Feedback Particle Filter: Design, Approximation, and Error Analysis

Presented at the University of California, Los Angeles

Amirhossein Taghvaei[†]
Joint work with Prashant G. Mehta[*] and Sean P. Meyn[+]

[†]Department of Mechanical and Aerospace Engineering, Univ. of California, Irvine
[*]Univ. of Illinois, [+]Univ. of Florida

Nov 19, 2019

UCI

Outline

Feedback particle filter: Numerical algorithm for the nonlinear filtering problem

Part I: Background

- Filtering problem in discrete time
- particle-based methods
- Filtering problem in continuous time
- Feedback particle filter (FPF)

Part II: Gain function approximation in FPF

Outline

Feedback particle filter: Numerical algorithm for the nonlinear filtering problem

Part I: Background

- Filtering problem in discrete time
- particle-based methods
- Filtering problem in continuous time
- Feedback particle filter (FPF)

Part II: Gain function approximation in FPF

$$\begin{tikzcd} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\ & Z_{k-1} & Z_k & Z_{k+1} & \end{tikzcd}$$

State process: $X_k \sim a(\cdot|X_{k-1}), \qquad X_0 \sim \pi_0(\cdot)$

Observation process: $Z_k \sim l(\cdot|X_k)$

Filtering objective: Compute the posterior distribution $\pi_k(\cdot) := \mathsf{P}(X_k \in \cdot|Z_{1:k})$

Solution:

$$\begin{aligned} \pi_{k-1}(x) \quad &\longrightarrow \quad \tilde{\pi}_k(x) = \int a(x|x')\pi_{k-1}(x')\mathrm{d}x' \\ &\longrightarrow \quad \pi_k(x) = \frac{l(Z_k|x)\tilde{\pi}_k(x)}{\gamma} \quad \text{(Bayes rule)} \end{aligned}$$

- In general, no finite-dimensional solution $\Rightarrow$ approximations

$$\begin{tikzcd} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\ & Z_{k-1} & Z_k & Z_{k+1} & \end{tikzcd}$$

State process: $X_k \sim a(\cdot|X_{k-1}), \qquad X_0 \sim \pi_0(\cdot)$

Observation process: $Z_k \sim l(\cdot|X_k)$

Filtering objective: Compute the posterior distribution $\pi_k(\cdot) := \mathsf{P}(X_k \in \cdot|Z_{1:k})$

Solution:

$$\pi_{k-1}(x) \quad \longrightarrow \quad \tilde{\pi}_k(x) = \int a(x|x')\pi_{k-1}(x')\mathrm{d}x'$$

$$\longrightarrow \quad \pi_k(x) = \frac{l(Z_k|x)\tilde{\pi}_k(x)}{\gamma} \quad \text{(Bayes rule)}$$

- In general, no finite-dimensional solution ⇒ approximations

$$
\begin{tikzcd}
{} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\
& Z_{k-1} & Z_k & Z_{k+1} &
\end{tikzcd}
$$

State process: $X_k \sim a(\cdot|X_{k-1}), \qquad X_0 \sim \pi_0(\cdot)$

Observation process: $Z_k \sim l(\cdot|X_k)$

Filtering objective: Compute the posterior distribution $\pi_k(\cdot) := \mathsf{P}(X_k \in \cdot|Z_{1:k})$

Solution:

$$
\begin{aligned}
\pi_{k-1}(x) \quad &\longrightarrow \quad \tilde{\pi}_k(x) = \int a(x|x')\pi_{k-1}(x')\mathrm{d}x' \\
&\longrightarrow \quad \pi_k(x) = \frac{l(Z_k|x)\tilde{\pi}_k(x)}{\gamma} \quad \text{(Bayes rule)}
\end{aligned}
$$

- In general, no finite-dimensional solution $\Rightarrow$ approximations

$$\begin{tikzcd} \arrow[r] & X_{k-1} \arrow[r] \arrow[d] & X_k \arrow[r] \arrow[d] & X_{k+1} \arrow[r] \arrow[d] & {} \\ & Z_{k-1} & Z_k & Z_{k+1} & \end{tikzcd}$$

State process: $X_k \sim a(\cdot|X_{k-1}), \qquad X_0 \sim \pi_0(\cdot)$

Observation process: $Z_k \sim l(\cdot|X_k)$

Filtering objective: Compute the posterior distribution $\pi_k(\cdot) := \mathsf{P}(X_k \in \cdot|Z_{1:k})$

Solution:

$$\pi_{k-1}(x) \quad \longrightarrow \quad \tilde{\pi}_k(x) = \int a(x|x')\pi_{k-1}(x')\mathrm{d}x'$$

$$\longrightarrow \quad \pi_k(x) = \frac{l(Z_k|x)\tilde{\pi}_k(x)}{\gamma} \quad \text{(Bayes rule)}$$

- In general, no finite-dimensional solution $\Rightarrow$ approximations

Background: Particle-based methods

- Approximate π_k with empirical distribution of particles $\{X_k^1, \ldots, X_k^N\}$

$$\pi_k \approx \frac{1}{N}\sum_{i=1}^{N} \delta_{X_k^i}$$

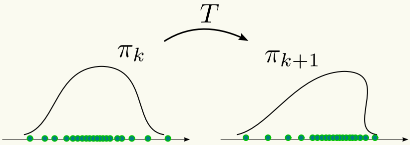
- Transform particles $\{X_k^i\}_{i=1}^N \sim \pi_k$ to particles $\{X_{k+1}^i\}_{i=1}^N \sim \pi_{k+1}$
- Different particle-based algorithms correspond to different transformations

Background: Particle-based methods

- Approximate π_k with empirical distribution of particles $\{X_k^1, \ldots, X_k^N\}$

$$\pi_k \approx \frac{1}{N} \sum_{i=1}^{N} \delta_{X_k^i}$$

- Transform particles $\{X_k^i\}_{i=1}^N \sim \pi_k$ to particles $\{X_{k+1}^i\}_{i=1}^N \sim \pi_{k+1}$



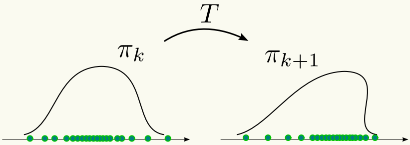
- Different particle-based algorithms correspond to different transformations

Background: Particle-based methods

- Approximate π_k with empirical distribution of particles $\{X_k^1, \dots, X_k^N\}$

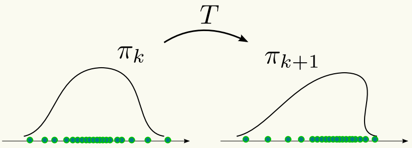
$$\pi_k \approx \frac{1}{N} \sum_{i=1}^{N} \delta_{X_k^i}$$

- Transform particles $\{X_k^i\}_{i=1}^N \sim \pi_k$ to particles $\{X_{k+1}^i\}_{i=1}^N \sim \pi_{k+1}$



- Different particle-based algorithms correspond to different transformations

Transportation viewpoint:



Given: $\{X_k^1, \dots, X_k^N\} \sim \pi_k$ and $l_k(\cdot)$

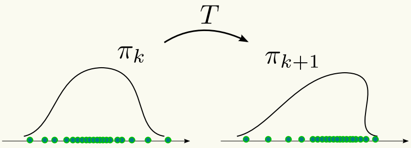
Approximate: a transportation map between π_k and $\pi_{k+1}(x) = \dfrac{\pi_k(x) l_k(x)}{\gamma}$

Questions:

- There are infinitely many transport maps. Which one to approximate?
- How to approximate it in terms of particles?
- What is the approximation error?

Y. Cheng and S. Reich, A McKean optimal transportation perspective on Feynman-Kac formulae with application to data assimilation, 2013

Transportation viewpoint:



Given: $\{X_k^1, \ldots, X_k^N\} \sim \pi_k$ and $l_k(\cdot)$

Approximate: a transportation map between π_k and $\pi_{k+1}(x) = \dfrac{\pi_k(x) l_k(x)}{\gamma}$

Questions:

- There are infinitely many transport maps. Which one to approximate?
- How to approximate it in terms of particles?
- What is the approximation error?

