

- $\rho:\mathbb{R}^d\to\mathbb{R}^+$ (prob. density)
- $h:\mathbb{R}^d\to\mathbb{R}$ (given function),
- $\hat{h} := \int h(x)\rho(x)\,\mathrm{d}x$
- $\phi:\mathbb{R}^d\to\mathbb{R}$ (solution)

Problem:

Given: $\{X^1,\ldots,X^N\}\overset{\text{i.i.d}}{\sim}\rho$

Find: $\{\nabla\phi(X^1),\ldots,\nabla\phi(X^N)\}$ (approximately)

Almost like a statistical learning problem

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson Equation in Nonlinear Filtering. SICON, 2015

Poisson equation: $$-\frac{1}{\rho(x)}\nabla\cdot(\rho(x)\nabla\phi(x)) = h(x) - \hat{h}$$
$$\int_{\mathbb{R}^d}\phi(x)\rho(x)\,\mathrm{d}x = 0$$

- $\rho : \mathbb{R}^d \to \mathbb{R}^+$ (prob. density)
- $h : \mathbb{R}^d \to \mathbb{R}$ (given function),
- $\hat{h} := \int h(x)\rho(x)\,\mathrm{d}x$
- $\phi : \mathbb{R}^d \to \mathbb{R}$ (solution)

Problem:

Given: $\{X^1,\ldots,X^N\} \overset{\text{i.i.d}}{\sim} \rho$

Find: $\{\nabla\phi(X^1),\ldots,\nabla\phi(X^N)\}$ (approximately)

Almost like a statistical learning problem

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson Equation in Nonlinear Filtering. SICON, 2015

Poisson equation:
$$-\frac{1}{\rho(x)}\nabla\cdot(\rho(x)\nabla\phi(x)) = h(x) - \hat{h}$$
$$\int_{\mathbb{R}^d} \phi(x)\rho(x)\,\mathrm{d}x = 0$$

- $\rho : \mathbb{R}^d \to \mathbb{R}^+$ (prob. density)
- $h : \mathbb{R}^d \to \mathbb{R}$ (given function),
- $\hat{h} := \int h(x)\rho(x)\,\mathrm{d}x$
- $\phi : \mathbb{R}^d \to \mathbb{R}$ (solution)

Problem:

Given: $\{X^1, \ldots, X^N\} \overset{\text{i.i.d}}{\sim} \rho$

Find: $\{\nabla\phi(X^1), \ldots, \nabla\phi(X^N)\}$ (approximately)

Almost like a statistical learning problem

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson Equation in Nonlinear Filtering. SICON, 2015

Feedback Particle Filter

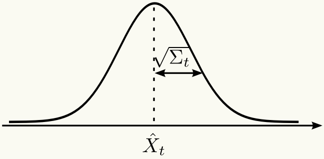
Generalization of the Kalman Filter

Kalman Filter:

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + \mathrm{d}B_t$$
$$\mathrm{d}Z_t = HX_t\,\mathrm{d}t + \mathrm{d}W_t$$

$\mathsf{P}(X_t|\mathcal{Z}_t)$ = Gaussian $N(\hat{X}_t, \Sigma_t)$,

$$\mathrm{d}\hat{X}_t = A\hat{X}_t\,\mathrm{d}t + \mathsf{K}_t(\,\mathrm{d}Z_t - H\hat{X}_t\,\mathrm{d}t)$$
$$\frac{\mathrm{d}\Sigma_t}{\mathrm{d}t} = \ldots \text{(Riccati equation)}$$



Feedback Particle Filter:

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \mathrm{d}B_t$$
$$\mathrm{d}Z_t = h(X_t)\,\mathrm{d}t + \mathrm{d}W_t$$

$\mathsf{P}(X_t|\mathcal{Z}_t) \approx$ empirical dist. $\{X^1, \ldots, X^N\}$,

$$\mathrm{d}X^i_t = a(X^i_t)\,\mathrm{d}t + \mathrm{d}B^i_t + \mathsf{K}_t(X^i_t) \circ (\,\mathrm{d}Z_t - \frac{h(X^i_t) + \hat{h}_t}{2}\,\mathrm{d}t)$$

Challenge: Compute the gain function $\mathsf{K}_t := \nabla\phi$ from Poisson eq.

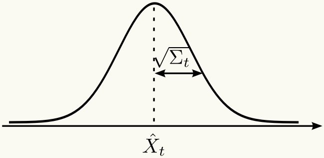
T. Yang, P. G. Mehta, and S. P. Meyn. feedback particle filter, *TAC*, 2013
T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter, *Automatica*, 2016

Kalman Filter:

$$dX_t = AX_t\,dt + dB_t$$
$$dZ_t = HX_t\,dt + dW_t$$

$\mathsf{P}(X_t|\mathcal{Z}_t) =$ Gaussian $N(\hat{X}_t, \Sigma_t)$,

$$d\hat{X}_t = A\hat{X}_t\,dt + \mathsf{K}_t(\,dZ_t - H\hat{X}_t\,dt)$$
$$\frac{d\Sigma_t}{dt} = \ldots \text{(Riccati equation)}$$

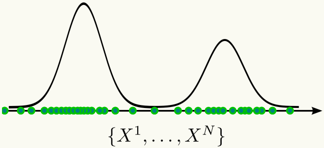


Feedback Particle Filter:

$$dX_t = a(X_t)\,dt + dB_t$$
$$dZ_t = h(X_t)\,dt + dW_t$$

$\mathsf{P}(X_t|\mathcal{Z}_t) \approx$ empirical dist. $\{X^1, \ldots, X^N\}$,

$$dX_t^i = a(X_t^i)\,dt + dB_t^i$$
$$+\,\mathsf{K}_t(X_t^i) \circ (\,dZ_t - \frac{h(X_t^i) + \hat{h}_t}{2}\,dt)$$



Challenge: Compute the gain function $\mathsf{K}_t := \nabla\phi$ from Poisson eq.

