Design and Analysis of Particle-based Algorithms for Nonlinear Filtering and Sampling

Doctoral Final Examination

Amirhossein Taghvaei

University of Illinois at Urbana-Champaign

June 25, 2019

Outline

Part I: Analysis of the feedback particle filter (FPF)

- Background (9 slides)
- (1) Gain function approximation (12 slides)
- (2) Optimal transport FPF (back-up slides)
- (3) Error analysis of linear FPF (back-up slides)

Part II: Accelerated flow for probability distributions (8 slides)

Common theme

- Control problem on the space of probability distributions (mean-field control)
- Numerical algorithm as a system of interacting particles

Outline

Part I: Analysis of the feedback particle filter (FPF)

- Background (9 slides)
- (1) Gain function approximation (12 slides)
- (2) Optimal transport FPF (back-up slides)
- (3) Error analysis of linear FPF (back-up slides)

Part II: Accelerated flow for probability distributions (8 slides)

Common theme

- Control problem on the space of probability distributions (mean-field control)
- Numerical algorithm as a system of interacting particles

Outline

Background:

- Filtering problem in discrete time
- Monte-Carlo approximation
- Filtering problem in continuous time
- Feedback particle filter (FPF)

Analysis of FPF

- Gain function approximation (this talk)
- Optimal transport FPF (back-up slides)
- Error analysis of linear FPF (back-up slides)

Objective: Motivate the three topics of contribution

Background: Filtering problem

Mathematical formulation in discrete-time

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

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: Filtering problem

Mathematical formulation in discrete-time

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

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: Filtering problem

Mathematical formulation in discrete-time

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

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

Background: Filtering problem

Mathematical formulation in discrete-time

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

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:

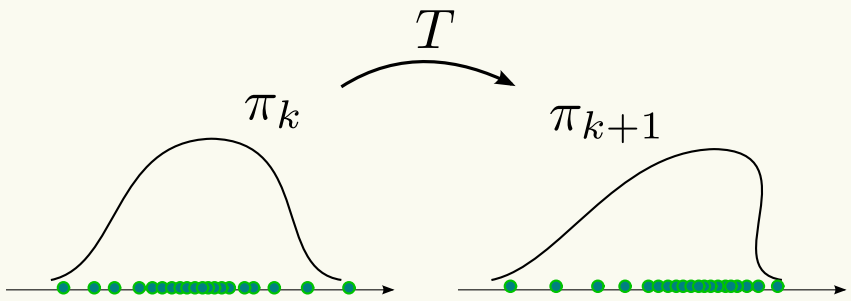
$$\begin{aligned}
\pi_{k-1}(x) \quad &\longrightarrow \quad \tilde{\pi}_k(x) = \int a(x|x')\pi_{k-1}(x')\mathrm{d}x' \\
&\longrightarrow \quad \pi_k(x) = \frac{l(Z_k|x)\tilde{\pi}_k(x)}{\gamma} \qquad \text{(Bayes rule)}
\end{aligned}$$

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

Background: Monte-Carlo approximation

Coupling viewpoint

- Approximate π_k with empirical distribution of particles $\frac{1}{N}\sum_{i=1}^{N}\delta_{X_k^i}$
- Transform particles $\{X_k^i\}_{i=1}^N \sim \pi_k$ to particles $\{X_{k+1}^i\}_{i=1}^N \sim \pi_{k+1}$



Coupling viewpoint:

- Find a coupling $\gamma(x, y)$ between π_k and π_{k+1}

$$\{X_k^i\}_{i=1}^N \longrightarrow \{X_{k+1}^i\}_{i=1}^N$$

- Update the particles with the transition kernel

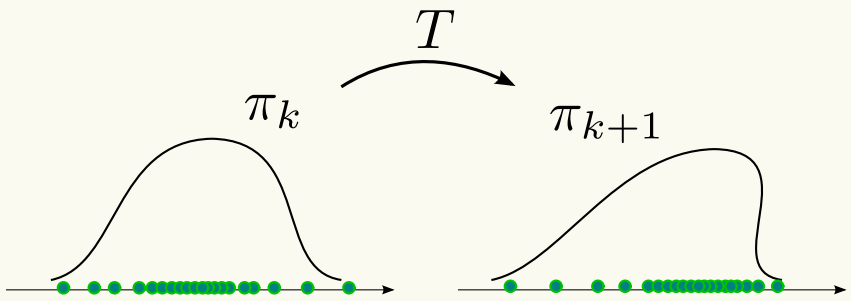
$$X_{k+1}^i \sim T(X_k^i, \cdot)$$

There are infinitely many couplings!

Background: Monte-Carlo approximation

Coupling viewpoint

- Approximate π_k with empirical distribution of particles $\frac{1}{N}\sum_{i=1}^{N}\delta_{X_k^i}$
- Transform particles $\{X_k^i\}_{i=1}^N \sim \pi_k$ to particles $\{X_{k+1}^i\}_{i=1}^N \sim \pi_{k+1}$



Coupling viewpoint:

- Find a coupling $\gamma(\cdot,\cdot)$ between π_k and π_{k+1}

$$\gamma(x,x') = T_k(x|x')\pi_k(x')$$

- Update the particles with the transition kernel

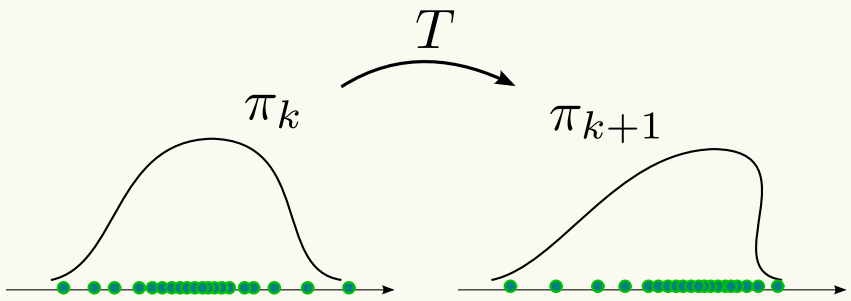
$$X_{k+1}^i \sim T_k(\cdot|X_k^i)$$

There are infinitely many couplings!

Background: Monte-Carlo approximation

Coupling viewpoint

- Approximate π_k with empirical distribution of particles $\frac{1}{N}\sum_{i=1}^{N}\delta_{X_k^i}$
- Transform particles $\{X_k^i\}_{i=1}^N \sim \pi_k$ to particles $\{X_{k+1}^i\}_{i=1}^N \sim \pi_{k+1}$



Coupling viewpoint:

- Find a coupling $\gamma(\cdot,\cdot)$ between π_k and π_{k+1}

$$\gamma(x,x') = T_k(x|x')\pi_k(x')$$

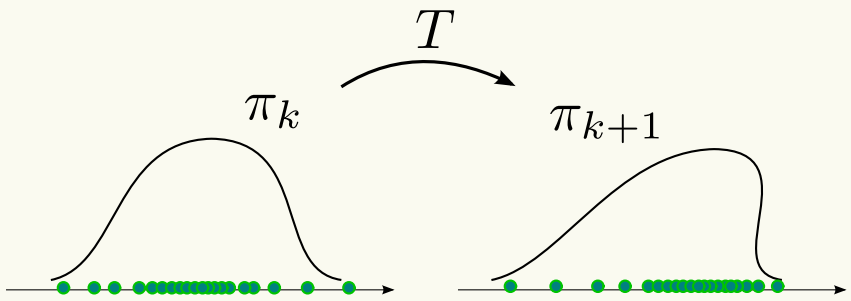
- Update the particles with the transition kernel

$$X_{k+1}^i \sim T_k(\cdot|X_k^i)$$

There are infinitely many couplings!

Background: Monte-Carlo approximation
Coupling viewpoint

- Approximate π_k with empirical distribution of particles $\frac{1}{N}\sum_{i=1}^{N}\delta_{X_k^i}$
- Transform particles $\{X_k^i\}_{i=1}^N \sim \pi_k$ to particles $\{X_{k+1}^i\}_{i=1}^N \sim \pi_{k+1}$



Coupling viewpoint:

- Find a coupling $\gamma(\cdot,\cdot)$ between π_k and π_{k+1}

$$\gamma(x,x') = T_k(x|x')\pi_k(x')$$

- Update the particles with the transition kernel

$$X_{k+1}^i \sim T_k(\cdot|X_k^i)$$

There are infinitely many couplings!

Background: Brief review of algorithms

Coupling viewpoint

Assume there is no dynamics: $\pi_{k+1}(x) = \dfrac{l_k(x)\pi_k(x)}{\gamma}$

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

- Sample from the independent coupling:

$$\mathbb{P}(x_k^i, x_k^j) = \pi_{k+1}(x_k^i)\pi_{k+1}(x_k^j) \quad \text{for } i, j = 1, \ldots, N$$

  But π_{k+1} is not explicitly known!

- Approximate $\pi_{k+1} \approx \sum_{i=1}^{N} w_k^i \delta_{x_k^i}$ where $w_k^i = \dfrac{l_k(x_k^i)}{\sum_{j=1}^{N} l_k(x_k^j)}$

(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)

Background: Brief review of algorithms
Coupling viewpoint

Assume there is no dynamics: $\pi_{k+1}(x) = \dfrac{l_k(x)\pi_k(x)}{\gamma}$

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

- Sample from the independent coupling:

$$\gamma(x,' x) = \pi_{k+1}(x)\pi_k(x') \quad \Rightarrow \quad X_{k+1}^i \sim \pi_{k+1}(\cdot)$$

  But π_{k+1} is not explicitly known!

- Approximate $\pi_{k+1} \approx \sum_{i=1}^{N} w_i \delta_{X_k^i}$ where $w_i = \dfrac{l_k(X_k^i)}{\sum_{j=1}^{N} l_k(X_k^j)}$

(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)

Background: Brief review of algorithms
Coupling viewpoint

Assume there is no dynamics: $\pi_{k+1}(x) = \dfrac{l_k(x)\pi_k(x)}{\gamma}$

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

- Sample from the independent coupling:

$$\gamma(x,' x) = \pi_{k+1}(x)\pi_k(x') \quad \Rightarrow \quad X_{k+1}^i \sim \pi_{k+1}(\cdot)$$

  But π_{k+1} is not explicitly known!

- Approximate $\pi_{k+1} \approx \sum_{i=1}^{N} w_i \delta_{X_k^i}$ where $w_i = \dfrac{l_k(X_k^i)}{\sum_{j=1}^{N} l_k(X_k^j)}$

(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)

Background: Brief review of algorithms
Coupling viewpoint

Assume there is no dynamics: $\pi_{k+1}(x) = \dfrac{l_k(x)\pi_k(x)}{\gamma}$

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

- Sample from the independent coupling:

$$\gamma(x,' x) = \pi_{k+1}(x)\pi_k(x') \quad \Rightarrow \quad X_{k+1}^i \sim \pi_{k+1}(\cdot)$$

  But π_{k+1} is not explicitly known!

- Approximate $\pi_{k+1} \approx \sum_{i=1}^{N} w_i \delta_{X_k^i}$ where $w_i = \dfrac{l_k(X_k^i)}{\sum_{j=1}^{N} l_k(X_k^j)}$

(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)

Background: Brief review of algorithms
Coupling viewpoint

Assume there is no dynamics: $\pi_{k+1}(x) = \dfrac{l_k(x)\pi_k(x)}{\gamma}$

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

- Sample from the independent coupling:

$$\gamma(x,' x) = \pi_{k+1}(x)\pi_k(x') \quad \Rightarrow \quad X_{k+1}^i \sim \pi_{k+1}(\cdot)$$

  But π_{k+1} is not explicitly known!

- Approximate $\pi_{k+1} \approx \sum_{i=1}^{N} w_i \delta_{X_k^i}$ where $w_i = \dfrac{l_k(X_k^i)}{\sum_{j=1}^{N} l_k(X_k^j)}$

(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)

Background: Monte-Carlo approximation

Three questions

1. Approximation: Numerical approximate of a coupling

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

$$\textbf{Approximate:} \quad \text{a coupling between } \pi_k \text{ and } \pi_{k+1}$$

2. Design: Which coupling to choose?

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

Background: Monte-Carlo approximation

Three questions

1. Approximation: Numerical approximate of a coupling

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

$$\textbf{Approximate:} \quad \text{a coupling between } \pi_k \text{ and } \pi_{k+1}$$

2. Design: Which coupling to choose?

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

Background: Monte-Carlo approximation

Three questions

1. Approximation: Numerical approximate of a coupling

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

$$\textbf{Approximate:} \quad \text{a coupling between } \pi_k \text{ and } \pi_{k+1}$$

2. Design: Which coupling to choose?

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

Background: Filtering problem

Mathematical formulation in continuous-time

State process: $\mathrm{d}X_t = a(X_t)\mathrm{d}t + \sigma_B(X_t)\mathrm{d}B_t, \qquad X_0 \sim p_0(\cdot)$

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

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

Solution:

$$\mathrm{d}\pi_t = (\text{Kushner-Stratonovic eq.})$$

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

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

Background: Filtering problem

Mathematical formulation in continuous-time

State process: $\mathrm{d}X_t = a(X_t)\mathrm{d}t + \sigma_B(X_t)\mathrm{d}B_t, \qquad X_0 \sim p_0(\cdot)$

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:

$$\mathrm{d}\pi_t = (\text{Kushner-Stratonovic eq.})$$

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

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

Background: Filtering problem

Mathematical formulation in continuous-time

State process: $\mathrm{d}X_t = a(X_t)\mathrm{d}t + \sigma_B(X_t)\mathrm{d}B_t, \qquad X_0 \sim p_0(\cdot)$