Y. Cheng and S. Reich, A McKean optimal transportation perspective on Feynman-Kac formulae with application to data assimilation, 2013

SIR particle filter: (Gordon et al., 1993, Doucet, 2009)

- Sample from the independent coupling:

(Variations of) SIR particle filter: (Del Moral 2004, Bain & Crisan 2009, Van Leeuwen 2015, Nakamura & Potthast 2015)

- Sample from a coupling that is optimal with respect to some optimality criteria

Ensemble transform particle filters:

- Sample from optimal transport coupling (Cheng & Reich, 2013)
- Sample from Schrodinger bridge coupling (Reich, 2018)

SIR particle filter: (Gordon et al., 1993, Doucet, 2009)

- Sample from the independent coupling:

(Variations of) SIR particle filter: (Del Moral 2004, Bain & Crisan 2009, Van Leeuwen 2015, Nakamura & Potthast 2015)

- Sample from a coupling that is optimal with respect to some optimality criteria

Ensemble transform particle filters:

- Sample from optimal transport coupling (Cheng & Reich, 2013)
- Sample from Schrodinger bridge coupling (Reich, 2018)

SIR particle filter: (Gordon et al., 1993, Doucet, 2009)

- Sample from the independent coupling:

(Variations of) SIR particle filter: (Del Moral 2004, Bain & Crisan 2009, Van Leeuwen 2015, Nakamura & Potthast 2015)

- Sample from a coupling that is optimal with respect to some optimality criteria

Ensemble transform particle filters:

- Sample from optimal transport coupling (Cheng & Reich, 2013)
- Sample from Schrodinger bridge coupling (Reich, 2018)

State process: $\mathrm{d}X_t = a(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \qquad X_0 \sim \pi_0$

Observation process: $\mathrm{d}Z_t = h(X_t)\mathrm{d}t + \mathrm{d}W_t$

Filtering objective: Compute the posterior distribution $\pi_t(\cdot) := \mathsf{P}(X_t \in \cdot | Z_{[0,t]})$

Solution:

$$\mathrm{d}\pi_t(x) = \underbrace{\mathcal{L}\pi_t(x)\mathrm{d}t}_{\text{Fokker-Planck}} + \underbrace{\pi_t(x)(h(x) - \hat{h}_t)\mathrm{d}I_t}_{\text{Bayes rule}}, \quad \text{(Kushner-Stratonovic spde)}$$

- Innovation error process $\mathrm{d}I_t = \mathrm{d}Z_t - \hat{h}_t\mathrm{d}t$
- $\hat{h}_t = \displaystyle\int h(x)\pi_t(x)\mathrm{d}x$

J. Xiong, An introduction to stochastic filtering theory, 2008

Background: Filtering problem
Continuous-time setting

$$\textbf{State process:} \quad \mathrm{d}X_t = a(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \qquad X_0 \sim \pi_0$$

$$\textbf{Observation process:} \quad \mathrm{d}Z_t = h(X_t)\mathrm{d}t + \mathrm{d}W_t$$

Filtering objective: Compute the posterior distribution $\pi_t(\cdot) := \mathsf{P}(X_t \in \cdot | Z_{[0,t]})$

Solution:

$$\mathrm{d}\pi_t(x) = \underbrace{\mathcal{L}\pi_t(x)\mathrm{d}t}_{\text{Fokker-Planck}} + \underbrace{\pi_t(x)(h(x) - \hat{h}_t)\mathrm{d}I_t}_{\text{Bayes rule}}, \quad \text{(Kushner-Stratonovic spde)}$$

- Innovation error process $\mathrm{d}I_t = \mathrm{d}Z_t - \hat{h}_t\mathrm{d}t$
- $\hat{h}_t = \int h(x)\pi_t(x)\mathrm{d}x$

J. Xiong, An introduction to stochastic filtering theory, 2008

State process: $\mathrm{d}X_t = a(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \qquad X_0 \sim \pi_0$

Observation process: $\mathrm{d}Z_t = h(X_t)\mathrm{d}t + \mathrm{d}W_t$

Filtering objective: Compute the posterior distribution $\pi_t(\cdot) := \mathsf{P}(X_t \in \cdot | Z_{[0,t]})$

Solution:

$$\mathrm{d}\pi_t(x) = \underbrace{\mathcal{L}\pi_t(x)\mathrm{d}t}_{\text{Fokker-Planck}} + \underbrace{\pi_t(x)(h(x) - \hat{h}_t)\mathrm{d}I_t}_{\text{Bayes rule}}, \quad \text{(Kushner-Stratonovic spde)}$$

- Innovation error process $\mathrm{d}I_t = \mathrm{d}Z_t - \hat{h}_t\mathrm{d}t$
- $\hat{h}_t = \displaystyle\int h(x)\pi_t(x)\mathrm{d}x$
- In general, no finite-dimensional solution $\Rightarrow$ approximations

J. Xiong, An introduction to stochastic filtering theory, 2008

$$\textbf{State process:} \quad \mathrm{d}X_t = a(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \qquad X_0 \sim \pi_0$$

$$\textbf{Observation process:} \quad \mathrm{d}Z_t = h(X_t)\mathrm{d}t + \mathrm{d}W_t$$

Filtering objective: Compute the posterior distribution $\pi_t(\cdot) := \mathsf{P}(X_t \in \cdot | Z_{[0,t]})$

Solution:

$$\mathrm{d}\pi_t(x) = \underbrace{\mathcal{L}\pi_t(x)\mathrm{d}t}_{\text{Fokker-Planck}} + \underbrace{\pi_t(x)(h(x) - \hat{h}_t)\mathrm{d}I_t}_{\text{Bayes rule}}, \quad \text{(Kushner-Stratonovic spde)}$$

- Innovation error process $\mathrm{d}I_t = \mathrm{d}Z_t - \hat{h}_t\mathrm{d}t$
- $\hat{h}_t = \int h(x)\pi_t(x)\mathrm{d}x$
- In general, no finite-dimensional solution $\Rightarrow$ approximations

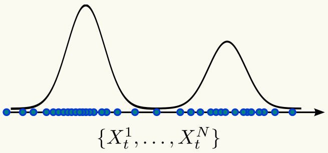
J. Xiong, An introduction to stochastic filtering theory, 2008

Idea:

- Approximate π_k with empirical distribution of particles $\{X_t^1, \dots, X_t^N\}$

$$\pi_t \approx \frac{1}{N} \sum_{i=1}^{N} \delta_{X_t^i}$$

- Continuously update particles such that $\{X_t^i\}_{i=1}^N \sim \pi_t$



Update formula:

$$\mathrm{d}X_t^i = \text{(dynamic model)} + \underbrace{\nabla\phi_t(X_t^i) \circ \mathrm{d}I_t^i}_{\text{correction}}, \quad X_0^i \sim \rho_0$$

- where ϕ_t solves the (weighted) Poisson eq.

$$\frac{1}{\rho_t(x)} \nabla \cdot (\rho_t(x) \nabla\phi_t(x)) = h(x) - \hat{h}_t$$

- ρ_t is probability density for X_t^i (in mean-field limit)

Consistency of FPF:

$$\text{if} \quad \rho_0 = \pi_0, \quad \text{then} \quad \rho_t = \pi_t$$

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

Update formula:

$$\mathrm{d}X_t^i = \text{(dynamic model)} + \underbrace{\nabla\phi_t(X_t^i) \circ \mathrm{d}I_t^i}_{\text{correction}}, \quad X_0^i \sim \rho_0$$

- where ϕ_t solves the (weighted) Poisson eq.

$$\frac{1}{\rho_t(x)}\nabla \cdot (\rho_t(x)\nabla\phi_t(x)) = h(x) - \hat{h}_t$$

- ρ_t is probability density for X_t^i (in mean-field limit)

Consistency of FPF:

$$\text{if} \quad \rho_0 = \pi_0, \quad \text{then} \quad \rho_t = \pi_t$$

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

Update formula:

$$\mathrm{d}X_t^i = \text{(dynamic model)} + \underbrace{\nabla\phi_t(X_t^i) \circ \mathrm{d}I_t^i}_{\text{correction}}, \quad X_0^i \sim \rho_0$$

- where ϕ_t solves the (weighted) Poisson eq.

$$\frac{1}{\rho_t(x)}\nabla\cdot(\rho_t(x)\nabla\phi_t(x)) = h(x) - \hat{h}_t$$

- ρ_t is probability density for X_t^i (in mean-field limit)

Consistency of FPF:

$$\text{if} \quad \rho_0 = \pi_0, \quad \text{then} \quad \rho_t = \pi_t$$

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

1. Approximation: Numerical approximation of the gain

$$\textbf{Given:} \quad \{X_t^1, \ldots, X_t^N\}_{i=1}^N \sim \rho_t$$
$$\textbf{Approximate:} \quad \{\nabla\phi_t(X_t^1), \ldots, \nabla\phi_t(X_t^N)\}_{i=1}^N$$

2. Design: How to construct a control law that is exact?

3. Error analysis: Analysis of the error between empirical distribution $\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$ and exact filter π_t