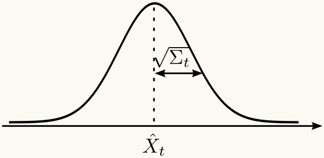
T. Yang, P. G. Mehta, and S. P. Meyn. feedback particle filter, *TAC*, 2013
T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter, *Automatica*, 2016

Kalman Filter:

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + \mathrm{d}B_t$$
$$\mathrm{d}Z_t = HX_t\,\mathrm{d}t + \mathrm{d}W_t$$

$\mathsf{P}(X_t|\mathcal{Z}_t) =$ Gaussian $N(\hat{X}_t, \Sigma_t)$,

$$\mathrm{d}\hat{X}_t = A\hat{X}_t\,\mathrm{d}t + \mathsf{K}_t(\mathrm{d}Z_t - H\hat{X}_t\,\mathrm{d}t)$$
$$\frac{\mathrm{d}\Sigma_t}{\mathrm{d}t} = \ldots \text{(Riccati equation)}$$

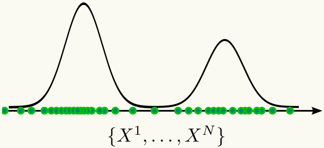


Feedback Particle Filter:

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \mathrm{d}B_t$$
$$\mathrm{d}Z_t = h(X_t)\,\mathrm{d}t + \mathrm{d}W_t$$

$\mathsf{P}(X_t|\mathcal{Z}_t) \approx$ empirical dist. $\{X^1, \ldots, X^N\}$,

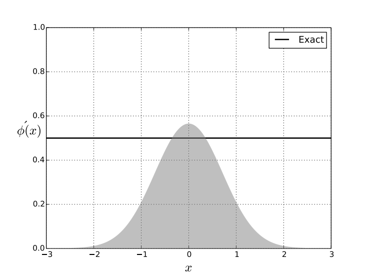
$$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \mathrm{d}B_t^i$$
$$+ \mathsf{K}_t(X_t^i) \circ (\mathrm{d}Z_t - \frac{h(X_t^i) + \hat{h}_t}{2}\,\mathrm{d}t)$$



Challenge: Compute the gain function $\mathsf{K}_t := \nabla\phi$ from Poisson eq.

T. Yang, P. G. Mehta, and S. P. Meyn. feedback particle filter, *TAC*, 2013
T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter, *Automatica*, 2016

Gaussian distribution
Linear observation

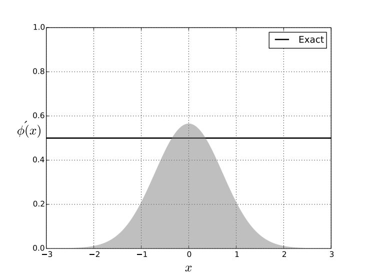


$\mathsf{K}_t(x) = \text{constant} \quad \text{(Kalman gain)}$

Non-Gaussian distribution
Nonlinear observation

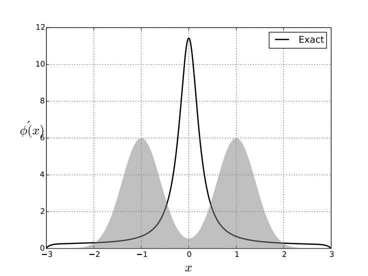
$\mathsf{K}_t(x) = \ldots \text{(Nonlinear gain)}$

Gaussian distribution
Linear observation



$\mathsf{K}_t(x) =$ constant (Kalman gain)

Non-Gaussian distribution
Nonlinear observation



$\mathsf{K}_t(x) = \ldots$ (Nonlinear gain)

Poisson equation: $$-\frac{1}{\rho}\nabla\cdot(\rho\nabla\phi) = h - \hat{h}$$

Weighted Laplacian: $$\Delta_\rho\phi := \frac{1}{\rho}\nabla\cdot(\rho\nabla\phi) = \Delta\phi + \nabla\log\rho\cdot\nabla\phi$$

PDE

- Markov Diffusion operators [D. Bakry, et. al. 2013]
- Heat kernels [A. Grigoryan, 2009]
- Optimal transportation [C. Villani, 2003]

Stochastic analysis

- Simulation and optimization theory for Markov models [S. Meyn, R. Tweedie, 2012]

Statistical learning

- Nonlinear dimensionality reduction [M. Belkin, 2003]
- Diffusion maps [R. Coifman, S. Lafon, 2006]
- Spectral clustering [M. Hein, et. al. 2006]

FPF (theory and application):

- T. Yang, et. al. Automatica, 2015.
- K. Berntorp, Fusion, 2015.
- P. M. Stano, et. al., 2014.