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:

$$\mathrm{d}\pi_t = \text{(Kushner-Stratonovic eq.)}$$

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

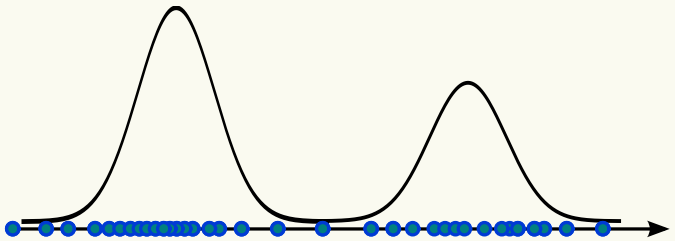
J. Xiong, An introduction to stochastic filtering theory, 2008

Background: Monte-Carlo approximation

Continuous-time setting

Idea:

- Approximate π_t with empirical distribution of particles $\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}$
- Continuously update particles such that $\{X_t^i\}_{i=1}^N \sim \pi_t$



$\{X_t^1, \ldots, X_t^N\}$

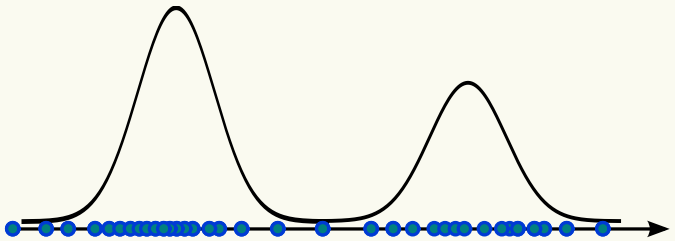
Control viewpoint: Find a control law

$$\mathrm{d}X_t^i = (\text{control law}) \quad \text{such that} \quad \underbrace{X_t^i \sim \pi_t}_{\text{exactness}}$$

Background: Monte-Carlo approximation
Continuous-time setting

Idea:

- Approximate π_t with empirical distribution of particles $\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^i}$
- Continuously update particles such that $\{X_t^i\}_{i=1}^N \sim \pi_t$



$\{X_t^1, \ldots, X_t^N\}$

Control viewpoint: Find a control law

$$\mathrm{d}X_t^i = (\text{control law}) \quad \text{such that} \quad \underbrace{X_t^i \sim \pi_t}_{\text{exactness}}$$

Background: Feedback particle filter (FPF)

Update formula:

$$\mathrm{d}X_t^i = \text{(dynamic model)} + \underbrace{\mathsf{K}_t(X_t^i) \circ (\mathrm{d}Z_t - \frac{h(X_t^i) + \hat{h}_t}{2}\mathrm{d}t)}_{\text{correction}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} \pi_0$$

- Gain function $\mathsf{K}_t(x) = \nabla\phi_t(x)$ where ϕ_t solves the Poisson eq.

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

- $\hat{h}_t = \mathsf{E}[h(X_t^i)|\mathcal{Z}_t]$
- ρ_t is probability density for X_t^i

FPF is exact: $X_t^i \sim \pi_t$

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

Background: Feedback particle filter (FPF)

Update formula:

$$\mathrm{d}X_t^i = \text{(dynamic model)} + \underbrace{\mathsf{K}_t(X_t^i) \circ (\mathrm{d}Z_t - \frac{h(X_t^i) + \hat{h}_t}{2}\mathrm{d}t)}_{\text{correction}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} \pi_0$$

- Gain function $\mathsf{K}_t(x) = \nabla\phi_t(x)$ where ϕ_t solves the Poisson eq.

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

- $\hat{h}_t = \mathsf{E}[h(X_t^i)|\mathcal{Z}_t]$
- ρ_t is probability density for X_t^i

FPF is exact: $X_t^i \sim \pi_t$

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

Background: Feedback particle filter (FPF)

Update formula:

$$\mathrm{d}X_t^i = \text{(dynamic model)} + \underbrace{\mathsf{K}_t(X_t^i) \circ (\mathrm{d}Z_t - \frac{h(X_t^i) + \hat{h}_t}{2}\mathrm{d}t)}_{\text{correction}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} \pi_0$$

- Gain function $\mathsf{K}_t(x) = \nabla\phi_t(x)$ where ϕ_t solves the Poisson eq.

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

- $\hat{h}_t = \mathsf{E}[h(X_t^i)|\mathcal{Z}_t]$
- ρ_t is probability density for X_t^i

FPF is exact: $X_t^i \sim \pi_t$

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

Background: Feedback particle filter

Three questions

1. Approximation: Numerical approximation of the gain function

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

$$\textbf{Approximate:} \quad \{\mathsf{K}_t(X_t^1), \ldots, \mathsf{K}(X_t^N)\}_{i=1}^N$$

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

Background: Feedback particle filter

Three questions

1. Approximation: Numerical approximation of the gain function

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

$$\textbf{Approximate:} \quad \{\mathsf{K}_t(X_t^1), \ldots, \mathsf{K}(X_t^N)\}_{i=1}^N$$

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

Background: Feedback particle filter

Three questions

1. Approximation: Numerical approximation of the gain function

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

$$\textbf{Approximate:} \quad \{\mathsf{K}_t(X_t^1), \ldots, \mathsf{K}(X_t^N)\}_{i=1}^N$$

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: Analysis of the feedback particle filter (FPF)

- Background
- Gain function approximation
  - A. Taghvaei and P. G. Mehta, Gain function approximation in the feedback particle filter, in Conference on Decision and Control (CDC), 2016, pp. 5446–5452.
  - A. Taghvaei, P. G. Mehta, and S. P. Meyn, Error estimates for the kernel gain function approximation in the feedback particle filter, in American Control Conference (ACC), 2017, pp. 4576–4582.
  - A. Taghvaei, P. G. Mehta, and S. P. Meyn, Gain function approximation in the feedback particle filter, SIAM journal on Uncertainty Quantification (under review),
- Optimal transport FPF
- Error analysis of linear FPF

Part II: Accelerated flow for probability distributions

FPF update formula:

$$\mathrm{d}X_t^i = \text{(dynamic model)} + \mathsf{K}_t(X_t^i) \circ (\mathrm{d}Z_t - \frac{1}{2}(h(X_t^i) + \hat{h}_t)\mathrm{d}t)$$

Gain function $\quad \mathsf{K}_t(x) = \nabla\phi_t(x) \quad$ where ϕ_t solves the Poisson eq.

Poisson equation:

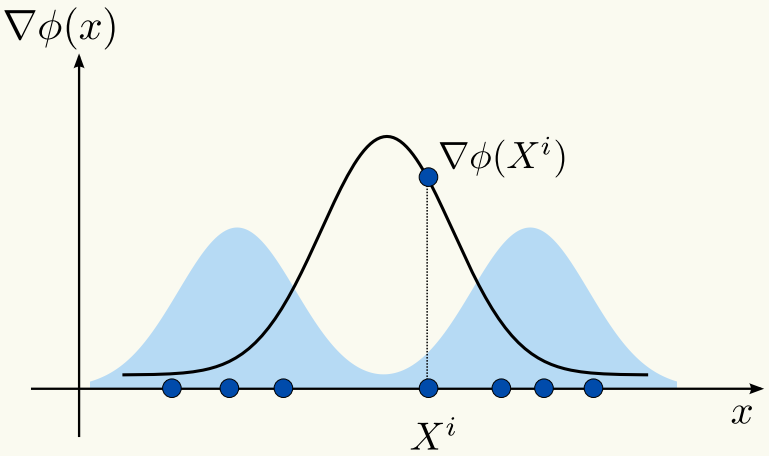
$$-\Delta_\rho \phi(x) = h(x) - \hat{h}$$

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

Computational problem:

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

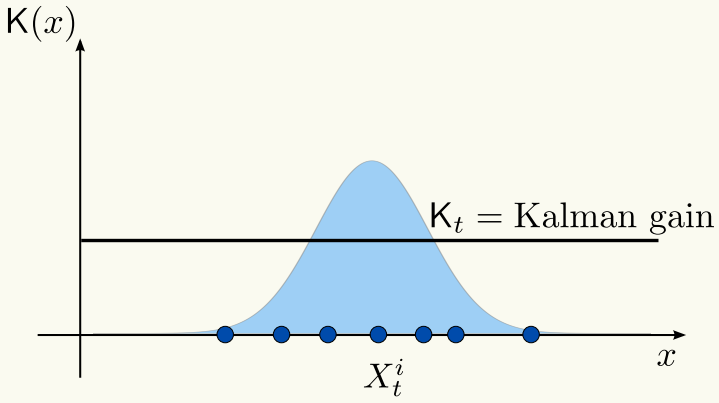
Approximate: $\{\mathsf{K}(X^1), \ldots, \mathsf{K}(X^N)\}$



Linear Gaussian setting

- ρ is Gaussian $\mathcal{N}(m, \Sigma)$
- $h(x) = Hx$

$$\mathsf{K}(x) = \Sigma H^\top \quad \text{(Kalman gain)}$$



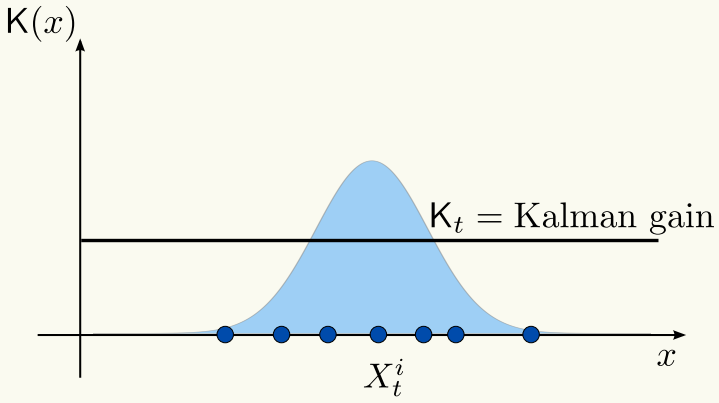
FPF $\xrightarrow{\text{linear Gaussian}}$ Ensemble Kalman filter (EnKF)

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

Linear Gaussian setting

- ρ is Gaussian $\mathcal{N}(m, \Sigma)$
- $h(x) = Hx$

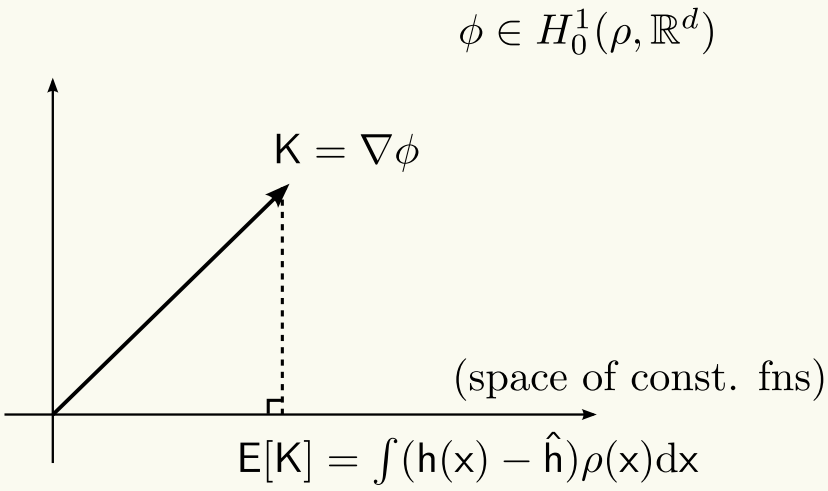
$$\mathsf{K}(x) = \Sigma H^\top \quad \text{(Kalman gain)}$$

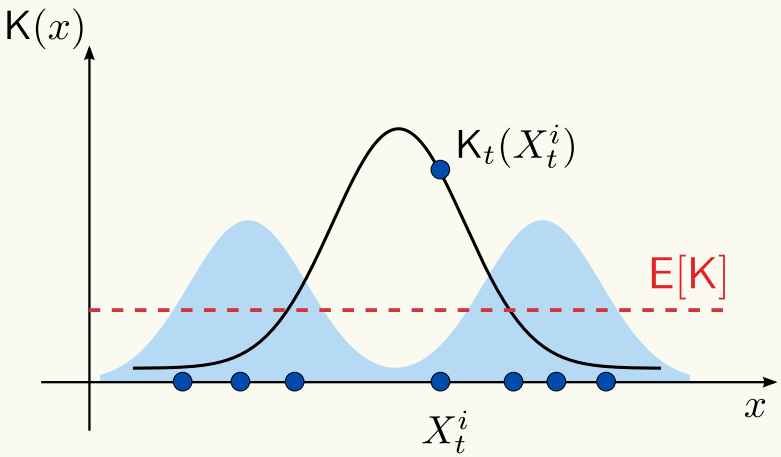


$$\boxed{\text{FPF} \xrightarrow{\text{linear Gaussian}} \text{Ensemble Kalman filter (EnKF)}}$$

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

Idea: Projection to space of constant functions



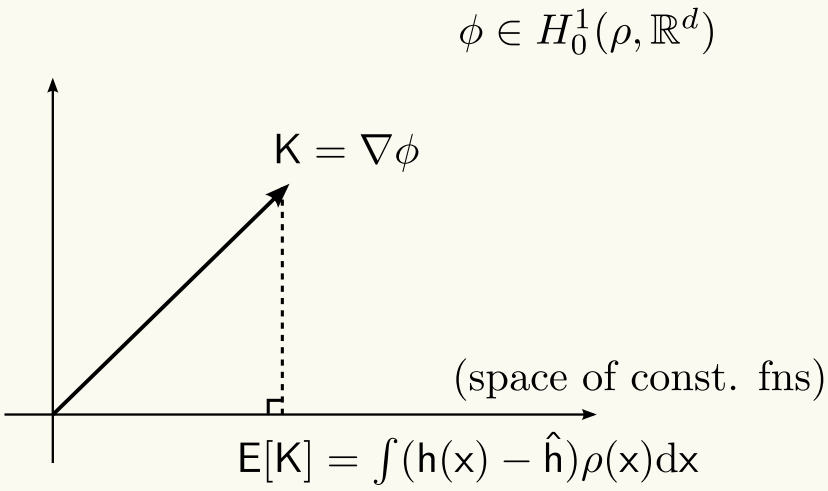


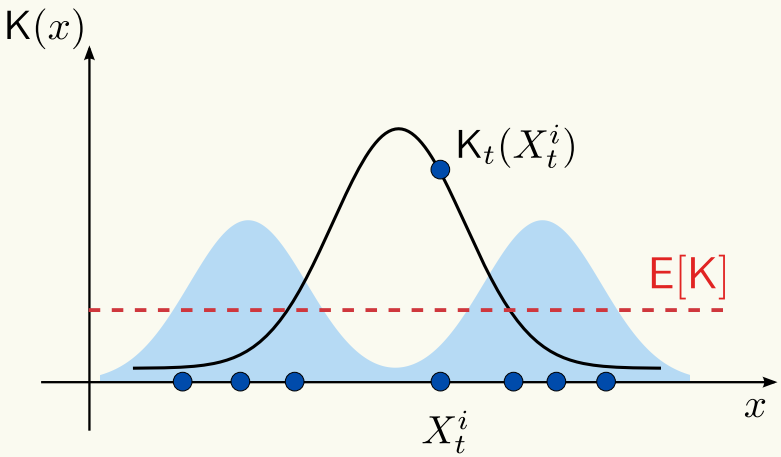
A closed-form formula:

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

Can we improve this approximation?