1. Approximation: Numerical approximation of the gain

$$\textbf{Given:} \quad \{X_t^1, \ldots, X_t^N\}_{i=1}^N \sim \rho_t$$
$$\textbf{Approximate:} \quad \{\nabla\phi_t(X_t^1), \ldots, \nabla\phi_t(X_t^N)\}_{i=1}^N$$

2. Design: How to construct a control law that is exact?

3. Error analysis: Analysis of the error between empirical distribution $\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$ and exact filter π_t

1. Approximation: Numerical approximation of the gain

$$\textbf{Given:} \quad \{X_t^1, \ldots, X_t^N\}_{i=1}^N \sim \rho_t$$
$$\textbf{Approximate:} \quad \{\nabla\phi_t(X_t^1), \ldots, \nabla\phi_t(X_t^N)\}_{i=1}^N$$

2. Design: How to construct a control law that is exact?

3. Error analysis: Analysis of the error between empirical distribution $\dfrac{1}{N}\displaystyle\sum_{i=1}^N \delta_{X_t^i}$ and exact filter π_t

1. Approximation: Numerical approximation of the gain $\leftarrow$ this talk

   Given: $\{X_t^1, \ldots, X_t^N\}_{i=1}^N \sim \rho_t$

   Approximate: $\{\nabla\phi_t(X_t^1), \ldots, \nabla\phi_t(X_t^N)\}_{i=1}^N$

2. Design: How to construct a control law that is exact?

3. Error analysis: Analysis of the error between empirical distribution $\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}$ and exact filter π_t

Outline

Part I: Background

part II: Gain function approximation in FPF

- Constant gain approximation
- Galerkin approximation
- Diffusion-map approximation
- Error analysis

FPF update formula:

$$\mathrm{d}X_t^i = (\text{dynamic model}) + \nabla\phi_t(X_t^i) \circ \mathrm{d}I_t^i$$

where ϕ_t solves the Poisson eq.

Poisson equation:

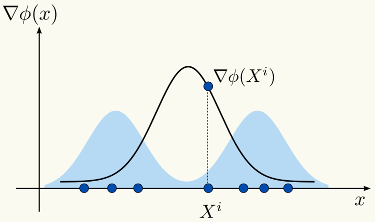
$$-\Delta_\rho \phi(x) = h(x) - \hat{h}$$

where $\Delta_\rho \phi := \frac{1}{\rho}\nabla \cdot (\rho \nabla \phi)$

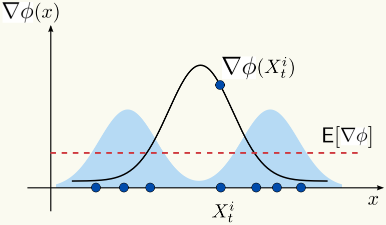
Computational problem:

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

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



$$\mathsf{K}_{\mathsf{const}} := \arg\min_{\mathsf{K}\in\mathbb{R}^d} \int |\nabla\phi(x) - \mathsf{K}|^2 \rho(x)\mathrm{d}x$$
$$= \int \nabla\phi(x)\rho(x)\mathrm{d}x$$



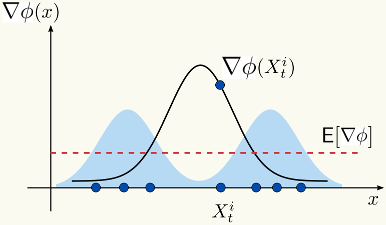
A closed-form formula:

$$\mathsf{K}_{\mathsf{const}} = \int (h(x) - \hat{h})x\rho(x)\mathrm{d}x \approx \frac{1}{N}\sum_{i=1} (h(X^i) - \hat{h}^{(N)})X^i$$

- FPF $\xrightarrow{\text{const. gain approx.}}$ Ensemble Kalman filter (EnKF)
- Can we improve this approximation?

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model . . . 1994.
A. Taghvaei, J de Wiljes, P. G. Mehta, and S. Reich. Kalman filter and its modern extensions... ASME, 2017

$$\mathsf{K}_{\text{const}} := \arg\min_{\mathsf{K}\in\mathbb{R}^d} \int |\nabla\phi(x) - \mathsf{K}|^2 \rho(x)\mathrm{d}x$$
$$= \int \nabla\phi(x)\rho(x)\mathrm{d}x$$



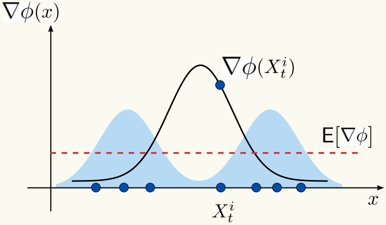
A closed-form formula:

$$\mathsf{K}_{\text{const}} = \int (h(x) - \hat{h})x\rho(x)\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}(h(X^i) - \hat{h}^{(N)})X^i$$

- FPF $\xrightarrow{\text{const. gain approx.}}$ Ensemble Kalman filter (EnKF)
- Can we improve this approximation?

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model . . . 1994.
A. Taghvaei, J de Wiljes, P. G. Mehta, and S. Reich. Kalman filter and its modern extensions... ASME, 2017

$$\mathsf{K}_{\text{const}} := \arg\min_{\mathsf{K}\in\mathbb{R}^d} \int |\nabla\phi(x) - \mathsf{K}|^2 \rho(x)\mathrm{d}x$$
$$= \int \nabla\phi(x)\rho(x)\mathrm{d}x$$



A closed-form formula:

$$\mathsf{K}_{\text{const}} = \int (h(x) - \hat{h}) x \rho(x)\mathrm{d}x \approx \frac{1}{N}\sum_{i=1} (h(X^i) - \hat{h}^{(N)}) X^i$$

- FPF $\xrightarrow{\text{const. gain approx.}}$ Ensemble Kalman filter (EnKF)
- Can we improve this approximation?

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model ...1994.
A. Taghvaei, J de Wiljes, P. G. Mehta, and S. Reich. Kalman filter and its modern extensions... ASME, 2017

Assumptions:

(A1) (Density has Gaussian tail) $\rho = e^{-V}$ where $V(x) = \dfrac{1}{2}(x-\bar{x})^\top \Sigma^{-1}(x-\bar{x}) + W(x)$ with $W \in C_b^\infty(\mathbb{R}^d)$

(A2) (h has linear growth) $h(x) = c^\top x + w(x)$ where $w \in C_b^\infty(\mathbb{R}^d)$

Function spaces: $L^2(\rho)$, $H^1(\rho)$, $L_0^2(\rho) := \left\{ f \in L^2(\rho); \int f\rho \mathrm{d}x = 0 \right\}$, $H_0^1(\rho)$

Poincaré inequality: Under Assumption (A1), $\exists \lambda > 0$ such that

$$\int f^2 \rho \mathrm{d}x \leq \frac{1}{\lambda} \int |\nabla f|^2 \rho \mathrm{d}x, \quad \forall f \in H_0^1(\rho)$$

Existence and uniqueness: Under Assumption (A1)-(A2) there exists a unique (weak) solution to the Poisson eq.

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson's equation in nonlinear filtering, *SIAM*, 2015

Assumptions:

(A1) (Density has Gaussian tail) $\rho = e^{-V}$ where $V(x) = \frac{1}{2}(x-\bar{x})^\top \Sigma^{-1}(x-\bar{x}) + W(x)$ with $W \in C_b^\infty(\mathbb{R}^d)$

(A2) (h has linear growth) $h(x) = c^\top x + w(x)$ where $w \in C_b^\infty(\mathbb{R}^d)$

Function spaces: $L^2(\rho)$, $H^1(\rho)$, $L^2_0(\rho) := \left\{ f \in L^2(\rho); \int f \rho \mathrm{d}x = 0 \right\}$, $H^1_0(\rho)$

Poincaré inequality: Under Assumption (A1), $\exists \lambda > 0$ such that

$$\int f^2 \rho \mathrm{d}x \leq \frac{1}{\lambda} \int |\nabla f|^2 \rho \mathrm{d}x, \quad \forall f \in H^1_0(\rho)$$

Existence and uniqueness: Under Assumption (A1)-(A2) there exists a unique (weak) solution to the Poisson eq.