} (Constant gain and Galerkin approximation)

Gain function approximation:

- K. Berntorp, P. Grover, ACC, 2016. (Data driven approach based on POD)
- Y. Matsuura, et. al. 2016. (Continuation method)
- A. Radhakrishnan, A. Devraj, and S. Meyn. CDC, 2016. (TD learning)

Also in other nonlinear filtering algorithms

- Particle flow filter [F. Daum, J. Huang, 2010]
- Approxiamte representation of SPDE [D. Crisan, J. Xiong, 2005]
- Dynamical systems framework for intermittent data assimilation [S. Riech 2011]
- Continuous-discrete time FPF [T. Yang, et. al. 2014]

Different Formulations of the Poisson Equation

Weak formulation: (Galerkin)

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Semigroup formulation: (kernel-based)

$$\phi = P\phi + \tilde{h}$$

where $P := e^{\epsilon\Delta_\rho}$ and $\tilde{h} := \int_0^t e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Variational formulation: (Neural net ?)

$$\min_{\phi \in H_0^1(\mathbb{R}^d, \rho)} \mathsf{E}\left[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)(h(X) - \hat{h})\right]$$

where $X \sim \rho$

Weak formulation: (Galerkin)

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Semigroup formulation: (kernel-based)

$$\phi = P\phi + \tilde{h}$$

where $P := e^{\epsilon\Delta_\rho}$ and $\tilde{h} := \int_0^t e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Variational formulation: (Neural net ?)

$$\min_{\phi \in H_0^1(\mathbb{R}^d, \rho)} \mathsf{E}\left[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)(h(X) - \hat{h})\right]$$

where $X \sim \rho$

Different Formulations of the Poisson Equation

Weak formulation: (Galerkin)

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Semigroup formulation: (kernel-based)

$$\phi = P\phi + \tilde{h}$$

where $P := e^{\epsilon\Delta_\rho}$ and $\tilde{h} := \int_0^t e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Variational formulation: (Neural net ?)

$$\min_{\phi \in H_0^1(\mathbb{R}^d,\rho)} \mathsf{E}\left[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)(h(X) - \hat{h})\right]$$

where $X \sim \rho$

Weak formulation: (Galerkin)

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x) g(x) \rho(x)\,\mathrm{d}x$

Semigroup formulation: (kernel-based)

$$\phi = P\phi + \tilde{h}$$

where $P := e^{\epsilon \Delta_\rho}$ and $\tilde{h} := \int_0^t e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Variational formulation: (Neural net ?)

$$\min_{\phi \in H_0^1(\mathbb{R}^d, \rho)} \mathsf{E}\left[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)(h(X) - \hat{h})\right]$$

where $X \sim \rho$

Weak formulation: (Galerkin)

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x) g(x) \rho(x) \, \mathrm{d}x$

Semigroup formulation: (kernel-based)

$$\phi = P\phi + \tilde{h}$$

where $P := e^{\epsilon \Delta_\rho}$ and $\tilde{h} := \int_0^t e^{s\Delta_\rho}(h - \hat{h}) \, \mathrm{d}s$

Variational formulation: (Neural net ?)

$$\min_{\phi \in H_0^1(\mathbb{R}^d, \rho)} \mathsf{E}\left[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)(h(X) - \hat{h})\right]$$

where $X \sim \rho$

Different Formulations of the Poisson Equation

Weak formulation: (Galerkin)

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Semigroup formulation: (kernel-based)

$$\phi = P\phi + \tilde{h}$$

where $P := e^{\epsilon\Delta_\rho}$ and $\tilde{h} := \int_0^t e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Variational formulation: (Neural net ?)

$$\min_{\phi \in H_0^1(\mathbb{R}^d, \rho)} \mathsf{E}\left[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)(h(X) - \hat{h})\right]$$

where $X \sim \rho$

Weak formulation: (Galerkin)

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Semigroup formulation: (kernel-based)

$$\phi = P\phi + \tilde{h}$$

where $P := e^{\epsilon\Delta_\rho}$ and $\tilde{h} := \int_0^t e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Variational formulation: (Neural net ?)

$$\min_{\phi \in H_0^1(\mathbb{R}^d,\rho)} \mathsf{E}\left[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)(h(X) - \hat{h})\right]$$

where $X \sim \rho$

Strong form:

$$-\frac{1}{\rho(x)}\nabla\cdot(\rho(x)\nabla\phi(x)) = h(x) - \hat{h}$$

Weak form:

$$\langle \nabla\phi, \nabla\psi\rangle = \langle h - \hat{h}, \psi\rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g\rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Galerkin approximation:

$$\langle \nabla\phi^{(M)}, \nabla\psi\rangle = \langle h - \hat{h}, \psi\rangle, \quad \forall \psi \in S$$

where $S = \text{span}\{\psi_1, \ldots, \psi_M\}$

Empirical approximation

$$\frac{1}{N}\sum_{i=1}^{N}\nabla\phi^{(M)}(X^i)\cdot\nabla\psi(X^i) = \frac{1}{N}\sum_{i=1}^{N}(h(X^i) - \hat{h})\psi(X^i), \quad \forall \psi \in S$$