Idea: Projection to space of constant functions



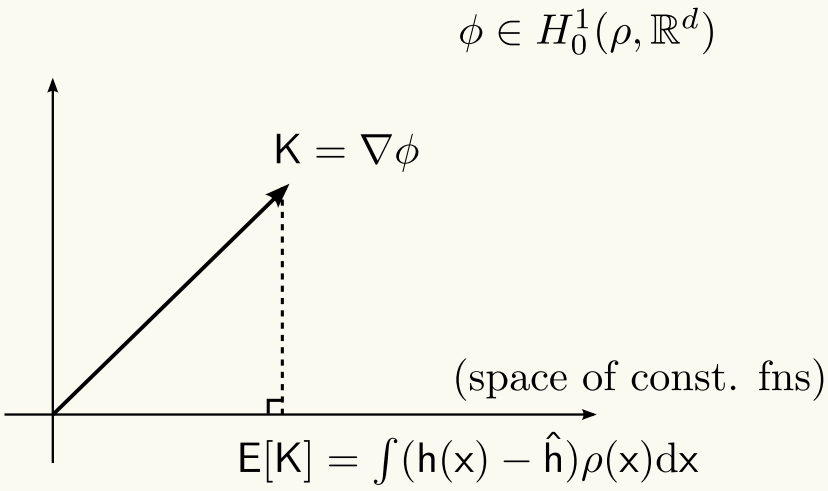


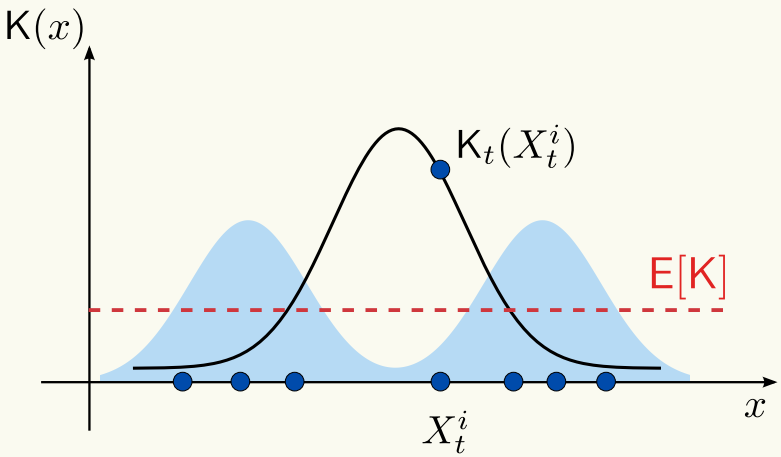
A closed-form formula:

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

Can we improve this approximation?

Idea: Projection to space of constant functions





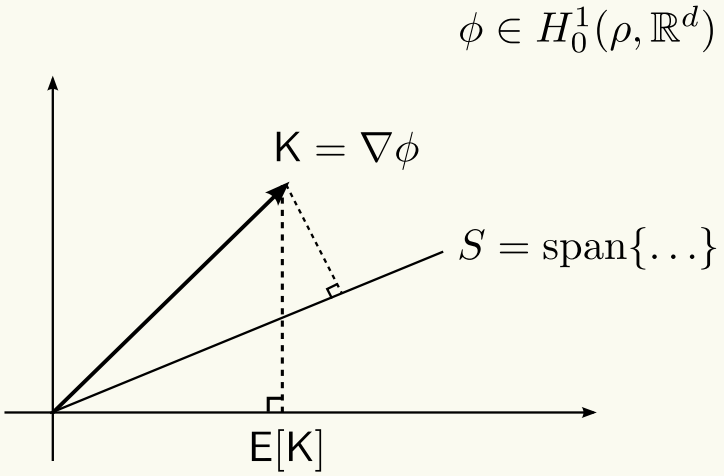
A closed-form formula:

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

$\boxed{\text{Can we improve this approximation?}}$

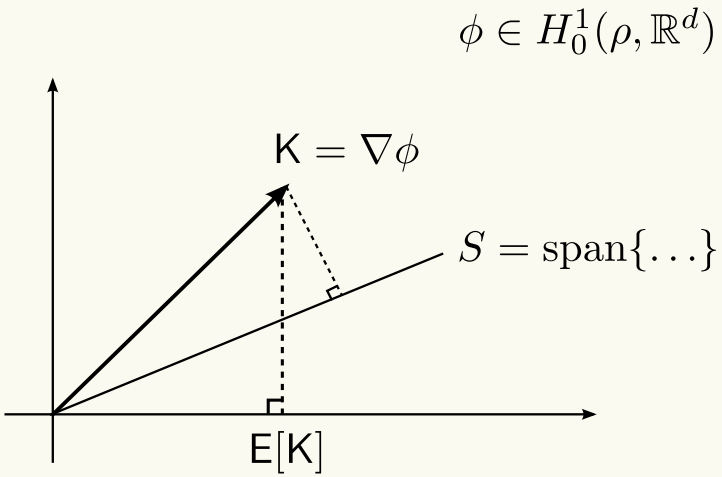
Gain function approximation
Galerkin approximation

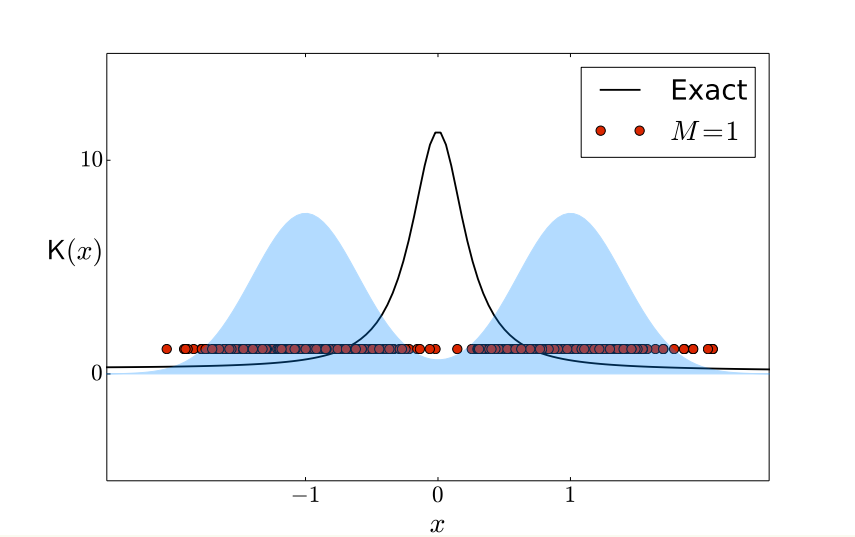
Idea: Projection into a finite-dim subspace