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson's equation in nonlinear filtering, *SIAM*, 2015

Assumptions:

(A1) (Density has Gaussian tail) $\rho = e^{-V}$ where $V(x) = \frac{1}{2}(x-\bar{x})^\top \Sigma^{-1}(x-\bar{x}) + W(x)$ with $W \in C_b^\infty(\mathbb{R}^d)$

(A2) (h has linear growth) $h(x) = c^\top x + w(x)$ where $w \in C_b^\infty(\mathbb{R}^d)$

Function spaces: $L^2(\rho)$, $H^1(\rho)$, $L_0^2(\rho) := \left\{ f \in L^2(\rho); \int f \rho \mathrm{d}x = 0 \right\}$, $H_0^1(\rho)$

Poincaré inequality: Under Assumption (A1), $\exists \lambda > 0$ such that

$$\int f^2 \rho \mathrm{d}x \leq \frac{1}{\lambda} \int |\nabla f|^2 \rho \mathrm{d}x, \quad \forall f \in H_0^1(\rho)$$

Existence and uniqueness: Under Assumption (A1)-(A2) there exists a unique (weak) solution to the Poisson eq.

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson's equation in nonlinear filtering, *SIAM*, 2015

Assumptions:

(A1) (Density has Gaussian tail) $\rho = e^{-V}$ where $V(x) = \frac{1}{2}(x-\bar{x})^\top \Sigma^{-1}(x-\bar{x}) + W(x)$ with $W \in C_b^\infty(\mathbb{R}^d)$

(A2) (h has linear growth) $h(x) = c^\top x + w(x)$ where $w \in C_b^\infty(\mathbb{R}^d)$

Function spaces: $L^2(\rho)$, $H^1(\rho)$, $L_0^2(\rho) := \left\{ f \in L^2(\rho); \int f\rho \mathrm{d}x = 0 \right\}$, $H_0^1(\rho)$

Poincaré inequality: Under Assumption (A1), $\exists \lambda > 0$ such that

$$\int f^2 \rho \mathrm{d}x \leq \frac{1}{\lambda} \int |\nabla f|^2 \rho \mathrm{d}x, \quad \forall f \in H_0^1(\rho)$$

Existence and uniqueness: Under Assumption (A1)-(A2) there exists a unique (weak) solution to the Poisson eq.

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson's equation in nonlinear filtering, *SIAM*, 2015

(I) Weak formulation: (Galerkin algorithm)

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

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

(II) Semigroup formulation: (Diffusion-map algorithm)

$$\phi = P_t \phi + \int_0^t P_s (h - \hat{h}) \mathrm{d}s$$

where $P_t := e^{t\Delta_\rho}$ is the semigroup

(III) Variational formulation: (Neural net. approx.)

$$\min_{\phi \in H_0^1(\rho)} \int \left[\frac{1}{2} |\nabla \phi(x)|^2 - \phi(x)(h(x) - \hat{h}) \right] \rho(x) \mathrm{d}x$$

(I) Weak formulation: (Galerkin algorithm)

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

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

(II) Semigroup formulation: (Diffusion-map algorithm)

$$\phi = P_t\phi + \int_0^t P_s(h - \hat{h})\mathrm{d}s$$

where $P_t := e^{t\Delta_\rho}$ is the semigroup

(III) Variational formulation: (Neural net. approx.)

$$\min_{\phi \in H_0^1(\rho)} \int \left[\frac{1}{2}|\nabla\phi(x)|^2 - \phi(x)(h(x) - \hat{h}) \right] \rho(x)\mathrm{d}x$$

(I) Weak formulation: (Galerkin algorithm)

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

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

(II) Semigroup formulation: (Diffusion-map algorithm)

$$\phi = P_t\phi + \int_0^t P_s(h - \hat{h})\mathrm{d}s$$

where $P_t := e^{t\Delta_\rho}$ is the semigroup

(III) Variational formulation: (Neural net. approx.)

$$\min_{\phi \in H_0^1(\rho)} \int \left[\frac{1}{2}|\nabla\phi(x)|^2 - \phi(x)(h(x) - \hat{h})\right]\rho(x)\mathrm{d}x$$

(I) Weak formulation: (Galerkin algorithm)

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

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

(II) Semigroup formulation: (Diffusion-map algorithm)

$$\phi = P_t\phi + \int_0^t P_s(h - \hat{h})\mathrm{d}s$$

where $P_t := e^{t\Delta_\rho}$ is the semigroup

(III) Variational formulation: (Neural net. approx.)

$$\min_{\phi \in H_0^1(\rho)} \int \left[\frac{1}{2}|\nabla\phi(x)|^2 - \phi(x)(h(x) - \hat{h}) \right] \rho(x)\mathrm{d}x$$

(I) Weak formulation: (Galerkin algorithm)

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

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

(II) Semigroup formulation: (Diffusion-map algorithm)

$$\phi = P_t\phi + \int_0^t P_s(h - \hat{h})\mathrm{d}s$$

where $P_t := e^{t\Delta_\rho}$ is the semigroup

(III) Variational formulation: (Neural net. approx.)

$$\min_{\phi \in H_0^1(\rho)} \int \left[\frac{1}{2} |\nabla\phi(x)|^2 - \phi(x)(h(x) - \hat{h}) \right] \rho(x) \mathrm{d}x$$

(I) Weak formulation: (Galerkin algorithm)

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

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

(II) Semigroup formulation: (Diffusion-map algorithm)

$$\phi = P_t\phi + \int_0^t P_s(h - \hat{h})\mathrm{d}s$$

where $P_t := e^{t\Delta_\rho}$ is the semigroup

(III) Variational formulation: (Neural net. approx.)

$$\min_{\phi \in H_0^1(\rho)} \int \left[\frac{1}{2} |\nabla\phi(x)|^2 - \phi(x)(h(x) - \hat{h}) \right] \rho(x)\mathrm{d}x$$

(I) Weak formulation: (Galerkin algorithm)

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

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

(II) Semigroup formulation: (Diffusion-map algorithm)

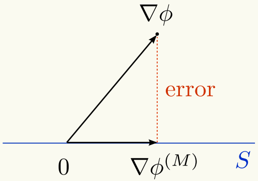
$$\phi = P_t\phi + \int_0^t P_s(h - \hat{h})\mathrm{d}s$$

where $P_t := e^{t\Delta_\rho}$ is the semigroup

(III) Variational formulation: (Neural net. approx.)

$$\min_{\phi \in H_0^1(\rho)} \int \left[\frac{1}{2} |\nabla\phi(x)|^2 - \phi(x)(h(x) - \hat{h}) \right] \rho(x)\mathrm{d}x$$

Idea: Projection into a finite-dim subspace
$S = \text{span}\{\psi_1, \ldots, \psi_M\} \subset H^1(\mathbb{R}^d, \rho)$



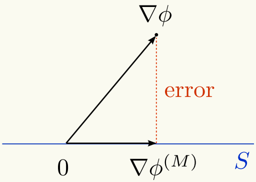
$\boxed{\text{Choice of basis function is difficult}}$

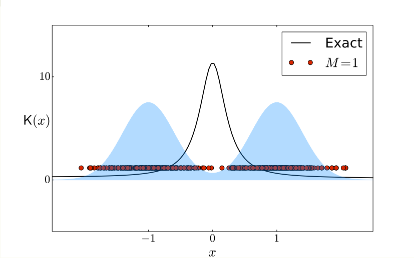
Ideal case: S is first M eigenfunctions of Δ_ρ

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

(I) Galerkin approximation

Idea: Projection into a finite-dim subspace $S = \text{span}\{\psi_1, \ldots, \psi_M\} \subset H^1(\mathbb{R}^d, \rho)$





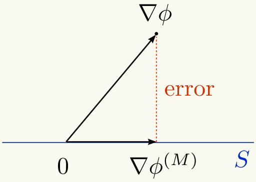
$$S = \text{span}\{1, x, \ldots, x^M\}$$

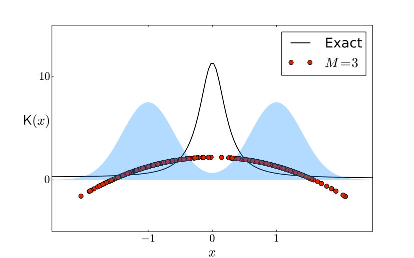
Choice of basis function is difficult

Ideal case: S is first M eigenfunctions of Δ_ρ

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

Idea: Projection into a finite-dim subspace $S = \text{span}\{\psi_1, \dots, \psi_M\} \subset H^1(\mathbb{R}^d, \rho)$





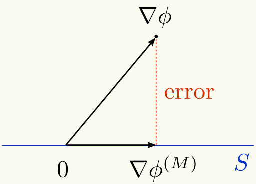
$S = \text{span}\{1, x, \dots, x^M\}$

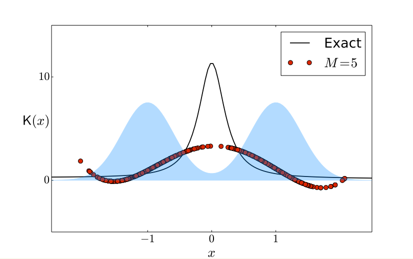
Choice of basis function is difficult

Ideal case: S is first M eigenfunctions of Δ_ρ

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

Idea: Projection into a finite-dim subspace $S = \text{span}\{\psi_1, \ldots, \psi_M\} \subset H^1(\mathbb{R}^d, \rho)$