where $X^i \sim \rho$

Strong form:

$$-\frac{1}{\rho(x)}\nabla\cdot(\rho(x)\nabla\phi(x)) = h(x) - \hat{h}$$

Weak form:

$$\langle\nabla\phi,\nabla\psi\rangle = \langle h-\hat{h},\psi\rangle,\quad \forall\psi\in H^1(\mathbb{R}^d,\rho)$$

where $\langle f,g\rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Galerkin approximation:

$$\langle\nabla\phi^{(M)},\nabla\psi\rangle = \langle h-\hat{h},\psi\rangle,\quad \forall\psi\in S$$

where $S = \text{span}\{\psi_1,\ldots,\psi_M\}$

Empirical approximation

$$\frac{1}{N}\sum_{i=1}^{N}\nabla\phi^{(M)}(X^i)\cdot\nabla\psi(X^i) = \frac{1}{N}\sum_{i=1}^{N}(h(X^i)-\hat{h})\psi(X^i),\quad \forall\psi\in S$$

where $X^i\sim\rho$

Strong form:

$$-\frac{1}{\rho(x)}\nabla \cdot (\rho(x)\nabla\phi(x)) = h(x) - \hat{h}$$

Weak form:

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Galerkin approximation:

$$\langle \nabla\phi^{(M)}, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in S$$

where $S = \text{span}\{\psi_1, \ldots, \psi_M\}$

Empirical approximation

$$\frac{1}{N}\sum_{i=1}^{N} \nabla\phi^{(M)}(X^i) \cdot \nabla\psi(X^i) = \frac{1}{N}\sum_{i=1}^{N}(h(X^i) - \hat{h})\psi(X^i), \quad \forall \psi \in S$$

where $X^i \sim \rho$

Strong form:

$$-\frac{1}{\rho(x)}\nabla\cdot(\rho(x)\nabla\phi(x)) = h(x) - \hat{h}$$

Weak form:

$$\langle \nabla\phi, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in H^1(\mathbb{R}^d, \rho)$$

where $\langle f, g \rangle := \int f(x)g(x)\rho(x)\,\mathrm{d}x$

Galerkin approximation:

$$\langle \nabla\phi^{(M)}, \nabla\psi \rangle = \langle h - \hat{h}, \psi \rangle, \quad \forall \psi \in S$$

where $S = \text{span}\{\psi_1, \ldots, \psi_M\}$

Empirical approximation

$$\frac{1}{N}\sum_{i=1}^{N} \nabla\phi^{(M)}(X^i)\cdot\nabla\psi(X^i) = \frac{1}{N}\sum_{i=1}^{N}(h(X^i) - \hat{h})\psi(X^i), \quad \forall \psi \in S$$

where $X^i \sim \rho$

- Select basis functions $\{\psi_1, \ldots, \psi_M\}$
- Express the approximate solution as

$$\phi^{(M,N)}(x) = \sum_{m=1}^{M} c_m \psi_m(x)$$

- Obtain $c = (c_1, \ldots, c_M)$ by solving

$$Ac = b$$

where

$$A_{ml} = \langle \nabla\psi_m, \nabla\psi_l \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \nabla\psi_m(X^i) \cdot \nabla\psi_l(X^i)$$

$$b_m = \langle \psi_m, h \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \psi_m(X^i) h(X^i) - \hat{h})$$

Special case: The basis functions are eigenfunctions of Δ_ρ

$$\mathsf{E}\left[\|\nabla\phi - \nabla\phi^{(M,N)}\|_{L^2}\right] \le \underbrace{\frac{1}{\sqrt{\lambda_M}}\|h - \Pi_S h\|_{L^2}}_{\text{Bias}} + \underbrace{\frac{1}{\sqrt{N}}\|h\|_\infty \sqrt{\sum_{m=1}^{M}\frac{1}{\lambda_m}}}_{\text{Variance}}$$

- ϕ: exact solution
- $\phi^{(M,N)}$: approximate solution
- $\{\lambda_m\}_{m=1}^\infty$ the eigenvalues

It is a projection:

$$\nabla\phi^{(M)} = \arg\min_{\nabla\psi\in S} \|\nabla\phi - \nabla\psi\|_2$$

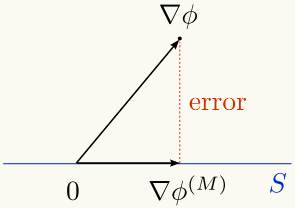
Special case: The basis functions are eigenfunctions of Δ_ρ

$$\mathsf{E}\left[\|\nabla\phi - \nabla\phi^{(M,N)}\|_{L^2}\right] \leq \underbrace{\frac{1}{\sqrt{\lambda_M}}\|h - \Pi_S h\|_{L^2}}_{\text{Bias}} + \underbrace{\frac{1}{\sqrt{N}}\|h\|_\infty \sqrt{\sum_{m=1}^{M} \frac{1}{\lambda_m}}}_{\text{Variance}}$$