Choice of basis function is difficult

Idea: Projection into a finite-dim subspace

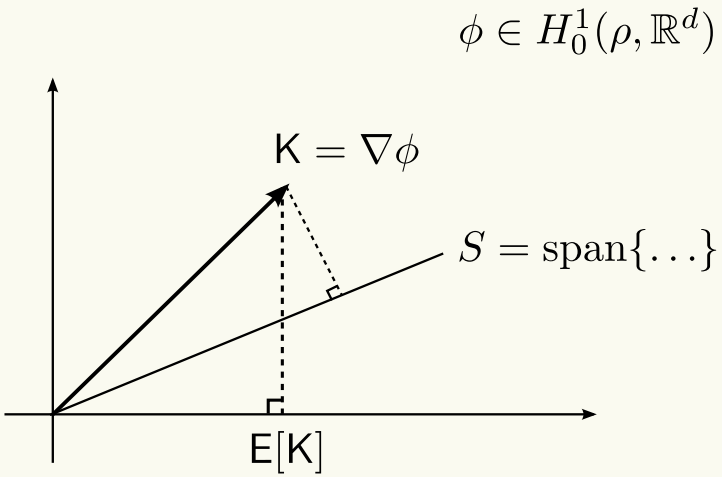


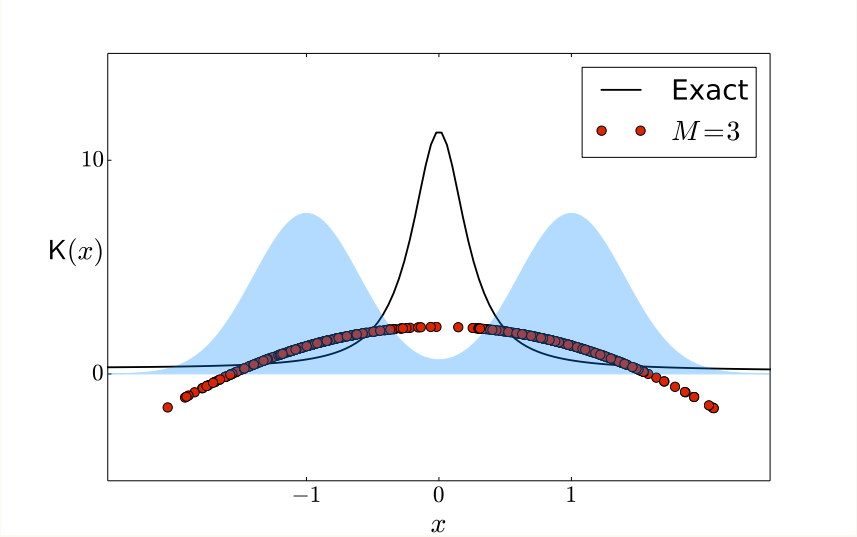


$$S = \mathsf{span}\{1, x, \ldots, x^M\}$$

Choice of basis function is difficult

Idea: Projection into a finite-dim subspace

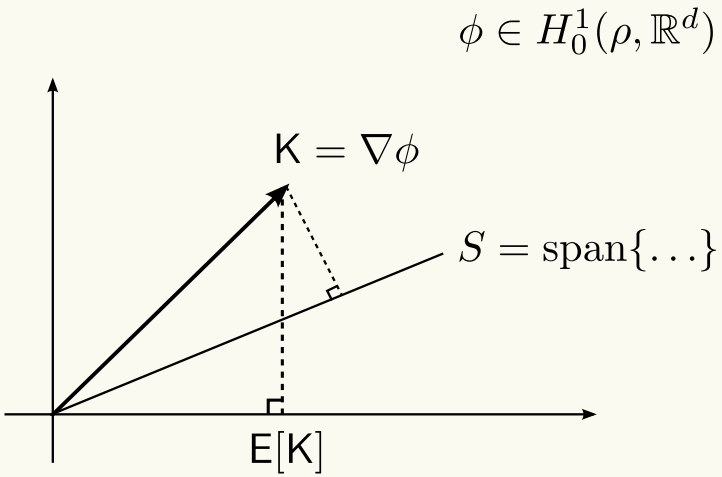


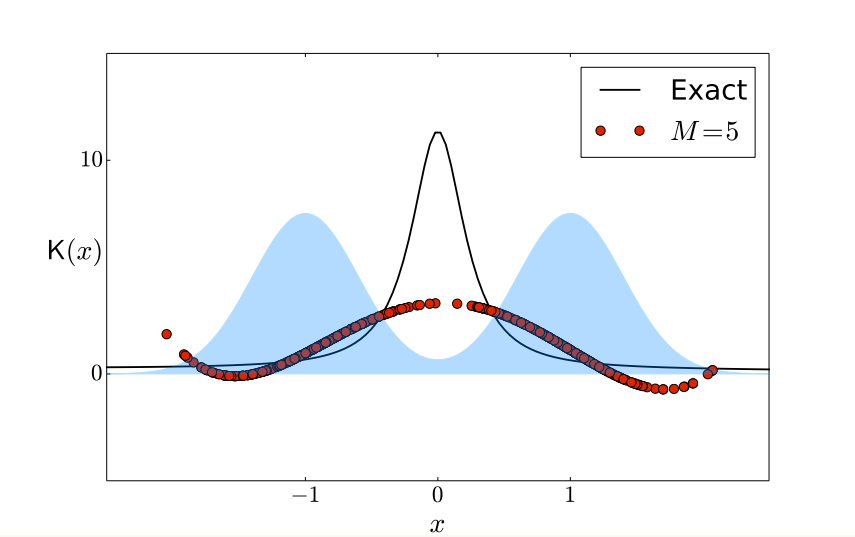


$$S = \mathsf{span}\{1, x, \ldots, x^M\}$$

Choice of basis function is difficult

Idea: Projection into a finite-dim subspace



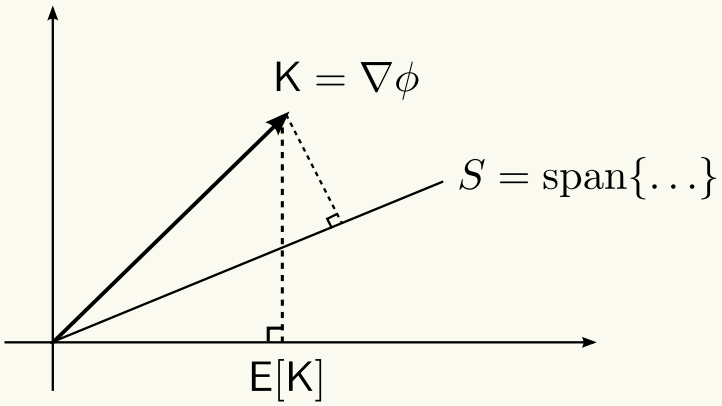


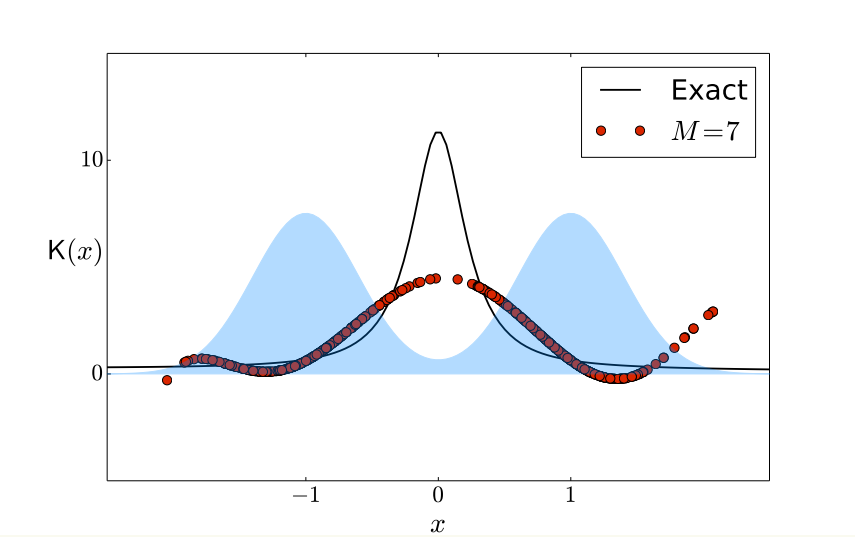
$$S = \mathsf{span}\{1, x, \ldots, x^M\}$$

Choice of basis function is difficult

Idea: Projection into a finite-dim subspace

$\phi \in H_0^1(\rho, \mathbb{R}^d)$

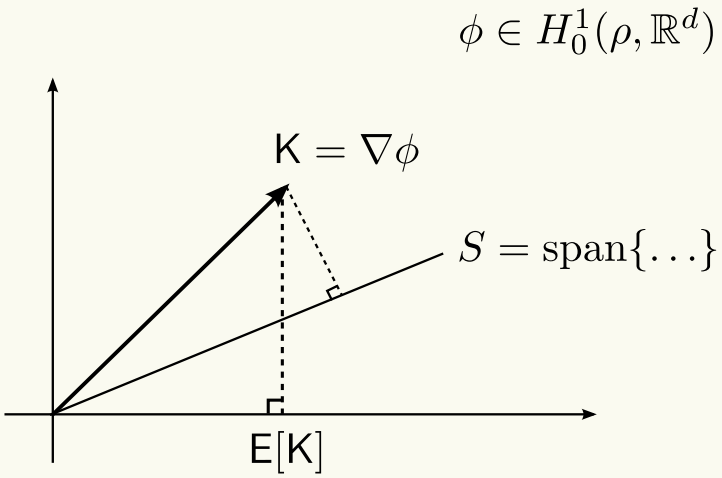


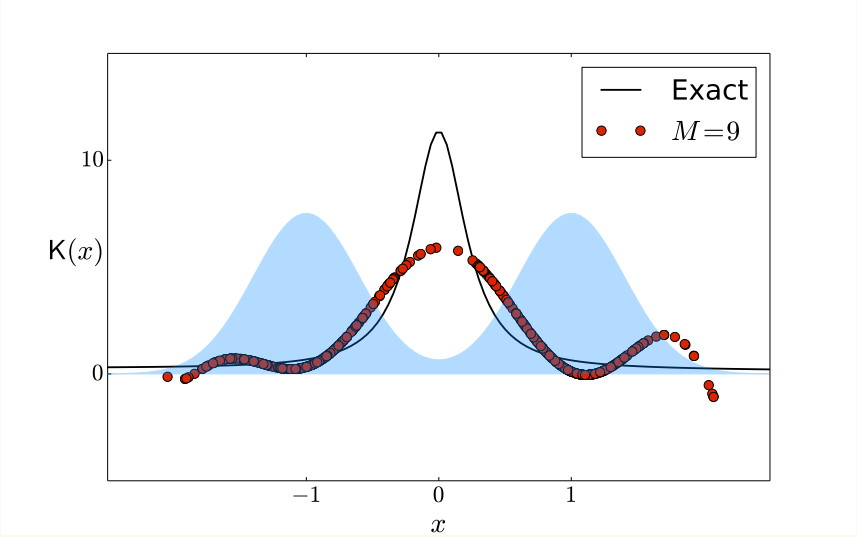


$$S = \text{span}\{1, x, \ldots, x^M\}$$

Choice of basis function is difficult

Idea: Projection into a finite-dim subspace

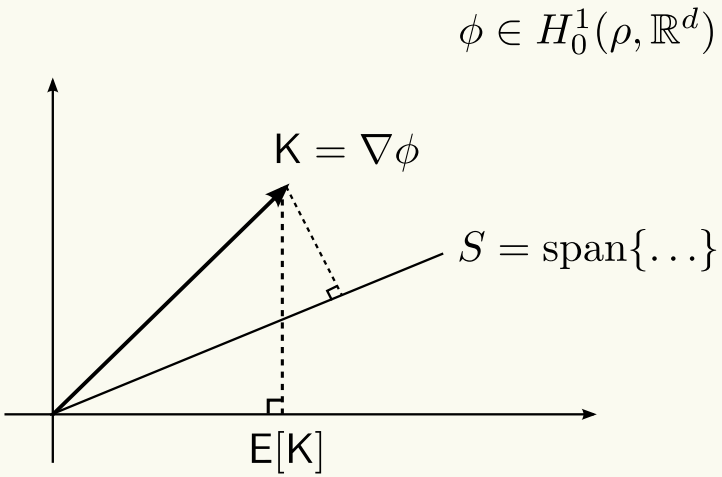


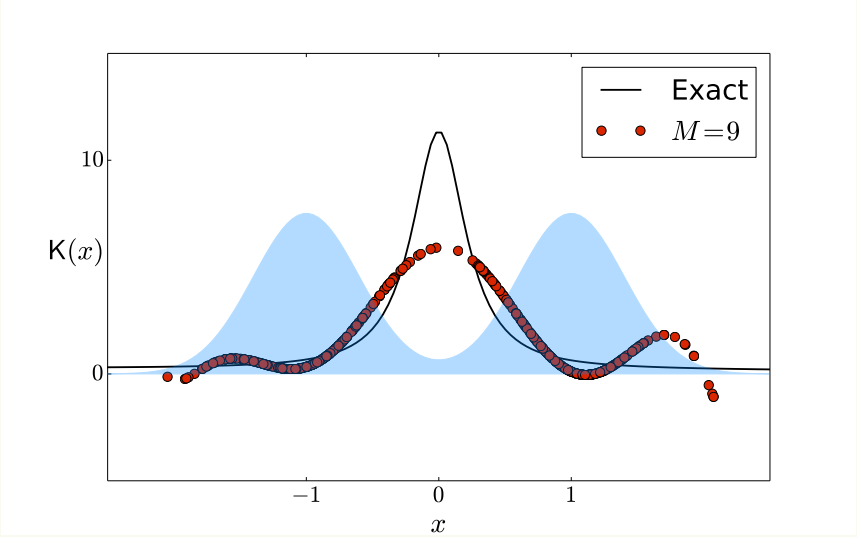


$$S = \mathsf{span}\{1, x, \ldots, x^M\}$$

Choice of basis function is difficult

Idea: Projection into a finite-dim subspace





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

Choice of basis function is difficult

Gain function approximation: Diffusion map-based algorithm

Overview

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

(approximation steps)

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

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

Gain function approximation: Diffusion map-based algorithm

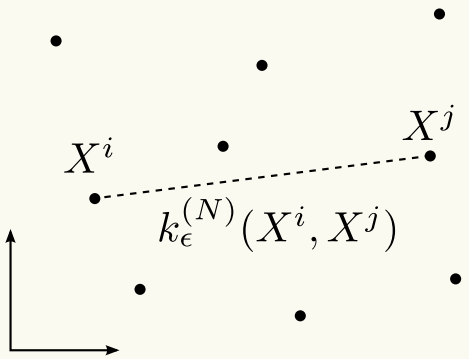
Overview

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

(approximation steps)

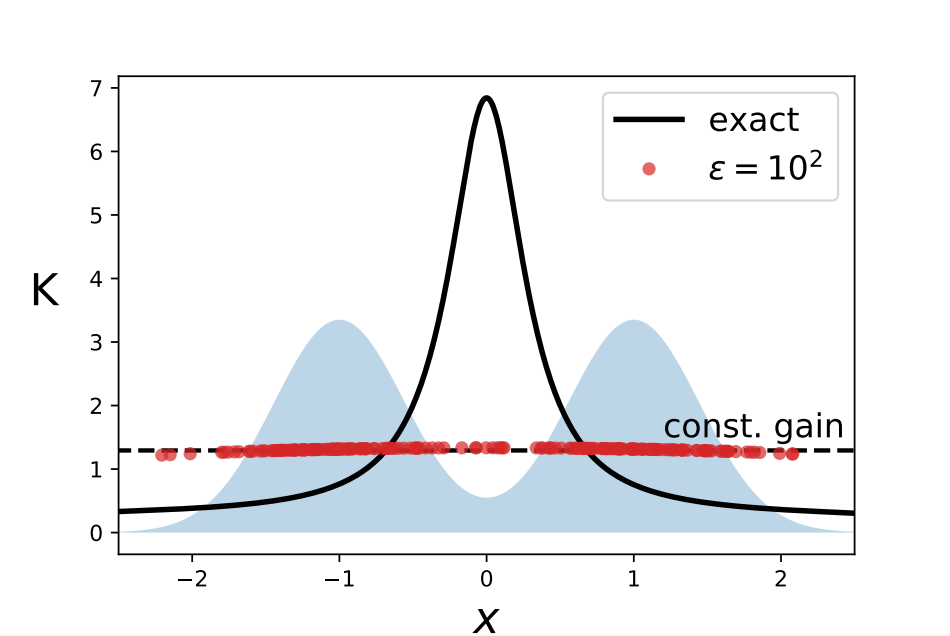
Finite-dim eq.: $\Phi = \mathsf{T}\Phi + \epsilon(\mathsf{h} - \pi(h)), \quad \Phi \in \mathbb{R}_0^N$

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



Gain function approximation: Diffusion map-based algorithm

Numerical analysis



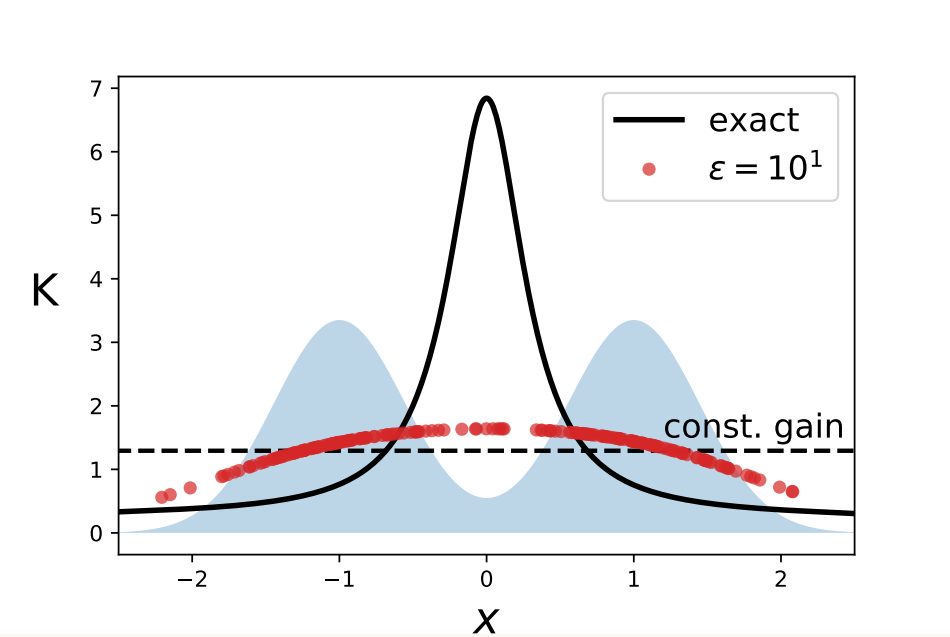
Properties:

- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Numerical analysis



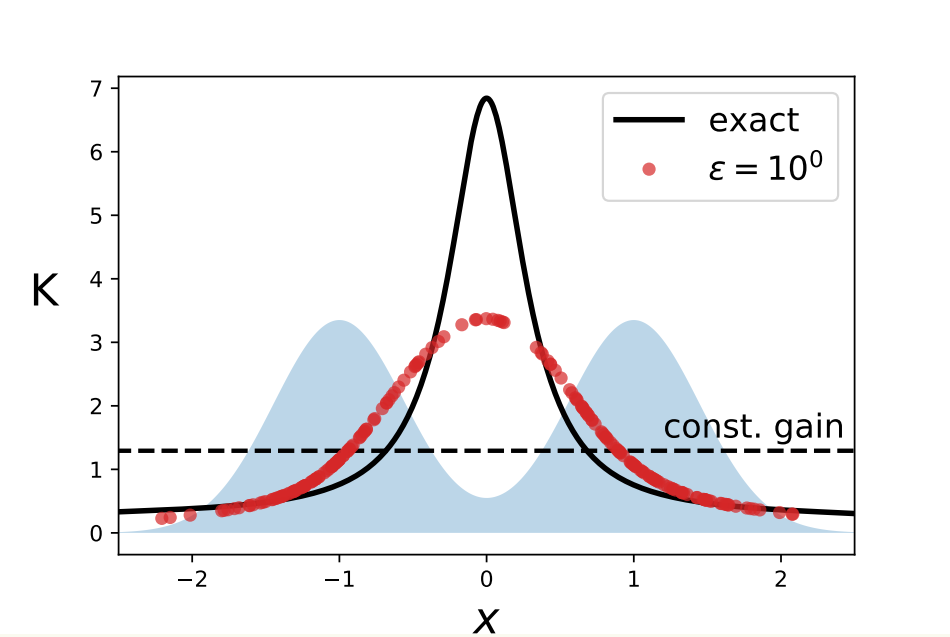
Properties:

- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Numerical analysis



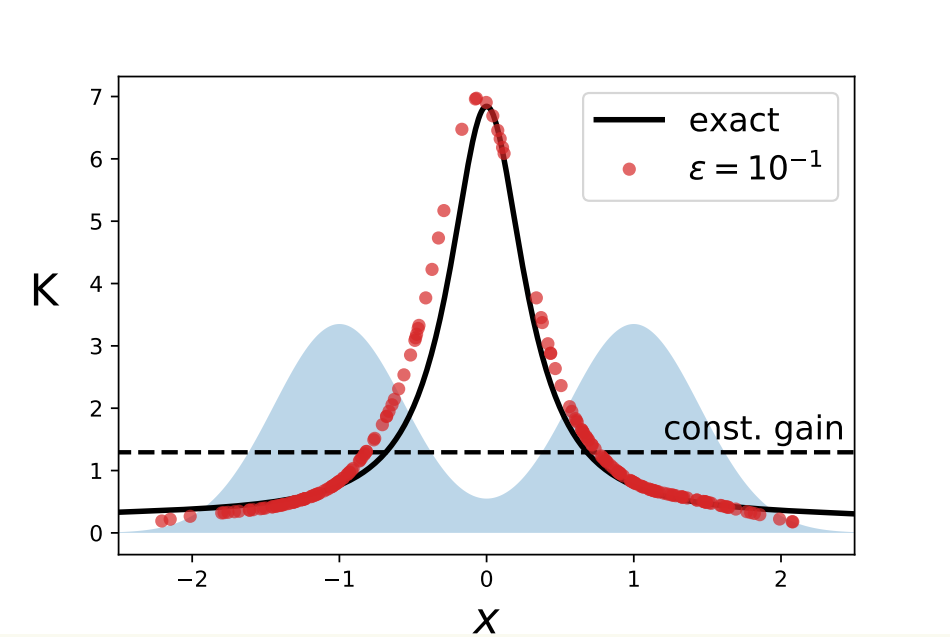
Properties:

- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Numerical analysis



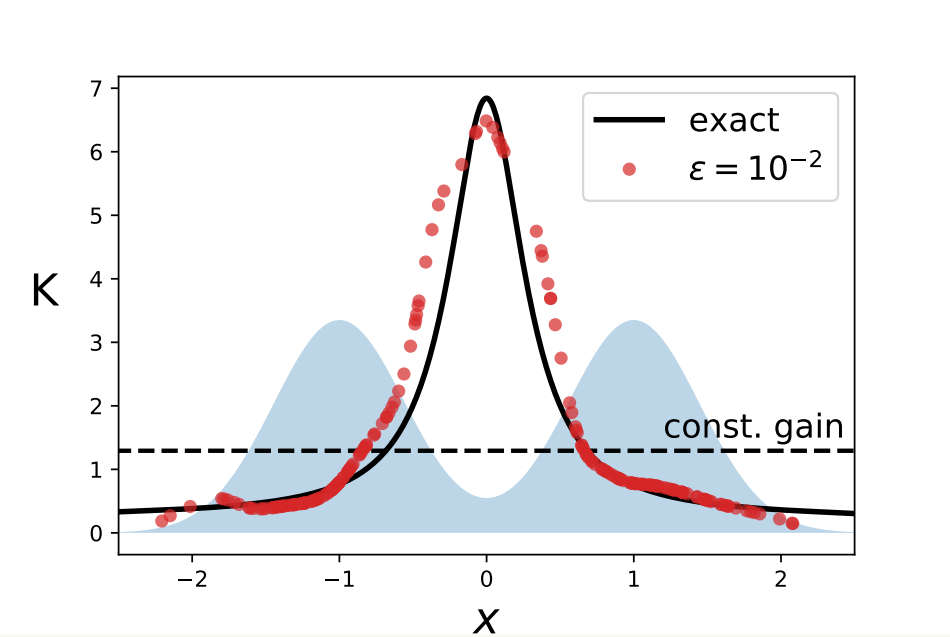
Properties:

- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Numerical analysis



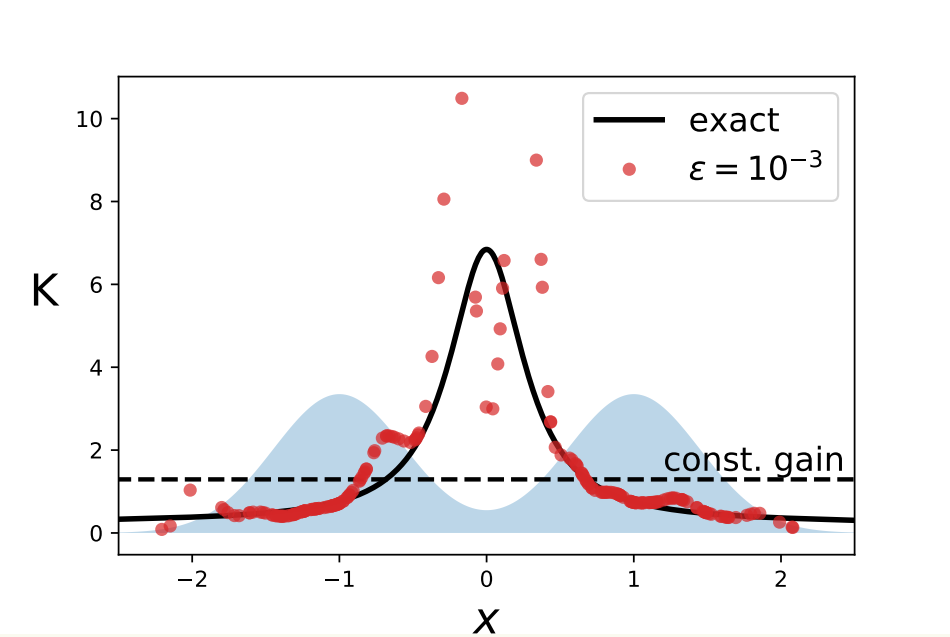
Properties:

- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Numerical analysis



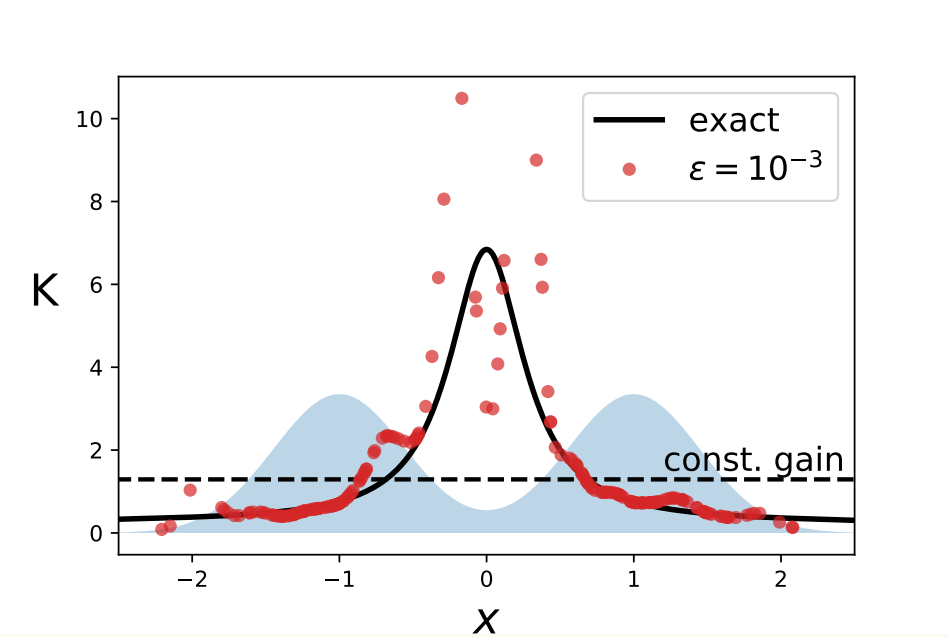
Properties:

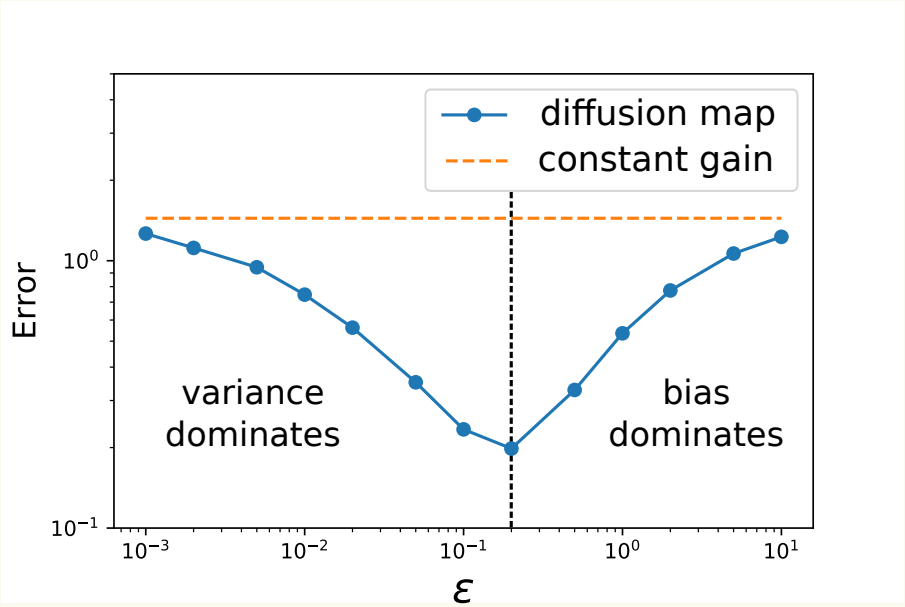
- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Numerical analysis





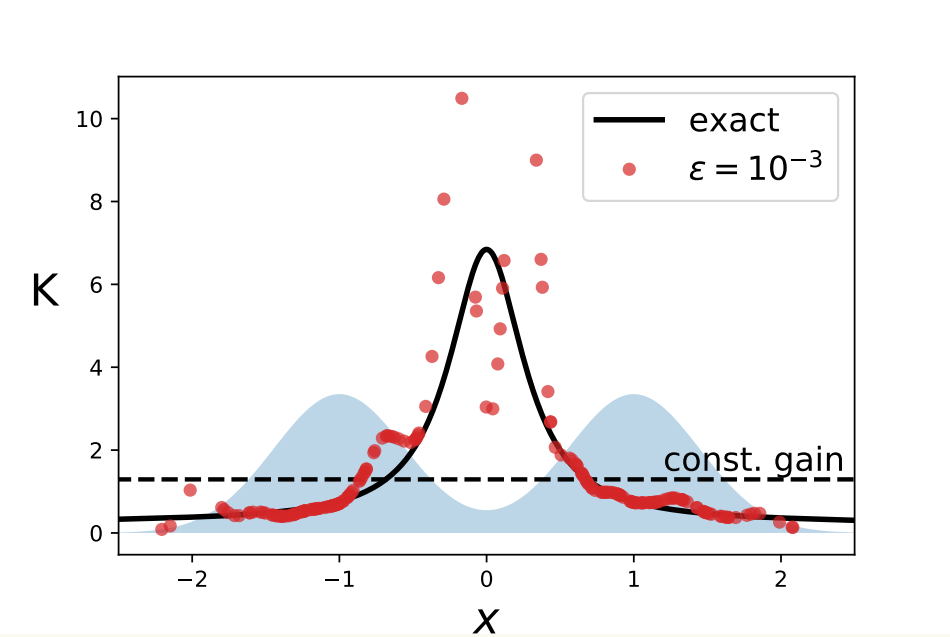
Properties:

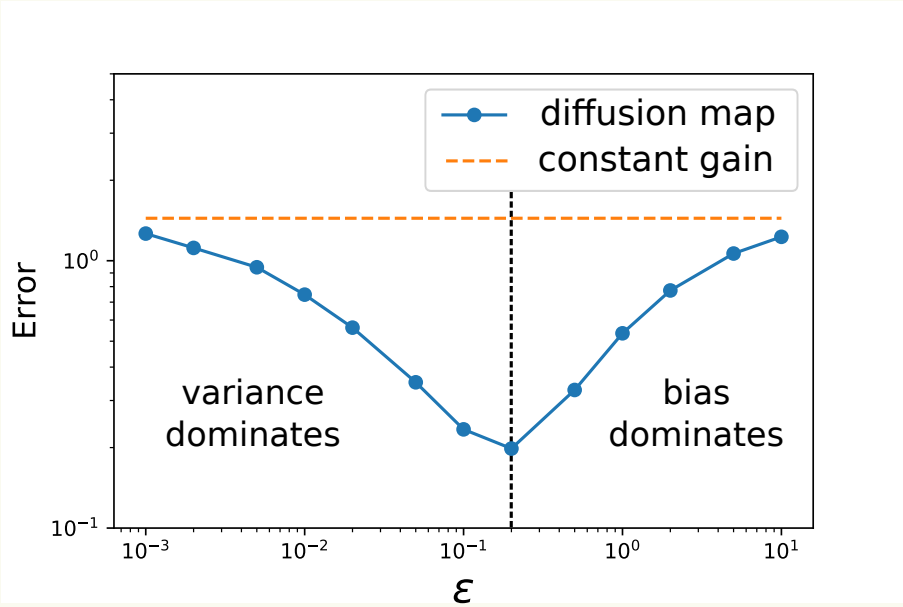
- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Numerical analysis





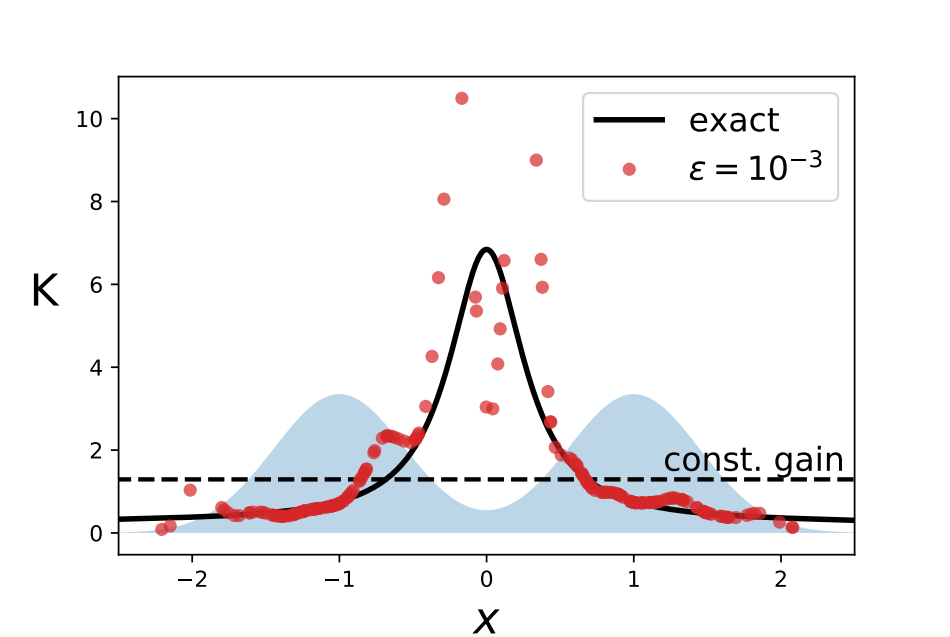
Properties:

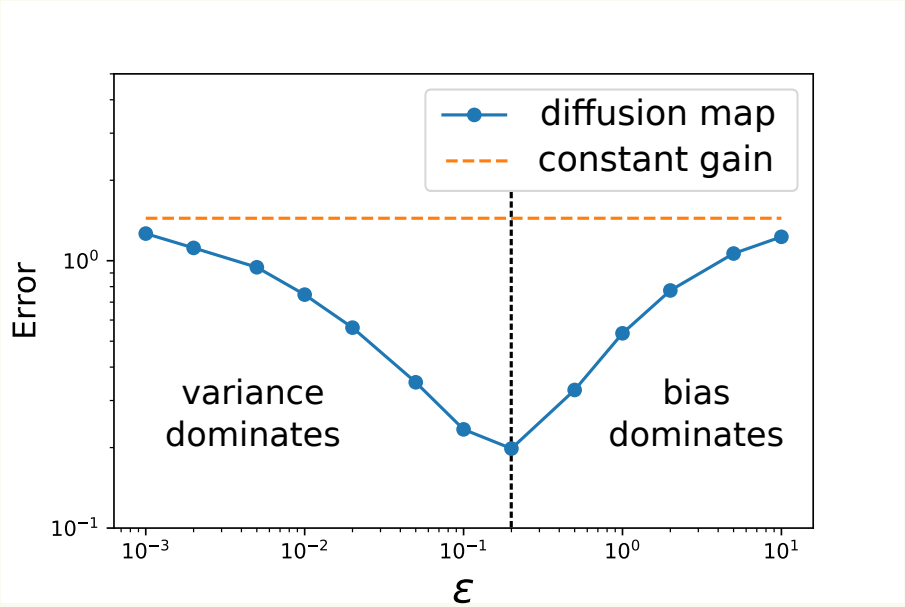
- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Numerical analysis





Properties:

- No basis function selection
- Reduces to constant gain in the limit as $\epsilon \to \infty$

Objective: Error analysis of bias and variance

Gain function approximation: Diffusion map-based algorithm

Four Markov operators

- Semigroup $P_t = e^{t\Delta_\rho}$

$$\frac{\partial}{\partial t} P_t f = P_t \Delta_\rho f$$

- Diffusion map T_ϵ

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

- Empirical approximation $T_\epsilon^{(N)}$

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

- $N \times N$ Markov matrix T

$$\mathsf{T}_{ij} := \frac{g_\epsilon(X^i, X^j)}{n_\epsilon^{(N)}(X^i)\sqrt{\sum_{l=1}^{N} g_\epsilon(X^j, X^l)}}$$

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)

Gain function approximation: Diffusion map-based algorithm

Four Markov operators

- Semigroup $P_t = e^{t\Delta_\rho}$

$$\frac{\partial}{\partial t} P_t f = P_t \Delta_\rho f$$

- Diffusion map T_ϵ

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

- Empirical approximation $T_\epsilon^{(N)}$

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

- $N \times N$ Markov matrix T

$$\mathsf{T}_{ij} := \frac{g_\epsilon(X^i, X^j)}{n_\epsilon^{(N)}(X^i)\sqrt{\sum_{l=1}^{N} g_\epsilon(X^j, X^l)}}$$

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)

Gain function approximation: Diffusion map-based algorithm

Four Markov operators

- Semigroup $P_t = e^{t\Delta_\rho}$

$$\frac{\partial}{\partial t} P_t f = P_t \Delta_\rho f$$

- Diffusion map T_ϵ

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

- Empirical approximation $T_\epsilon^{(N)}$

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

- $N \times N$ Markov matrix T

$$\mathsf{T}_{ij} := \frac{g_\epsilon(X^i, X^j)}{n_\epsilon^{(N)}(X^i)\sqrt{\sum_{l=1}^{N} g_\epsilon(X^j, X^l)}}$$

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)

Gain function approximation: Diffusion map-based algorithm

Four Markov operators

- Semigroup $P_t = e^{t\Delta_\rho}$

$$\frac{\partial}{\partial t} P_t f = P_t \Delta_\rho f$$

- Diffusion map T_ϵ

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

- Empirical approximation $T_\epsilon^{(N)}$

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

- $N \times N$ Markov matrix T

$$\mathsf{T}_{ij} := \frac{g_\epsilon(X^i, X^j)}{n_\epsilon^{(N)}(X^i)\sqrt{\sum_{l=1}^{N} g_\epsilon(X^j, X^l)}}$$

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)

Gain function approximation: Diffusion map-based algorithm

Approximation steps overview

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

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

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

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

(5) Finite-dim: $$\mathsf{\Phi} = \mathsf{T}\mathsf{\Phi} + \epsilon(\mathsf{h} - \hat{h}_\epsilon^{(N)})$$

$$(1) \Leftrightarrow (2) \overset{\epsilon\downarrow 0}{\approx} (3) \overset{N\uparrow\infty}{\approx} (4) \Leftrightarrow (5)$$

Error analysis:

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

Gain function approximation: Diffusion map-based algorithm

Approximation steps overview

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

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

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

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

(5) Finite-dim: $$\mathsf{\Phi} = \mathsf{T}\mathsf{\Phi} + \epsilon(\mathsf{h} - \hat{h}_\epsilon^{(N)})$$

$$(1) \Leftrightarrow (2) \overset{\epsilon\downarrow 0}{\approx} (3) \overset{N\uparrow\infty}{\approx} (4) \Leftrightarrow (5)$$

Error analysis:

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

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

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

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

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

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

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

Gain function approximation: Error analysis
Bias

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

Proposition

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.

Gain function approximation: Error analysis

Bias

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \sum_{k=0}^{n-1} \epsilon T_\epsilon^k (h - \hat{h}_\epsilon), \quad \epsilon = \frac{t}{n}$$

Proposition

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.

Gain function approximation: Error analysis

Bias

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \sum_{k=0}^{n-1} \epsilon T_\epsilon^k (h - \hat{h}_\epsilon), \quad \epsilon = \frac{t}{n}$$

Proposition

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.

Gain function approximation: Error analysis
Bias

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \sum_{k=0}^{n-1} \epsilon T_\epsilon^k (h - \hat{h}_\epsilon), \quad \epsilon = \frac{t}{n}$$

Proposition

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.

Gain function approximation: Error analysis

Bias

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \sum_{k=0}^{n-1} \epsilon T_\epsilon^k (h - \hat{h}_\epsilon), \quad \epsilon = \frac{t}{n}$$

Proposition

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.

Gain function approximation: Error analysis

Bias

$$\phi = P_t\phi + \int_0^t P_s(h-\hat{h})\mathrm{d}s \quad \text{and} \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \sum_{k=0}^{n-1} \epsilon T_\epsilon^k (h - \hat{h}_\epsilon), \quad \epsilon = \frac{t}{n}$$

Proposition

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.

Gain function approximation: Error analysis
Variance

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

Proposition

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

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

  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)

Gain function approximation: Error analysis
Variance

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

Proposition

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

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

  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)

Gain function approximation: Error analysis

Variance

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

Proposition

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

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

  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)

Gain function approximation: Numerical analysis

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.:

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

Gain function approximation: Numerical analysis

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.:

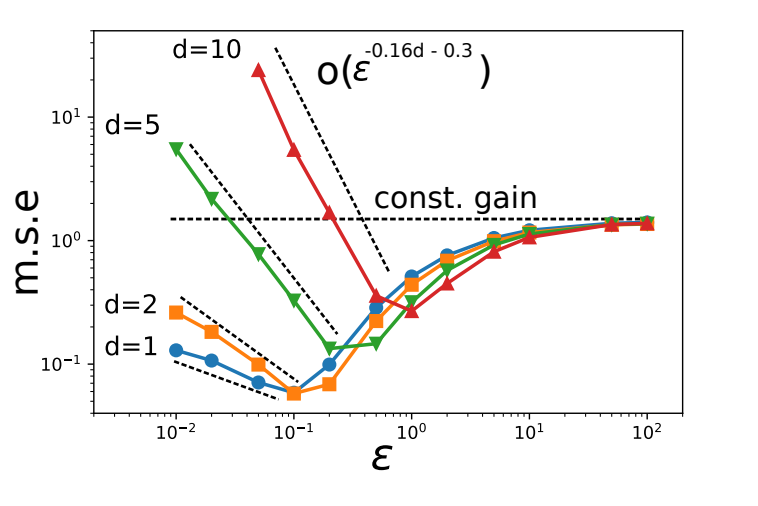
$$\lim_{\epsilon \to \infty} \mathsf{K}_{\epsilon}^{(N)} = \text{const. gain approximatoin}$$

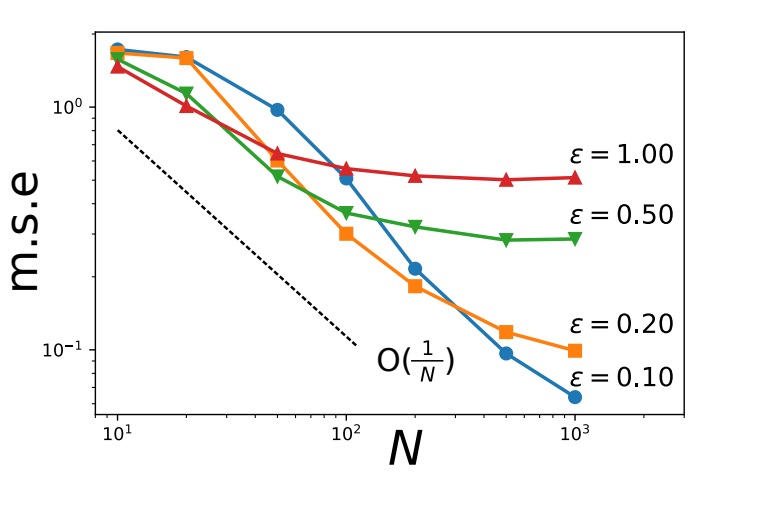
Gain function approximation: Numerical analysis

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.:

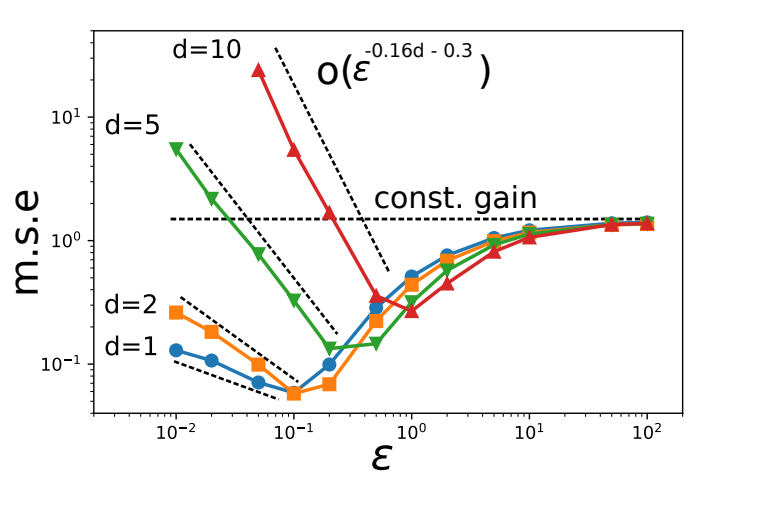
$$\lim_{\epsilon \to \infty} \mathrm{K}_{\epsilon}^{(N)} = \text{const. gain approximatoin}$$

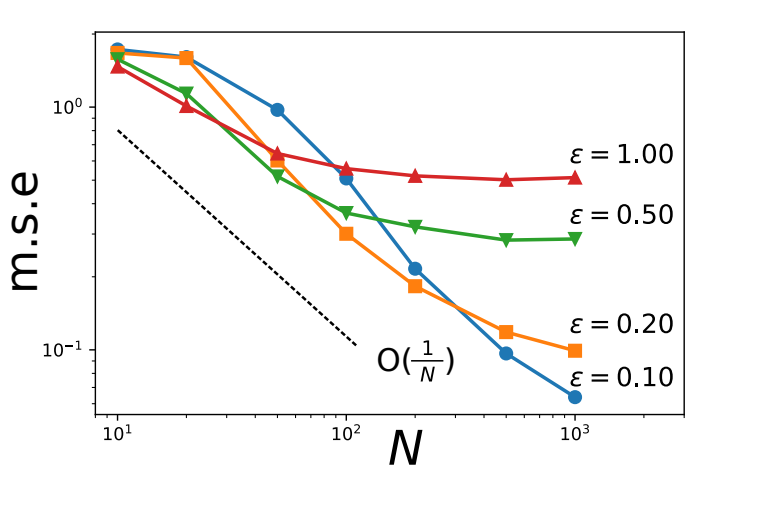
Gain function approximation: Numerical analysis

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.:

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

Outline

Part I: Analysis of the feedback particle filter (FPF)

- Introduction
- Gain function approximation
- Optimal transport FPF
  - A. Taghvaei and P. G. Mehta, An optimal transport formulation of the linear feedback particle filter, in American Control Conference (ACC), 2016, pp. 3614–3619.

    Summary:
  - Developed a principled approach to construct a unique control law
  - Carry out the approach in linear Gaussian setting
- Error analysis of linear FPF

Part II: Accelerated flow for probability distributions

Outline

Part I: Analysis of the feedback particle filter (FPF)

- Introduction
- Gain function approximation
- Optimal transport FPF
- Error analysis of linear FPF
  - A. Taghvaei and P. G. Mehta, Error analysis of the linear feedback particle filter, in American Control Conference (ACC), 2018, pp. 4261-4266
  - A. Taghvaei and P. G. Mehta, Error analysis of the linear feedback particle filter, in Conference on Decision and Control (CDC), 2018, pp. 7194-7199

  Summary:

  - Error analysis of two types of FPF algorithms in linear Gaussian setting
  - Convergence of the empirical distribution to the exact filter given by Kalman filter

Part II: Accelerated flow for probability distributions

Outline

Part I: Analysis of the feedback particle filter (FPF)

- Introduction
- Gain function approximation
- Optimal transport FPF
- Error analysis of linear FPF

Part II: Accelerated flow for probability distributions

- A. Taghvaei and P. G. Mehta, Accelerated flow for probability distributions, International Conference on Machine Learning (ICML), 2019

Part II: Motivation and objective

- Many machine learning problems are modeled as an optimization problem on the space of probability distributions
  - Bayesian inference
  - Learning generative models
  - Policy optimization in reinforcement learning
- This motivates application of optimization algorithms to solve these problems
- **Objective:** Construct accelerated gradient flows for probability distribution

Part II: Motivation and objective

- Many machine learning problems are modeled as an optimization problem on the space of probability distributions
  - Bayesian inference
  - Learning generative models
  - Policy optimization in reinforcement learning
- This motivates application of optimization algorithms to solve these problems
- **Objective:** Construct accelerated gradient flows for probability distribution

Part II: Motivation and objective

- Many machine learning problems are modeled as an optimization problem on the space of probability distributions
  - Bayesian inference
  - Learning generative models
  - Policy optimization in reinforcement learning
- This motivates application of optimization algorithms to solve these problems
- **Objective:** Construct accelerated gradient flows for probability distribution

Part II: Proposed Approach

Euclidean space	Space of probability distributions
Gradient descent	Wasserstein gradient flow
Accelerated methods	?

- (Wibisono, et. al. 2017) proposed a variational formulation to construct accelerated flows on Euclidean space
- Our approach is to extend the variational formulation for probability distributions

Part II: Proposed Approach

Euclidean space	Space of probability distributions
Gradient descent	Wasserstein gradient flow
Accelerated methods	?

- (Wibisono, et. al. 2017) proposed a variational formulation to construct accelerated flows on Euclidean space
- Our approach is to extend the variational formulation for probability distributions

Part II: Proposed Approach

Euclidean space	Space of probability distributions
Gradient descent	Wasserstein gradient flow
Accelerated methods	?

- (Wibisono, et. al. 2017) proposed a variational formulation to construct accelerated flows on Euclidean space
- Our approach is to extend the variational formulation for probability distributions

Part II: Review of variational formulation in Euclidean space

- Optimization problem:

$$\min_{x\in\mathbb{R}^d} \; f(x) \quad \text{(Assume } f \text{ is convex)}$$

- Gradient flow:

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = -\nabla f(x_t)$$

- Accelerated gradient flow (ode limit of Nesterov method) (Su, et. al. 2014):

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \frac{2}{t^3} y_t,$$
$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = -\frac{t^3}{2}\nabla f(x_t)$$

- Variational formulation (Wibinoso, et. al. 2016):

$$\begin{aligned} \text{Minimize} \quad & \int_0^\infty t^3\left(\frac{1}{2}|u_t|^2 - f(x_t)\right)\mathrm{d}t \\ \text{Subject to} \quad & \frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t \end{aligned}$$

Part II: Review of variational formulation in Euclidean space

- Optimization problem:

$$\min_{x\in\mathbb{R}^d} \; f(x) \quad \text{(Assume } f \text{ is convex)}$$

- Gradient flow:

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = -\nabla f(x_t)$$

- Accelerated gradient flow (ode limit of Nesterov method) (Su, et. al. 2014):

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \frac{2}{t^3} y_t,$$
$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = -\frac{t^3}{2}\nabla f(x_t)$$

- Variational formulation (Wibinoso, et. al. 2016):

$$\begin{aligned} \text{Minimize} \quad & \int_0^\infty t^3\left(\frac{1}{2}|u_t|^2 - f(x_t)\right)\mathrm{d}t \\ \text{Subject to} \quad & \frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t \end{aligned}$$

Part II: Review of variational formulation in Euclidean space

- Optimization problem:

$$\min_{x\in\mathbb{R}^d} \; f(x) \quad \text{(Assume } f \text{ is convex)}$$

- Gradient flow:

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = -\nabla f(x_t)$$

- Accelerated gradient flow (ode limit of Nesterov method) (Su, et. al. 2014):

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \frac{2}{t^3} y_t,$$
$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = -\frac{t^3}{2}\nabla f(x_t)$$

- Variational formulation (Wibinoso, et. al. 2016):

$$\begin{aligned} \text{Minimize} \quad & \int_0^\infty t^3\left(\frac{1}{2}|u_t|^2 - f(x_t)\right)\mathrm{d}t \\ \text{Subject to} \quad & \frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t \end{aligned}$$

Part II: Review of variational formulation in Euclidean space

- Optimization problem:

$$\min_{x\in\mathbb{R}^d} \; f(x) \quad \text{(Assume } f \text{ is convex)}$$

- Gradient flow:

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = -\nabla f(x_t)$$

- Accelerated gradient flow (ode limit of Nesterov method) (Su, et. al. 2014):

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \frac{2}{t^3} y_t,$$
$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = -\frac{t^3}{2}\nabla f(x_t)$$

- Variational formulation (Wibinoso, et. al. 2016):

$$\begin{aligned} &\text{Minimize} && \int_0^\infty t^3\left(\frac{1}{2}|u_t|^2 - f(x_t)\right)\mathrm{d}t \\ &\text{Subject to} && \frac{\mathrm{d}x_t}{\mathrm{d}t} = u_t \end{aligned}$$

Part II: Summary

	vector variables $\mathbb{R}^d$	probability distribution $\mathcal{P}_2(\mathbb{R}^d)$
Objective funct.	$f(x)$	?
Gradient flow	$\dot{x}_t = -\nabla f(x_t)$	?
Lagrangian	$t^3(\frac{1}{2}\lvert u_t\rvert^2 - f(x_t))$	?
Accelerated flow	$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \frac{2}{t^3} y_t$ $\frac{\mathrm{d}y_t}{\mathrm{d}t} = -\frac{t^3}{2}\nabla f(x_t)$	?

Part II: Wasserstein gradient flow

- Optimization problem:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \mathsf{F}(\rho) = D(\rho|\rho_\infty)$$

where $\rho_\infty = e^{-f}$.

- Wasserstein gradient flow (Jordan, et. al. 1998):

$$\text{pde form:} \quad \frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t, \quad \text{(Fokker-Planck eq.)}$$

$$\text{probabilistic form:} \quad \mathrm{d}X_t = -\nabla f(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad \text{(Langevin eq.)}$$

where $\rho_t = \mathsf{Law}(X_t)$.

- The goal is to build accelerated version of this sde

Part II: Wasserstein gradient flow

- Optimization problem:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \mathsf{F}(\rho) = D(\rho|\rho_\infty)$$

where $\rho_\infty = e^{-f}$.

- Wasserstein gradient flow (Jordan, et. al. 1998):

pde form: $$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t, \quad \text{(Fokker-Planck eq.)}$$

probabilistic form: $$\mathrm{d}X_t = -\nabla f(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad \text{(Langevin eq.)}$$

where $\rho_t = \mathsf{Law}(X_t)$.

- The goal is to build accelerated version of this sde

Part II: Wasserstein gradient flow

- Optimization problem:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \mathsf{F}(\rho) = D(\rho|\rho_\infty)$$

where $\rho_\infty = e^{-f}$.

- Wasserstein gradient flow (Jordan, et. al. 1998):

pde form: $$\frac{\partial \rho_t}{\partial t} = \nabla \cdot (\rho_t \nabla f) + \Delta \rho_t, \quad \text{(Fokker-Planck eq.)}$$

probabilistic form: $$\mathrm{d}X_t = -\nabla f(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad \text{(Langevin eq.)}$$

where $\rho_t = \mathsf{Law}(X_t)$.

- The goal is to build accelerated version of this sde

Part II: Summary

	vector variables $\mathbb{R}^d$	probability distribution $\mathcal{P}_2(\mathbb{R}^d)$
Objective funct.	$f(x)$	$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty)$
Gradient flow	$\dot{x}_t = -\nabla f(x_t)$	$\mathrm{d}X_t = -\nabla f(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$
Lagrangian	$t^3(\frac{1}{2}\lvert u_t\rvert^2 - f(x_t))$	?
Accelerated flow	$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \frac{2}{t^3}y_t$ $\frac{\mathrm{d}y_t}{\mathrm{d}t} = -\frac{t^3}{2}\nabla f(x_t)$	?