$S = \text{span}\{1, x, \ldots, x^M\}$

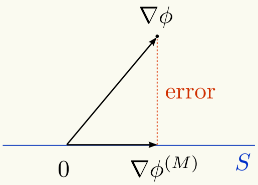
$\boxed{\text{Choice of basis function is difficult}}$

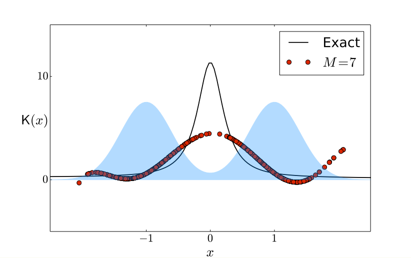
Ideal case: S is first M eigenfunctions of Δ_ρ

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

(I) Galerkin approximation

Idea: Projection into a finite-dim subspace $S = \text{span}\{\psi_1, \ldots, \psi_M\} \subset H^1(\mathbb{R}^d, \rho)$





$S = \text{span}\{1, x, \ldots, x^M\}$

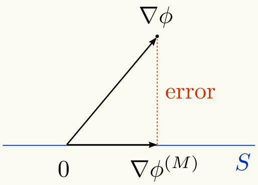
Choice of basis function is difficult

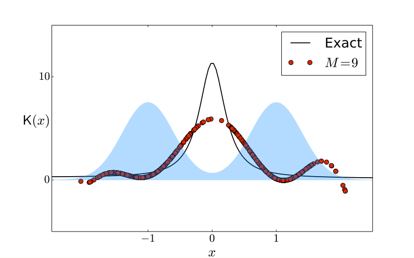
Ideal case: S is first M eigenfunctions of Δ_ρ

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

(I) Galerkin approximation

Idea: Projection into a finite-dim subspace $S = \text{span}\{\psi_1, \dots, \psi_M\} \subset H^1(\mathbb{R}^d, \rho)$





$$S = \text{span}\{1, x, \dots, x^M\}$$

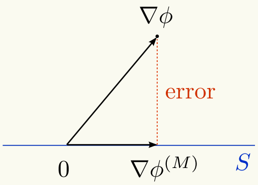
$\boxed{\text{Choice of basis function is difficult}}$

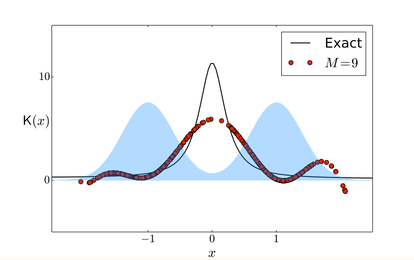
Ideal case: S is first M eigenfunctions of Δ_ρ

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

(I) Galerkin approximation

Idea: Projection into a finite-dim subspace
$S = \mathsf{span}\{\psi_1, \ldots, \psi_M\} \subset H^1(\mathbb{R}^d, \rho)$





$$S = \mathsf{span}\{1, x, \ldots, x^M\}$$

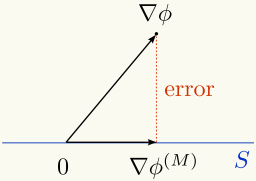
$\boxed{\text{Choice of basis function is difficult}}$

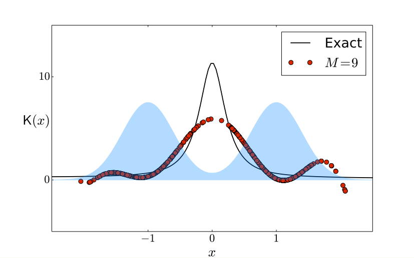
Ideal case: S is first M eigenfunctions of Δ_ρ

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

(I) Galerkin approximation

Idea: Projection into a finite-dim subspace
$S = \text{span}\{\psi_1, \ldots, \psi_M\} \subset H^1(\mathbb{R}^d, \rho)$





$$S = \text{span}\{1, x, \ldots, x^M\}$$

$$\boxed{\text{Choice of basis function is difficult}}$$

Ideal case: S is first M eigenfunctions of Δ_ρ

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

- Poisson equation

$$-\Delta_\rho \phi = h - \hat{h}$$

- Step 1: Semigroup formulation:

$$\phi = P_t \phi + \int_0^t P_s (h - \hat{h}) \mathrm{d}s$$

- Step 2: Diffusion map approximation:

$$T_\epsilon f(x) := \frac{1}{n_\epsilon(x)} \int g_\epsilon(x,y) \frac{f(y)\rho(y)}{\sqrt{(g_\epsilon * \rho)(y)}} \mathrm{d}y \overset{\epsilon \downarrow 0}{\approx} P_\epsilon f(x)$$

- Step 3: Empirical approximation

$$T_\epsilon^{(N)} f(x) := \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N g_\epsilon(x, X^i) \frac{f(X^i)}{\sqrt{\sum_{j=1}^N g_\epsilon(X^i, X^j)}} \overset{N \uparrow \infty}{\approx} T_\epsilon f(x)$$

R. Coifman and S. Lafon, Diffusion maps, Applied and computational harmonic analysis, 21 (2006)
M. Hein, J. Audibert, and U. Luxburg, Graph Laplacians and their convergence on random neighborhood graphs, JMLR, 8 (2007)

- Poisson equation

$$-\Delta_\rho \phi = h - \hat{h}$$

- Step 1: Semigroup formulation:

$$\phi = P_t \phi + \int_0^t P_s (h - \hat{h}) \mathrm{d}s$$

- Step 2: Diffusion map approximation:

$$T_\epsilon f(x) := \frac{1}{n_\epsilon(x)} \int g_\epsilon(x,y) \frac{f(y)\rho(y)}{\sqrt{(g_\epsilon * \rho)(y)}} \mathrm{d}y \overset{\epsilon \downarrow 0}{\approx} P_\epsilon f(x)$$

- Step 3: Empirical approximation

$$T_\epsilon^{(N)} f(x) := \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N g_\epsilon(x, X^i) \frac{f(X^i)}{\sqrt{\sum_{j=1}^N g_\epsilon(X^i, X^j)}} \overset{N \uparrow \infty}{\approx} T_\epsilon f(x)$$

R. Coifman and S. Lafon, Diffusion maps, Applied and computational harmonic analysis, 21 (2006)
M. Hein, J. Audibert, and U. Luxburg, Graph Laplacians and their convergence on random neighborhood graphs, JMLR, 8 (2007)

- Poisson equation

$$-\Delta_\rho \phi = h - \hat{h}$$

- Step 1: Semigroup formulation:

$$\phi = P_t \phi + \int_0^t P_s (h - \hat{h}) \mathrm{d}s$$

Proof: $\dfrac{\partial P_t \phi}{\partial t} = P_t \Delta_\rho \phi = -P_t (h - \hat{h})$

- Step 2: Diffusion map approximation:

$$T_\epsilon f(x) := \frac{1}{n_\epsilon(x)} \int g_\epsilon(x,y) \frac{f(y)\rho(y)}{\sqrt{(g_\epsilon * \rho)(y)}} \mathrm{d}y \overset{\epsilon\downarrow 0}{\approx} P_\epsilon f(x)$$

- Step 3: Empirical approximation

$$T_\epsilon^{(N)} f(x) := \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N g_\epsilon(x, X^i) \frac{f(X^i)}{\sqrt{\sum_{j=1}^N g_\epsilon(X^i, X^j)}} \overset{N\uparrow\infty}{\approx} T_\epsilon f(x)$$

R. Coifman and S. Lafon, Diffusion maps, Applied and computational harmonic analysis, 21 (2006)
M. Hein, J. Audibert, and U. Luxburg, Graph Laplacians and their convergence on random neighborhood graphs, JMLR, 8 (2007)

- Poisson equation

$$-\Delta_\rho \phi = h - \hat{h}$$

- Step 1: Semigroup formulation:

$$\phi = P_t \phi + \int_0^t P_s (h - \hat{h}) \mathrm{d}s$$

- Step 2: Diffusion map approximation:

$$T_\epsilon f(x) := \frac{1}{n_\epsilon(x)} \int g_\epsilon(x,y) \frac{f(y)\rho(y)}{\sqrt{(g_\epsilon * \rho)(y)}} \mathrm{d}y \overset{\epsilon \downarrow 0}{\approx} P_\epsilon f(x)$$