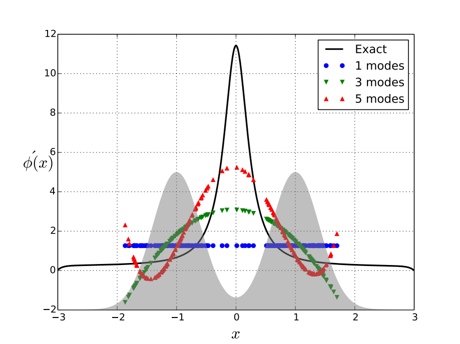
- ϕ: exact solution
- $\phi^{(M,N)}$: approximate solution
- $\{\lambda_m\}_{m=1}^{\infty}$ the eigenvalues

It is a projection:

$$\nabla\phi^{(M)} = \arg\min_{\nabla\psi \in S} \|\nabla\phi - \nabla\psi\|_2$$



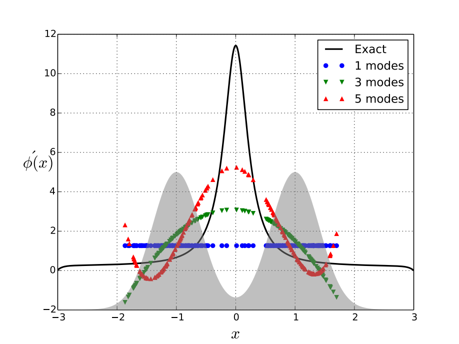
Galerkin Approximation
Numerical result



Issues:

- Choice of basis functions
- Singularity of A
- Computationally scales with $O(Nd^p)$

Galerkin Approximation
Numerical result



Issues:

- Choice of basis functions
- Singularity of A
- Computationally scales with $O(Nd^p)$

Poisson equation: $-\Delta_\rho \phi = h - \hat{h}$

Semigroup identity: $e^{\epsilon \Delta_\rho} = I + \int_0^\epsilon e^{s\Delta_\rho} \Delta_\rho \, \mathrm{d}s$

Semigroup formulation:

$$\phi = e^{\epsilon \Delta_\rho} \phi + \tilde{h}$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho} (h - \hat{h}) \, \mathrm{d}s$

Kernel representation:

$$\phi(x) = \int \tilde{k}_\epsilon(x,y) \phi(y) \rho(y) \, \mathrm{d}y + \tilde{h}(x)$$

Empirical approximation:

$$\phi(x) = \frac{1}{N} \sum_{i=1}^{N} \tilde{k}_\epsilon(x, X^i) \phi(X^i) + \tilde{h}(x)$$

But $\tilde{k}_\epsilon(x,y) = ?$

Poisson equation: $-\Delta_\rho \phi = h - \hat{h}$

Semigroup identity: $e^{\epsilon\Delta_\rho} = I + \int_0^\epsilon e^{s\Delta_\rho}\Delta_\rho \, \mathrm{d}s$

Semigroup formulation:

$$\phi = e^{\epsilon\Delta_\rho}\phi + \tilde{h}$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h}) \, \mathrm{d}s$

Kernel representation:

$$\phi(x) = \int \tilde{k}_\epsilon(x,y)\phi(y)\rho(y) \, \mathrm{d}y + \tilde{h}(x)$$

Empirical approximation:

$$\phi(x) = \frac{1}{N}\sum_{i=1}^{N} \tilde{k}_\epsilon(x, X^i)\phi(X^i) + \tilde{h}(x)$$

But $\tilde{k}_\epsilon(x,y) = ?$

Poisson equation: $-\Delta_\rho \phi = h - \hat{h}$

Semigroup identity: $e^{\epsilon \Delta_\rho} = I + \int_0^\epsilon e^{s\Delta_\rho} \Delta_\rho \, \mathrm{d}s$

Semigroup formulation:

$$\phi = e^{\epsilon \Delta_\rho} \phi + \tilde{h}$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho} (h - \hat{h}) \, \mathrm{d}s$

Kernel representation:

$$\phi(x) = \int \tilde{k}_\epsilon(x,y) \phi(y) \rho(y) \, \mathrm{d}y + \tilde{h}(x)$$

Empirical approximation:

$$\phi(x) = \frac{1}{N} \sum_{i=1}^N \tilde{k}_\epsilon(x, X^i) \phi(X^i) + \tilde{h}(x)$$

But $\tilde{k}_\epsilon(x,y) = ?$

Poisson equation: $-\Delta_\rho \phi = h - \hat{h}$

Semigroup identity: $e^{\epsilon\Delta_\rho} = I + \int_0^\epsilon e^{s\Delta_\rho}\Delta_\rho \,\mathrm{d}s$

Semigroup formulation:

$$\phi = e^{\epsilon\Delta_\rho}\phi + \tilde{h}$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Kernel representation:

$$\phi(x) = \int \tilde{k}_\epsilon(x,y)\phi(y)\rho(y)\,\mathrm{d}y + \tilde{h}(x)$$

Empirical approximation:

$$\phi(x) = \frac{1}{N}\sum_{i=1}^{N} \tilde{k}_\epsilon(x, X^i)\phi(X^i) + \tilde{h}(x)$$