- Step 3: Empirical approximation

$$T_\epsilon^{(N)} f(x) := \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N g_\epsilon(x, X^i) \frac{f(X^i)}{\sqrt{\sum_{j=1}^N g_\epsilon(X^i, X^j)}} \overset{N \uparrow \infty}{\approx} T_\epsilon f(x)$$

R. Coifman and S. Lafon, Diffusion maps, Applied and computational harmonic analysis, 21 (2006)
M. Hein, J. Audibert, and U. Luxburg, Graph Laplacians and their convergence on random neighborhood graphs, JMLR, 8 (2007)

- Poisson equation

$$-\Delta_\rho \phi = h - \hat{h}$$

- Step 1: Semigroup formulation:

$$\phi = P_t \phi + \int_0^t P_s (h - \hat{h}) \mathrm{d}s$$

- Step 2: Diffusion map approximation:

$$T_\epsilon f(x) := \frac{1}{n_\epsilon(x)} \int g_\epsilon(x,y) \frac{f(y)\rho(y)}{\sqrt{(g_\epsilon * \rho)(y)}} \mathrm{d}y \overset{\epsilon \downarrow 0}{\approx} P_\epsilon f(x)$$

- Step 3: Empirical approximation

$$T_\epsilon^{(N)} f(x) := \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^{N} g_\epsilon(x, X^i) \frac{f(X^i)}{\sqrt{\sum_{j=1}^{N} g_\epsilon(X^i, X^j)}} \overset{N \uparrow \infty}{\approx} T_\epsilon f(x)$$

R. Coifman and S. Lafon, Diffusion maps, Applied and computational harmonic analysis, 21 (2006)
M. Hein, J. Audibert, and U. Luxburg, Graph Laplacians and their convergence on random neighborhood graphs, JMLR, 8 (2007)

Poisson eq.: $-\Delta_\rho \phi = h - \hat{h}, \quad \phi \in H^1_0(\rho)$

(approximation steps)

Finite-dim eq.: $\Phi = \mathsf{T}\Phi + \epsilon(\mathsf{h} - \pi(h)), \quad \Phi \in \mathbb{R}^N_0$

- T is a $N \times N$ Markov matrix
- $k_\epsilon^{(N)}(x,y)$ is the diffusion map kernel (Coifman & Lafon, 2006)
- The solution $\Phi \approx (\phi(X^1), \ldots, \phi(X^N))$

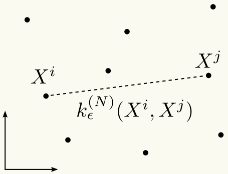
Gain function approximation: Diffusion map-based algorithm
Overview

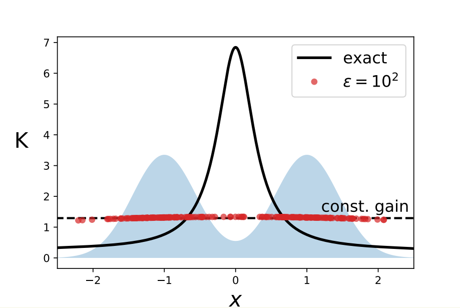
$$\textbf{Poisson eq.:} \quad -\Delta_\rho \phi = h - \hat{h}, \quad \phi \in H_0^1(\rho)$$

(approximation steps)

$$\textbf{Finite-dim eq.:} \quad \Phi = \mathsf{T}\Phi + \epsilon(\mathsf{h} - \pi(h)), \quad \Phi \in \mathbb{R}_0^N$$

- T is a $N \times N$ Markov matrix
- $k_\epsilon^{(N)}(x, y)$ is the diffusion map kernel (Coifman & Lafon, 2006)
- The solution $\Phi \approx (\phi(X^1), \ldots, \phi(X^N))$

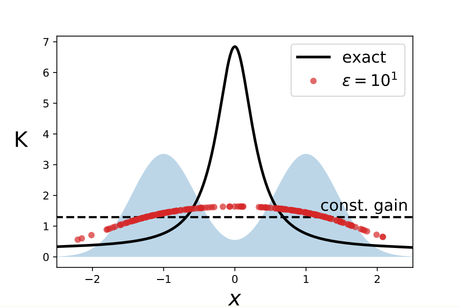




Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

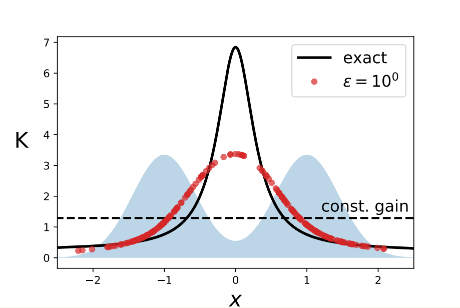
Objective: Error analysis of bias and variance



Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

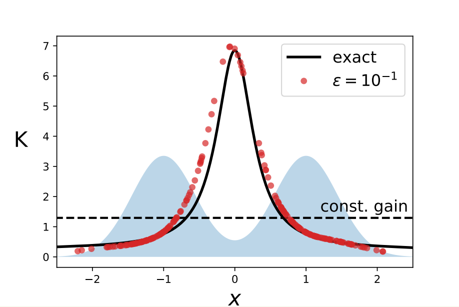
Objective: Error analysis of bias and variance



Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

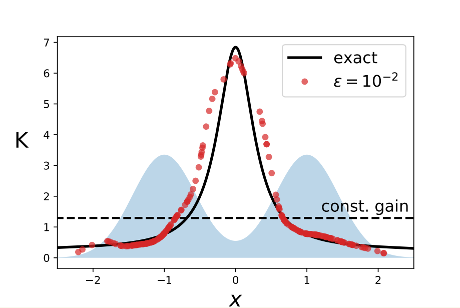
Objective: Error analysis of bias and variance



Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

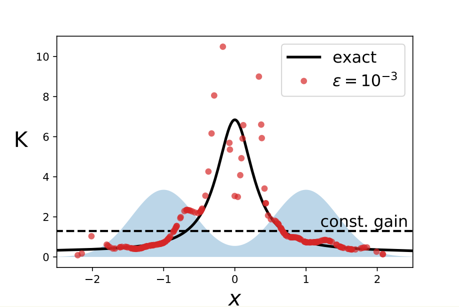
Objective: Error analysis of bias and variance



Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

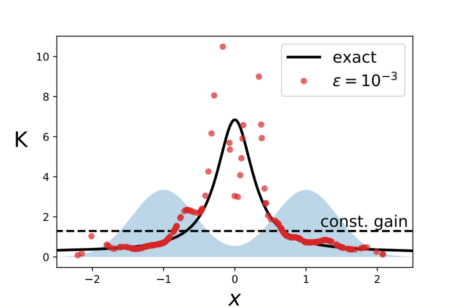
Objective: Error analysis of bias and variance

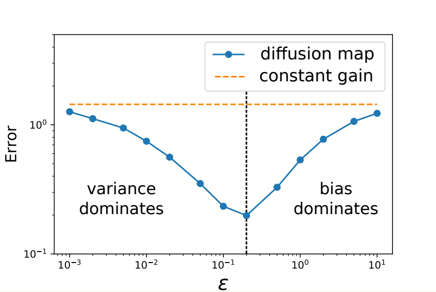


Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

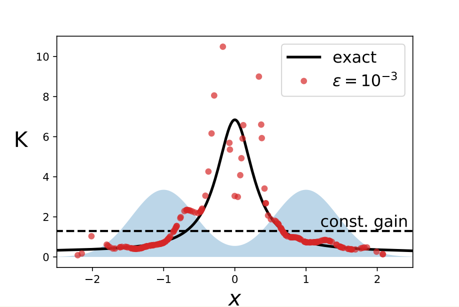


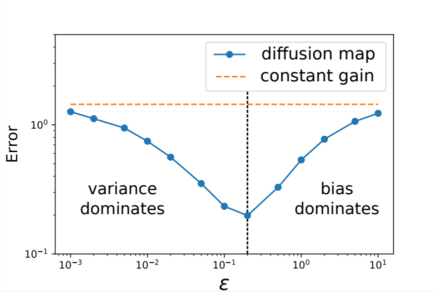


Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

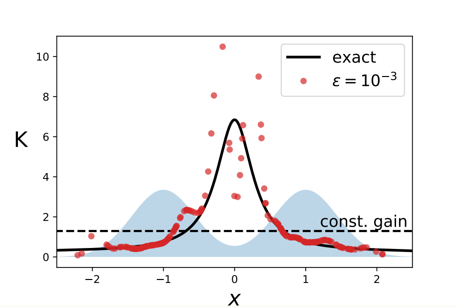


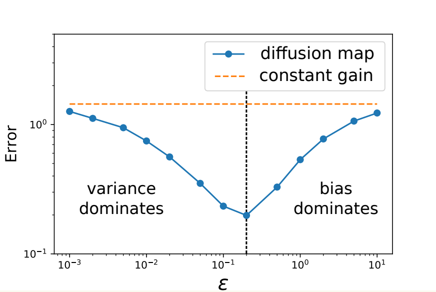


Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance





Properties:

- No basis function selection
- Computationally scales well with dimension $O(N^2 d)$.
- Converges to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

(1) Semigroup formulation: $$\phi = P_\epsilon \phi + \int_0^\epsilon P_s(h - \hat{h})\mathrm{d}s$$

(2) Diffusion map approx: $$\phi_\epsilon = T_\epsilon \phi_\epsilon + \epsilon(h - \hat{h}_\epsilon)$$

(3) Empirical approx: $$\phi_\epsilon^{(N)} = T_\epsilon^{(N)} \phi_\epsilon^{(N)} + \epsilon(h - \hat{h}_\epsilon^{(N)})$$

Error analysis:

- Bias: Study convergence $\phi_\epsilon \to \phi$ as $\epsilon \to 0$
- Variance: Study convergence $\phi_\epsilon^{(N)} \to \phi_\epsilon$ as $N \to \infty$

(1) Semigroup formulation: $$\phi = P_\epsilon \phi + \int_0^\epsilon P_s(h - \hat{h})\mathrm{d}s$$

(2) Diffusion map approx: $$\phi_\epsilon = T_\epsilon \phi_\epsilon + \epsilon(h - \hat{h}_\epsilon)$$

(3) Empirical approx: $$\phi_\epsilon^{(N)} = T_\epsilon^{(N)} \phi_\epsilon^{(N)} + \epsilon(h - \hat{h}_\epsilon^{(N)})$$

Error analysis:

- Bias: Study convergence $\phi_\epsilon \to \phi$ as $\epsilon \to 0$
- Variance: Study convergence $\phi_\epsilon^{(N)} \to \phi_\epsilon$ as $N \to \infty$

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon\phi_\epsilon + \epsilon(h-\hat{h}_\epsilon)$$

Proposition (TMM'19)

1. For all functions $f, \nabla f \in L^4(\rho)$

$$\|T_{\frac{t}{n}}^n f - P_t f\|_{L^2(\rho)} \leq \frac{\text{(const.)}}{n}(\|f\|_{L^4(\rho)} + \|\nabla f\|_{L^4(\rho)})$$

2. $T_{\frac{t}{n}}^n$ admits a uniform spectral gap on $L_0^2(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \leq \text{(const.)}\|h\|_{L^2(\rho_\epsilon)}$$

3. In the asymptotic limit as $\epsilon \to 0$

$$\|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq O(\epsilon)$$

1. Proof of 1: Feynman-Kac representation of semigroup
2. Proof of 2: Foster-Lyapunov condition from stochastic stability theory

---

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon\phi_\epsilon + \epsilon(h-\hat{h}_\epsilon)$$

Proposition (TMM'19)

1. For all functions $f, \nabla f \in L^4(\rho)$

$$\|T_{\frac{t}{n}}^n f - P_t f\|_{L^2(\rho)} \le \frac{\text{(const.)}}{n}(\|f\|_{L^4(\rho)} + \|\nabla f\|_{L^4(\rho)})$$

2. $T_{\frac{t}{n}}^n$ admits a uniform spectral gap on $L_0^2(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \le \text{(const.)}\|h\|_{L^2(\rho_\epsilon)}$$

3. In the asymptotic limit as $\epsilon \to 0$

$$\|\phi_\epsilon - \phi\|_{L^2(\rho)} \le O(\epsilon)$$

1. Proof of 1: Feynman-Kac representation of semigroup
2. Proof of 2: Foster-Lyapunov condition from stochastic stability theory

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon\phi_\epsilon + \epsilon(h-\hat{h}_\epsilon)$$

Proposition (TMM'19)

1. For all functions $f, \nabla f \in L^4(\rho)$

$$\|T^n_{\frac{t}{n}} f - P_t f\|_{L^2(\rho)} \leq \frac{\text{(const.)}}{n}(\|f\|_{L^4(\rho)} + \|\nabla f\|_{L^4(\rho)})$$

2. $T^n_{\frac{t}{n}}$ admits a uniform spectral gap on $L^2_0(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \leq \text{(const.)}\|h\|_{L^2(\rho_\epsilon)}$$

3. In the asymptotic limit as $\epsilon \to 0$

$$\|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq O(\epsilon)$$

1. Proof of 1: Feynman-Kac representation of semigroup
2. Proof of 2: Foster-Lyapunov condition from stochastic stability theory

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon\phi_\epsilon + \epsilon(h-\hat{h}_\epsilon)$$

Proposition (TMM'19)

1. For all functions $f, \nabla f \in L^4(\rho)$

$$\|T^n_{\frac{t}{n}} f - P_t f\|_{L^2(\rho)} \leq \frac{\text{(const.)}}{n}(\|f\|_{L^4(\rho)} + \|\nabla f\|_{L^4(\rho)})$$

2. $T^n_{\frac{t}{n}}$ admits a uniform spectral gap on $L^2_0(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \leq \text{(const.)}\|h\|_{L^2(\rho_\epsilon)}$$

3. In the asymptotic limit as $\epsilon \to 0$

$$\|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq O(\epsilon)$$

1. Proof of 1: Feynman-Kac representation of semigroup
2. Proof of 2: Foster-Lyapunov condition from stochastic stability theory

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon\phi_\epsilon + \epsilon(h-\hat{h}_\epsilon)$$

Proposition (TMM'19)

1. For all functions $f, \nabla f \in L^4(\rho)$

$$\|T_{\frac{t}{n}}^n f - P_t f\|_{L^2(\rho)} \leq \frac{\text{(const.)}}{n}(\|f\|_{L^4(\rho)} + \|\nabla f\|_{L^4(\rho)})$$

2. $T_{\frac{t}{n}}^n$ admits a uniform spectral gap on $L_0^2(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \leq \text{(const.)}\|h\|_{L^2(\rho_\epsilon)}$$

3. In the asymptotic limit as $\epsilon \to 0$

$$\|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq O(\epsilon)$$

1. Proof of 1: Feynman-Kac representation of semigroup
2. Proof of 2: Foster-Lyapunov condition from stochastic stability theory

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.

$$\phi_\epsilon^{(N)} = T_\epsilon^{(N)}\phi_\epsilon^{(N)} + \epsilon(h - \hat{h}_\epsilon^{(N)}) \quad \text{and} \quad \phi_\epsilon = T_\epsilon\phi_\epsilon + \epsilon(h - \hat{h}_\epsilon)$$

Proposition (TMM'19)

- For any bounded function f and for all $x \in \mathbb{R}^d$:

$$\lim_{N\to\infty} T_\epsilon^{(N)} f(x) = T_\epsilon f(x), \quad \text{a.s}$$

- For any $\delta \in (0,1)$

$$\|T_\epsilon^{(N)} f - T_\epsilon f\|_{L^2(\rho)}^2 \leq O\left(\frac{\log(\frac{N}{\delta})}{N\epsilon^d}\right)$$

with probability larger than $1-\delta$

- If the distribution ρ has compact support

$$\lim_{N\to\infty} \|\phi_\epsilon^{(N)} - \phi_\epsilon\|_\infty = 0, \quad \text{a.s}$$

Approach: Numerical analysis of integral equations on a grid (Anselone, 1971, Atkinson, 1976)

$$\phi_\epsilon^{(N)} = T_\epsilon^{(N)}\phi_\epsilon^{(N)} + \epsilon(h - \hat{h}_\epsilon^{(N)}) \quad \text{and} \quad \phi_\epsilon = T_\epsilon\phi_\epsilon + \epsilon(h - \hat{h}_\epsilon)$$

Proposition (TMM'19)

- For any bounded function f and for all $x \in \mathbb{R}^d$:

$$\lim_{N\to\infty} T_\epsilon^{(N)} f(x) = T_\epsilon f(x), \quad \text{a.s}$$

- For any $\delta \in (0,1)$

$$\|T_\epsilon^{(N)} f - T_\epsilon f\|_{L^2(\rho)}^2 \leq O\left(\frac{\log(\frac{N}{\delta})}{N\epsilon^d}\right)$$

  with probability larger than $1-\delta$

- If the distribution ρ has compact support

$$\lim_{N\to\infty} \|\phi_\epsilon^{(N)} - \phi_\epsilon\|_\infty = 0, \quad \text{a.s}$$