But $\tilde{k}_\epsilon(x,y) = ?$

Poisson equation: $-\Delta_\rho \phi = h - \hat{h}$

Semigroup identity: $e^{\epsilon \Delta_\rho} = I + \int_0^\epsilon e^{s\Delta_\rho} \Delta_\rho \, \mathrm{d}s$

Semigroup formulation:

$$\phi = e^{\epsilon \Delta_\rho} \phi + \tilde{h}$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho} (h - \hat{h}) \, \mathrm{d}s$

Kernel representation:

$$\phi(x) = \int \tilde{k}_\epsilon(x, y) \phi(y) \rho(y) \, \mathrm{d}y + \tilde{h}(x)$$

Empirical approximation:

$$\phi(x) = \frac{1}{N} \sum_{i=1}^{N} \tilde{k}_\epsilon(x, X^i) \phi(X^i) + \tilde{h}(x)$$

But $\tilde{k}_\epsilon(x, y) = ?$

Kernel-based Approximation

Special case: $\rho = 1$

$$e^{\epsilon\Delta} f(x) = \int g_\epsilon(x,y) f(y) \, \mathrm{d}y. \quad \text{(for all } \epsilon > 0)$$

where g_ϵ is the Gaussian kernel.

In general:

$$e^{\epsilon\Delta_\rho} f(x) \approx \int \frac{1}{n_\epsilon(x)} \frac{g_\epsilon(x,y)}{\sqrt{\int g_\epsilon(y,z)\rho(z)\,\mathrm{d}z}} f(y)\rho(y)\,\mathrm{d}y := T_\epsilon f(x) \quad \text{(for } \epsilon \downarrow 0)$$

where n_ϵ is normalizing constant.

Empirical apprximation:

$$e^{\epsilon\Delta_\rho} f(x) \approx \sum_{j=1}^{N} \frac{1}{n_\epsilon^{(N)}(x)} \frac{g_\epsilon(x, X^j)}{\sqrt{\frac{1}{N}\sum_{l=1}^{N} g_\epsilon(X^j, X^l)}} f(X^j) := T_\epsilon^{(N)} f(x)$$

where $n_\epsilon^{(N)}$ is normalizing constant.

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,
M. Hein, J. Audibert, U. Von Luxburg, Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

Kernel-based Approximation

Special case: $\rho = 1$

$$e^{\epsilon\Delta} f(x) = \int g_\epsilon(x,y) f(y)\,\mathrm{d}y. \quad \text{(for all } \epsilon > 0\text{)}$$

where g_ϵ is the Gaussian kernel.

In general:

$$e^{\epsilon\Delta_\rho} f(x) \approx \int \frac{1}{n_\epsilon(x)} \frac{g_\epsilon(x,y)}{\sqrt{\int g_\epsilon(y,z)\rho(z)\,\mathrm{d}z}} f(y)\rho(y)\,\mathrm{d}y := T_\epsilon f(x) \quad \text{(for } \epsilon \downarrow 0\text{)}$$

where n_ϵ is normalizing constant.

Empirical apprximation:

$$e^{\epsilon\Delta_\rho} f(x) \approx \sum_{j=1}^{N} \frac{1}{n_\epsilon^{(N)}(x)} \frac{g_\epsilon(x,X^j)}{\sqrt{\frac{1}{N}\sum_{l=1}^{N} g_\epsilon(X^j,X^l)}} f(X^j) := T_\epsilon^{(N)} f(x)$$

where $n_\epsilon^{(N)}$ is normalizing constant.

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,
M. Hein, J. Audibert, U. Von Luxburg, Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

Kernel-based Approximation

Special case: $\rho = 1$

$$e^{\epsilon\Delta} f(x) = \int g_\epsilon(x,y) f(y)\,\mathrm{d}y. \quad \text{(for all } \epsilon > 0\text{)}$$

where g_ϵ is the Gaussian kernel.

In general:

$$e^{\epsilon\Delta_\rho} f(x) \approx \int \frac{1}{n_\epsilon(x)} \frac{g_\epsilon(x,y)}{\sqrt{\int g_\epsilon(y,z)\rho(z)\,\mathrm{d}z}} f(y)\rho(y)\,\mathrm{d}y := T_\epsilon f(x) \quad \text{(for } \epsilon \downarrow 0\text{)}$$

where n_ϵ is normalizing constant.

Empirical apprximation:

$$e^{\epsilon\Delta_\rho} f(x) \approx \sum_{j=1}^{N} \frac{1}{n_\epsilon^{(N)}(x)} \frac{g_\epsilon(x,X^j)}{\sqrt{\frac{1}{N}\sum_{l=1}^{N} g_\epsilon(X^j,X^l)}} f(X^j) := T_\epsilon^{(N)} f(x)$$

where $n_\epsilon^{(N)}$ is normalizing constant.

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,
M. Hein, J. Audibert, U. Von Luxburg, Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

Exact solution:

$$\phi(x) = e^{\epsilon\Delta_\rho}\phi(x) + \tilde{h}(x)$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Approximation:

$$\phi_\epsilon^{(N)}(x) := T_\epsilon^{(N)}\phi_\epsilon^{(N)}(x) + \epsilon(h(x) - \hat{h}),$$

Numerics:

$$\Phi = \mathbf{T}\Phi + \epsilon(\mathbf{h} - \hat{h})$$

- $\Phi = (\Phi_\epsilon^{(N)}(X^1), \ldots, \Phi_\epsilon^{(N)}(X^N))$
- $\mathbf{T}_{ij} = \frac{1}{n_\epsilon(X^i)} \frac{g_\epsilon(X^i, X^j)}{\sqrt{\frac{1}{N}\sum_{l=1}^N g_\epsilon(X^i, X^l)}}$ (Markov matrix)

Gain function:

$$\nabla\phi_\epsilon^{(N)}(x) = \nabla(T_\epsilon^{(N)}\phi_\epsilon^{(N)})(x) + \epsilon\nabla h(x)$$

Exact solution:

$$\phi(x) = e^{\epsilon \Delta_\rho} \phi(x) + \tilde{h}(x)$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Approximation:

$$\phi_\epsilon^{(N)}(x) := T_\epsilon^{(N)} \phi_\epsilon^{(N)}(x) + \epsilon(h(x) - \hat{h}),$$

Numerics:

$$\Phi = \mathbf{T}\Phi + \epsilon(\mathbf{h} - \hat{h})$$

- $\Phi = (\Phi_\epsilon^{(N)}(X^1), \ldots, \Phi_\epsilon^{(N)}(X^N))$
- $\mathbf{T}_{ij} = \dfrac{1}{n_\epsilon(X^i)} \dfrac{g_\epsilon(X^i, X^j)}{\sqrt{\frac{1}{N}\sum_{l=1}^N g_\epsilon(X^i, X^l)}}$ (Markov matrix)

Gain function:

$$\nabla \phi_\epsilon^{(N)}(x) = \nabla (T_\epsilon^{(N)} \phi_\epsilon^{(N)})(x) + \epsilon \nabla h(x)$$

Exact solution:

$$\phi(x) = e^{\epsilon\Delta_\rho}\phi(x) + \tilde{h}(x)$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Approximation:

$$\phi_\epsilon^{(N)}(x) := T_\epsilon^{(N)}\phi_\epsilon^{(N)}(x) + \epsilon(h(x) - \hat{h}),$$

Numerics:

$$\Phi = \mathbf{T}\Phi + \epsilon(\mathbf{h} - \hat{h})$$

- $\Phi = (\Phi_\epsilon^{(N)}(X^1), \ldots, \Phi_\epsilon^{(N)}(X^N))$
- $\mathbf{T}_{ij} = \frac{1}{n_\epsilon(X^i)} \frac{g_\epsilon(X^i, X^j)}{\sqrt{\frac{1}{N}\sum_{l=1}^N g_\epsilon(X^i, X^l)}}$ (Markov matrix)

Gain function:

$$\nabla\phi_\epsilon^{(N)}(x) = \nabla(T_\epsilon^{(N)}\phi_\epsilon^{(N)})(x) + \epsilon\nabla h(x)$$

Exact solution:

$$\phi(x) = e^{\epsilon\Delta_\rho}\phi(x) + \tilde{h}(x)$$

where $\tilde{h} := \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\,\mathrm{d}s$

Approximation:

$$\phi_\epsilon^{(N)}(x) := T_\epsilon^{(N)}\phi_\epsilon^{(N)}(x) + \epsilon(h(x) - \hat{h}),$$

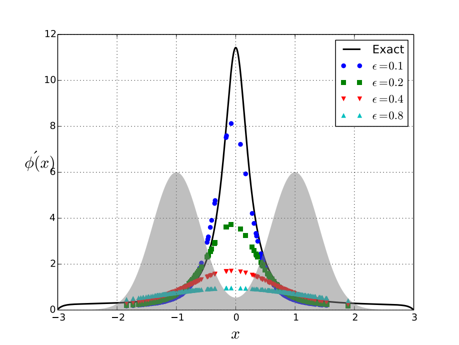
Numerics:

$$\Phi = \mathbf{T}\Phi + \epsilon(\mathbf{h} - \hat{h})$$

- $\Phi = (\Phi_\epsilon^{(N)}(X^1), \ldots, \Phi_\epsilon^{(N)}(X^N))$
- $\mathbf{T}_{ij} = \dfrac{1}{n_\epsilon(X^i)} \dfrac{g_\epsilon(X^i, X^j)}{\sqrt{\frac{1}{N}\sum_{l=1}^N g_\epsilon(X^i, X^l)}}$ (Markov matrix)

Gain function:

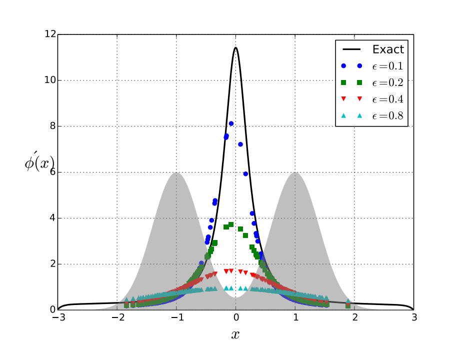
$$\nabla\phi_\epsilon^{(N)}(x) = \nabla(T_\epsilon^{(N)}\phi_\epsilon^{(N)})(x) + \epsilon\nabla h(x)$$