Approach: Numerical analysis of integral equations on a grid (Anselone, 1971, Atkinson, 1976)

$$\phi_\epsilon^{(N)} = T_\epsilon^{(N)}\phi_\epsilon^{(N)} + \epsilon(h - \hat{h}_\epsilon^{(N)}) \quad \text{and} \quad \phi_\epsilon = T_\epsilon \phi_\epsilon + \epsilon(h - \hat{h}_\epsilon)$$

Proposition (TMM'19)

- For any bounded function f and for all $x \in \mathbb{R}^d$:

$$\lim_{N\to\infty} T_\epsilon^{(N)} f(x) = T_\epsilon f(x), \quad \text{a.s}$$

- For any $\delta \in (0,1)$

$$\|T_\epsilon^{(N)} f - T_\epsilon f\|^2_{L^2(\rho)} \leq O(\frac{\log(\frac{N}{\delta})}{N\epsilon^d})$$

  with probability larger than $1-\delta$

- If the distribution ρ has compact support

$$\lim_{N\to\infty} \|\phi_\epsilon^{(N)} - \phi_\epsilon\|_\infty = 0, \quad \text{a.s}$$

Approach: Numerical analysis of integral equations on a grid (Anselone, 1971, Atkinson, 1976)

Error estimate:

$$\text{r.m.s.e} \leq \underbrace{O(\epsilon)}_{\text{bias}} + \underbrace{O(\frac{1}{\epsilon^{1+d/2}N^{1/2}})}_{\text{variance}}$$

Setup: $\rho(x) = \rho_{\text{bimodal}}(x_1)\prod_{n=2}^{d}\rho_{\text{Gaussian}}(x_n)$ and $h(x) = x_1$.

Convergence to const. gain approx.:

$$\lim_{\epsilon\to\infty} \mathsf{K}_\epsilon^{(N)} = \text{const. gain approximatoin}$$

Error estimate:

$$\text{r.m.s.e} \leq \underbrace{O(\epsilon)}_{\text{bias}} + \underbrace{O\Big(\frac{1}{\epsilon^{1+d/2}N^{1/2}}\Big)}_{\text{variance}}$$

Setup: $\rho(x) = \rho_{\text{bimodal}}(x_1)\prod_{n=2}^{d}\rho_{\text{Gaussian}}(x_n)$ and $h(x) = x_1$.

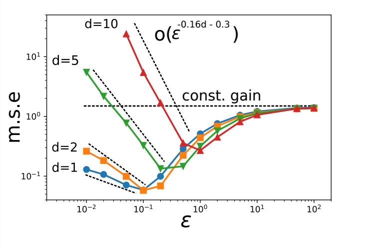
Convergence to const. gain approx.:

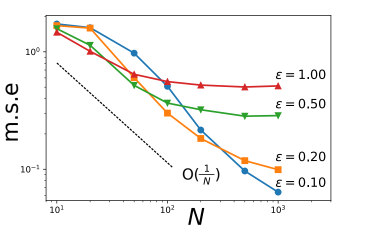
$$\lim_{\epsilon\to\infty} \mathsf{K}_\epsilon^{(N)} = \text{const. gain approximatoin}$$

Error estimate:

$$\text{r.m.s.e} \le \underbrace{O(\epsilon)}_{\text{bias}} + \underbrace{O(\frac{1}{\epsilon^{1+d/2}N^{1/2}})}_{\text{variance}}$$

Setup: $\rho(x) = \rho_{\text{bimodal}}(x_1)\prod_{n=2}^{d}\rho_{\text{Gaussian}}(x_n)$ and $h(x) = x_1$.





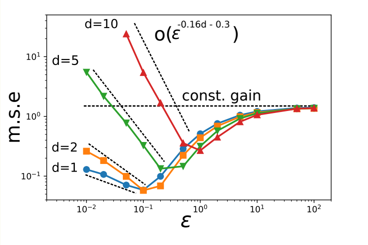
Convergence to const. gain approx.:

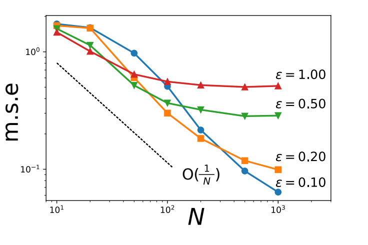
$$\lim_{\epsilon\to\infty} \mathsf{K}_\epsilon^{(N)} = \text{const. gain approximatoin}$$

Error estimate:

$$\text{r.m.s.e} \leq \underbrace{O(\epsilon)}_{\text{bias}} + \underbrace{O\left(\frac{1}{\epsilon^{1+d/2}N^{1/2}}\right)}_{\text{variance}}$$

Setup: $\rho(x) = \rho_{\text{bimodal}}(x_1)\prod_{n=2}^{d}\rho_{\text{Gaussian}}(x_n)$ and $h(x) = x_1$.





Convergence to const. gain approx.:

$$\lim_{\epsilon\to\infty} \mathsf{K}_\epsilon^{(N)} = \text{const. gain approximatoin}$$

Model:

$$dX_t = 0, \quad X_0 \sim \pi_0$$

$$dZ_t = |X_t| dt + \sigma_w dW_t$$

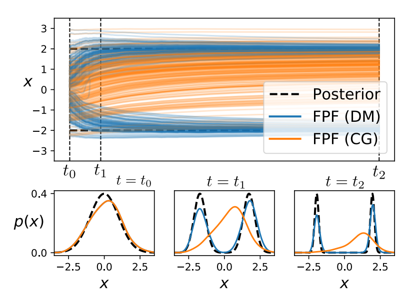
Numerical result:

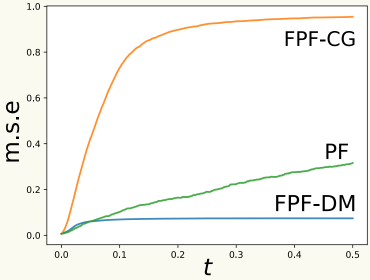
Model:

$$\mathrm{d}X_t = 0, \quad X_0 \sim \pi_0$$

$$\mathrm{d}Z_t = |X_t|\mathrm{d}t + \sigma_w \mathrm{d}W_t$$

Numerical result:





Conclusion

Feedback Particle filter:

- Gain function approximation: Diffusion-map based algorithm and its error analysis
- Optimality properties of control law: An optimal transport formulation of FPF
- Error analysis of the overall algorithm: Error analysis in linear Gaussian setting

Question: Can the three questions addressed in a single framework?

References:

- A. Taghvaei,P. G. Mehta, S. P. Meyn, Diffusion map-based algorithm for gain function approximation in the feedback particle filter, 2019
- A. Taghvaei, P. G. Mehta, An optimal transport formulation of the ensemble Kalman filter, 2019
- A. Taghvaei, J de Wiljes, P. G. Mehta, and S. Reich. Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem. (ASME), 2017

Feedback Particle filter:

- Gain function approximation: Diffusion-map based algorithm and its error analysis
- Optimality properties of control law: An optimal transport formulation of FPF
- Error analysis of the overall algorithm: Error analysis in linear Gaussian setting

Question: Can the three questions addressed in a single framework?

References:

- A. Taghvaei,P. G. Mehta, S. P. Meyn, Diffusion map-based algorithm for gain function approximation in the feedback particle filter, 2019
- A. Taghvaei, P. G. Mehta, An optimal transport formulation of the ensemble Kalman filter, 2019
- A. Taghvaei, J de Wiljes, P. G. Mehta, and S. Reich. Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem. (ASME), 2017

Conclusion

Feedback Particle filter:

- Gain function approximation: Diffusion-map based algorithm and its error analysis
- Optimality properties of control law: An optimal transport formulation of FPF
- Error analysis of the overall algorithm: Error analysis in linear Gaussian setting

Question: Can the three questions addressed in a single framework?

References:

- A. Taghvaei,P. G. Mehta, S. P. Meyn, Diffusion map-based algorithm for gain function approximation in the feedback particle filter, 2019
- A. Taghvaei, P. G. Mehta, An optimal transport formulation of the ensemble Kalman filter, 2019
- A. Taghvaei, J de Wiljes, P. G. Mehta, and S. Reich. Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem. (ASME), 2017

Stochastic thermodynamics:

- Optimality analysis of the power for stochastic thermodynamic systems

Optimal transport map estimation:

- Numerical estimation of the optimal transport map using input convex neural networks.

Accelerated flow for probability distribution:

- Variational formulation of accelerated gradient flows for probability distributions

Thanks for your attention!