Properties

1. No singularity
2. Easy extension to Manifolds [C. Zhang, et. al. CDC 2015]
3. Better error bounds
4. Computational cost $O(N^2)$ (good in high dimensions)

Kernel-based approximation
Numerical result



Properties

1. No singularity
2. Easy extension to Manifolds [C. Zhang, et. al. CDC 2015]
3. Better error bounds
4. Computational cost $O(N^2)$ (good in high dimensions)

Kernel-based approximation
Error Analysis

$$\phi_\epsilon^{(N)} \xrightarrow[\text{Variance}]{N \uparrow \infty} \phi_\epsilon \xrightarrow[\text{Bias}]{\epsilon \downarrow 0} \phi$$

(Exact solution) $\phi(x) = e^{\epsilon\Delta_\rho}\phi(x) + \tilde{h}(x)$

(Kernel approx.) $\phi_\epsilon(x) = T_\epsilon \phi_\epsilon(x) + \epsilon(h(x) - \hat{h})$

(Empirical approx.) $\phi_\epsilon^{(N)}(x) = T_\epsilon^{(N)} \phi_\epsilon^{(N)}(x) + \epsilon(h(x) - \hat{h})$

Error bound:

$$\mathsf{E}\left[\|\nabla\phi - \nabla\phi_\epsilon^{(N)}\|_2\right] \leq \underbrace{O(\epsilon)}_{\text{Bias}} + \underbrace{O\left(\frac{1}{\epsilon^{1+d/4}\sqrt{N}}\right)}_{\text{Variance}}$$

on bounded domain.

Future work: Extension to unbounded domain

$$\phi_\epsilon^{(N)} \xrightarrow[\text{Variance}]{N \uparrow \infty} \phi_\epsilon \xrightarrow[\text{Bias}]{\epsilon \downarrow 0} \phi$$

(Exact solution) $\quad \phi(x) = e^{\epsilon \Delta_\rho} \phi(x) + \tilde{h}(x)$

(Kernel approx.) $\quad \phi_\epsilon(x) = T_\epsilon \phi_\epsilon(x) + \epsilon(h(x) - \hat{h})$

(Empirical approx.) $\quad \phi_\epsilon^{(N)}(x) = T_\epsilon^{(N)} \phi_\epsilon^{(N)}(x) + \epsilon(h(x) - \hat{h})$

Error bound:

$$\mathsf{E}\left[\|\nabla\phi - \nabla\phi_\epsilon^{(N)}\|_2\right] \leq \underbrace{O(\epsilon)}_{\text{Bias}} + \underbrace{O\left(\frac{1}{\epsilon^{1+d/4}\sqrt{N}}\right)}_{\text{Variance}}$$

on bounded domain.

Future work: Extension to unbounded domain

$$\phi_\epsilon^{(N)} \xrightarrow[\text{Variance}]{N \uparrow \infty} \phi_\epsilon \xrightarrow[\text{Bias}]{\epsilon \downarrow 0} \phi$$

(Exact solution) $\phi(x) = e^{\epsilon\Delta_\rho}\phi(x) + \tilde{h}(x)$

(Kernel approx.) $\phi_\epsilon(x) = T_\epsilon\phi_\epsilon(x) + \epsilon(h(x) - \hat{h})$

(Empirical approx.) $\phi_\epsilon^{(N)}(x) = T_\epsilon^{(N)}\phi_\epsilon^{(N)}(x) + \epsilon(h(x) - \hat{h})$

Error bound:

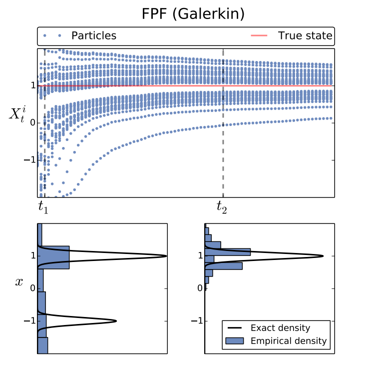
$$\mathsf{E}\left[\|\nabla\phi - \nabla\phi_\epsilon^{(N)}\|_2\right] \leq \underbrace{O(\epsilon)}_{\text{Bias}} + \underbrace{O(\frac{1}{\epsilon^{1+d/4}\sqrt{N}})}_{\text{Variance}}$$

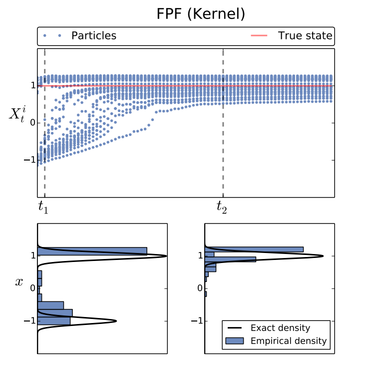
on bounded domain.

Future work: Extension to unbounded domain

Numerical result
Nonlinear filtering example

$$\mathrm{d}X_t = 0, \quad X_0 \sim \frac{1}{2}N(-1,\sigma^2) + \frac{1}{2}N(+1,\sigma^2)$$
$$\mathrm{d}Z_t = X_t\,\mathrm{d}t + \sigma_w\,\mathrm{d}W_t$$





Final slide

Thank you!