PDE form:

$$\begin{aligned} &\text{Minimize} && \int_0^\infty t^3 \left(\int_{\mathbb{R}^d} \frac{1}{2} |u_t(x)|^2 \rho_t(x) \mathrm{d}x - D(\rho_t | \rho_\infty) \right) \mathrm{d}t \\ &\text{Subject to} && \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u_t) = 0 \end{aligned}$$

Probabilistic form:

$$\begin{aligned} &\text{Minimize} && \mathsf{E}\left[\int_0^\infty t^3 \left(\frac{1}{2} |U_t|^2 - \log\left(\frac{\rho_t(X_t)}{\rho_\infty(X_t)}\right) \right) \mathrm{d}t \right] \\ &\text{Subject to} && \frac{\mathrm{d}X_t}{\mathrm{d}t} = U_t \end{aligned}$$

- It is a mean-field optimal control problem (Bensoussan, et al. 2013, Carmona & Delarue, 2017)

Part II: Variational formulation for probability distributions

PDE form:

$$\begin{aligned} \text{Minimize} \quad & \int_0^\infty t^3 \left(\int_{\mathbb{R}^d} \frac{1}{2} |u_t(x)|^2 \rho_t(x) \mathrm{d}x - D(\rho_t | \rho_\infty) \right) \mathrm{d}t \\ \text{Subject to} \quad & \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u_t) = 0 \end{aligned}$$

Probabilistic form:

$$\begin{aligned} \text{Minimize} \quad & \mathsf{E} \left[\int_0^\infty t^3 \left(\frac{1}{2} |U_t|^2 - \log(\frac{\rho_t(X_t)}{\rho_\infty(X_t)}) \right) \mathrm{d}t \right] \\ \text{Subject to} \quad & \frac{\mathrm{d}X_t}{\mathrm{d}t} = U_t \end{aligned}$$

- It is a mean-field optimal control problem (Bensoussan, et al. 2013, Carmona & Delarue, 2017)

Part II: Variational formulation for probability distributions

PDE form:

$$\begin{aligned} \text{Minimize} \quad & \int_0^\infty t^3 \left(\int_{\mathbb{R}^d} \frac{1}{2} |u_t(x)|^2 \rho_t(x) \mathrm{d}x - D(\rho_t | \rho_\infty) \right) \mathrm{d}t \\ \text{Subject to} \quad & \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u_t) = 0 \end{aligned}$$

Probabilistic form:

$$\begin{aligned} \text{Minimize} \quad & \mathsf{E} \left[\int_0^\infty t^3 \left(\frac{1}{2} |U_t|^2 - \log(\frac{\rho_t(X_t)}{\rho_\infty(X_t)}) \right) \mathrm{d}t \right] \\ \text{Subject to} \quad & \frac{\mathrm{d}X_t}{\mathrm{d}t} = U_t \end{aligned}$$

- It is a mean-field optimal control problem (Bensoussan, et al. 2013, Carmona & Delarue, 2017)

Part II: Main result

Proposition (ICML'19)

1. Accelerated flow (probabilistic form):

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \frac{2}{t^3} Y_t$$
$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = -\frac{t^3}{2}(\nabla f(X_t) + \nabla \log(\rho_t(X_t)))$$

   where $\rho_t = \mathsf{Law}(X_t)$

2. (Convergence) If $\mathsf{F}(\rho) = D(\rho|\rho_\infty)$ is displacement convex, and the dimension $d = 1$. Then

$$D(\rho_t|\rho_\infty) \leq O(\frac{1}{t^2})$$

3. If ρ_∞ is Gaussian, then X_t is also Gaussian and the mean evolves according to the accelerated gradient flow in Euclidean space

Part II: Main result

Proposition (ICML'19)

1. Accelerated flow (probabilistic form):

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \frac{2}{t^3} Y_t$$
$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = -\frac{t^3}{2}(\nabla f(X_t) + \nabla \log(\rho_t(X_t)))$$

   where $\rho_t = \mathsf{Law}(X_t)$

2. (Convergence) If $\mathsf{F}(\rho) = D(\rho|\rho_\infty)$ is displacement convex, and the dimension $d = 1$. Then

$$D(\rho_t|\rho_\infty) \leq O(\frac{1}{t^2})$$

3. If ρ_∞ is Gaussian, then X_t is also Gaussian and the mean evolves according to the accelerated gradient flow in Euclidean space

Part II: Main result

Proposition (ICML'19)

1. Accelerated flow (probabilistic form):

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \frac{2}{t^3} Y_t$$
$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = -\frac{t^3}{2}(\nabla f(X_t) + \nabla \log(\rho_t(X_t)))$$

   where $\rho_t = \mathsf{Law}(X_t)$

2. (Convergence) If $\mathsf{F}(\rho) = D(\rho|\rho_\infty)$ is displacement convex, and the dimension $d = 1$. Then

$$D(\rho_t|\rho_\infty) \leq O(\frac{1}{t^2})$$

3. If ρ_∞ is Gaussian, then X_t is also Gaussian and the mean evolves according to the accelerated gradient flow in Euclidean space

Part II: Summary

	vector variables $\mathbb{R}^d$	probability distribution $\mathcal{P}_2(\mathbb{R}^d)$
Objective funct.	$f(x)$	$\mathsf{F}(\rho) = \mathsf{D}(\rho \| e^{-f})$
Gradient flow	$\dot{x}_t = -\nabla f(x_t)$	$\mathrm{d}X_t = -\nabla f(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$
Lagrangian	$t^3(\frac{1}{2}\lvert u_t\rvert^2 - f(x_t))$	$\mathsf{E}[t^3(\frac{1}{2}\lvert U_t\rvert^2 - f(X_t) - \log(\rho(X_t)))]$
Accelerated flow	$\frac{\mathrm{d}x_t}{\mathrm{d}t} = \frac{2}{t^3}y_t$ $\frac{\mathrm{d}y_t}{\mathrm{d}t} = -\frac{t^3}{2}\nabla f(x_t)$	$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \frac{2}{t^3}Y_t$ $\frac{\mathrm{d}Y_t}{\mathrm{d}t} = -\frac{t^3}{2}(\nabla f(X_t) + \nabla \log(\rho_t(X_t)))$

Procedure: Simulate N particles $\{(X_t^1, Y_t^1), \ldots, (X_t^N, Y_t^N)\}$

$$\frac{\mathrm{d}X_t^i}{\mathrm{d}t} = \frac{2}{t^3} Y_t^i, \quad X_0^i \sim \rho_0$$
$$\frac{\mathrm{d}Y_t^i}{\mathrm{d}t} = -\frac{t^3}{2}(\nabla f(X_t^i) + \underbrace{\nabla \log(\rho_t(X_t^i))}_{\text{interaction term}})$$

Approximation: Interaction term is approximated with particles

- (parametric) Gaussian approximation

$$\nabla \log(\rho(x)) \approx -\Sigma^{-1}(x - m), \quad m, \Sigma = \text{empirical mean and covariance}$$

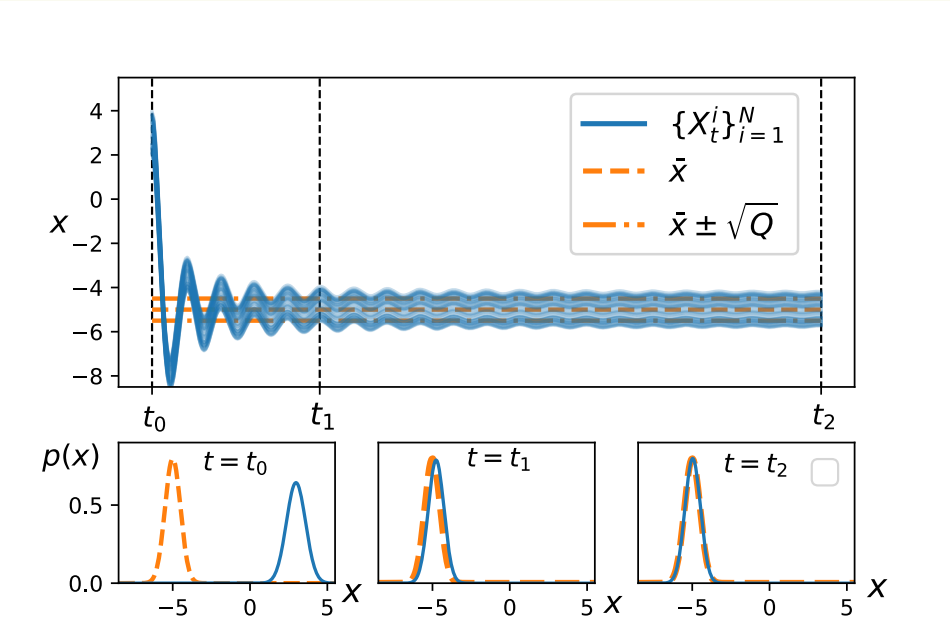
- (non-parametric) Diffusion-map approximation

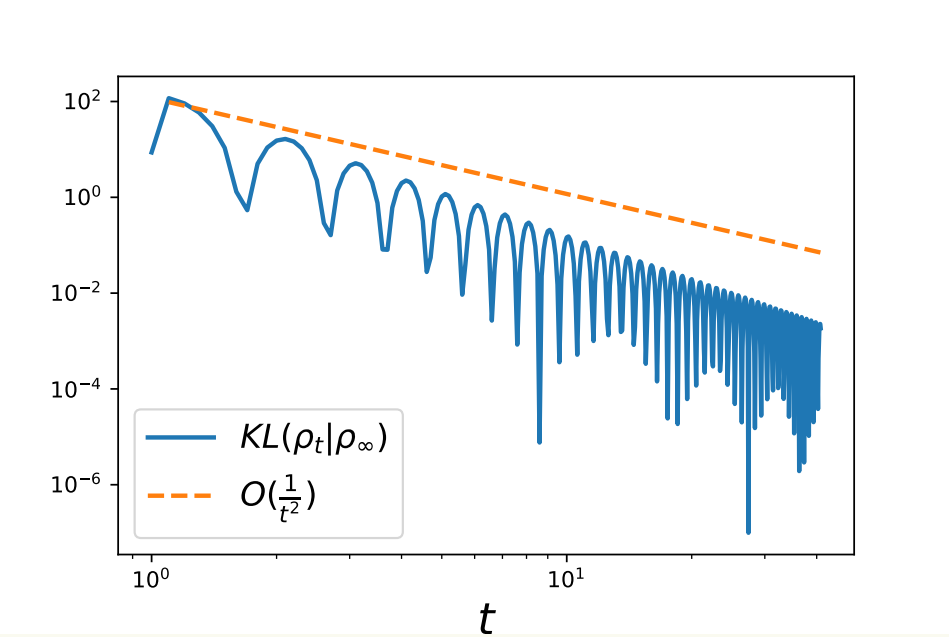
$$\nabla \log(\rho(x)) = \Delta_\rho(x) \approx -\frac{1}{\epsilon} \frac{\sum_{i=1}^N k_\epsilon(x, X^i)(x - X^i)}{\sum_{i=1}^N k_\epsilon(x, X^i)}$$

where $k_\epsilon(\cdot, \cdot)$ is the <u>diffusion-map kernel</u>

Part II: Numerical example

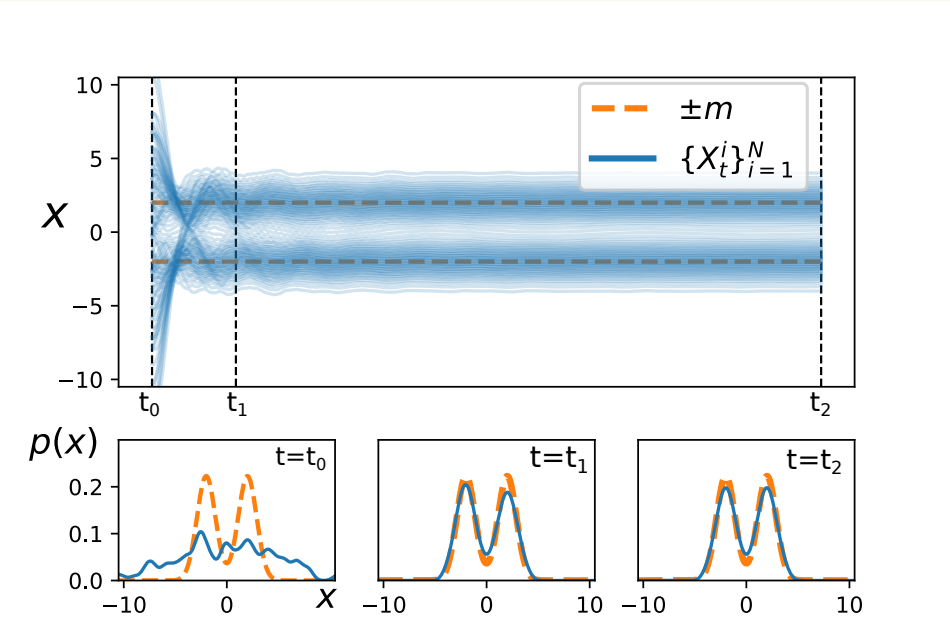
- The target distribution is Gaussian
- The target distribution is mixture of two Gaussians

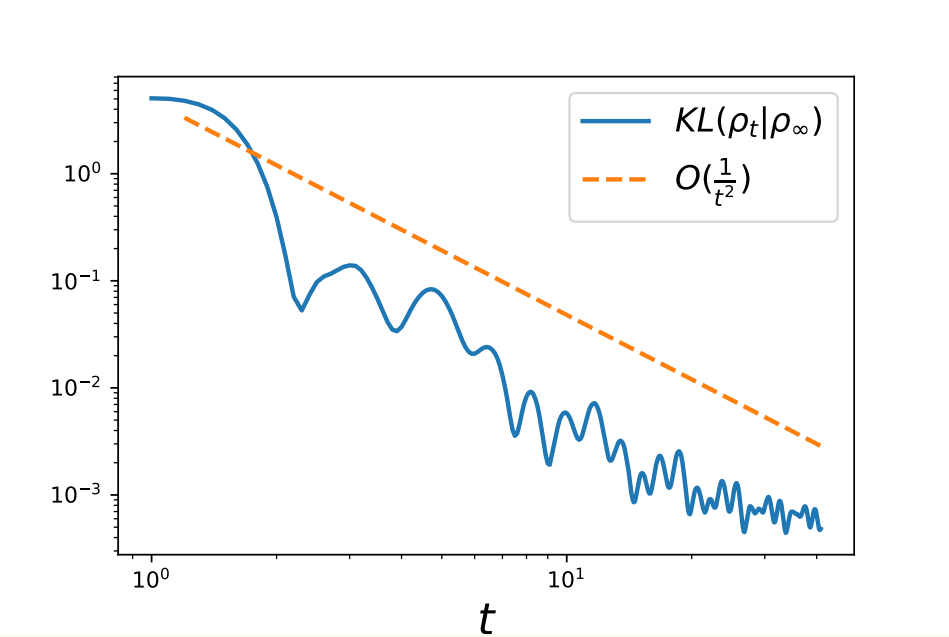




Part II: Numerical example

- The target distribution is Gaussian
- The target distribution is mixture of two Gaussians





Summary and future work

Part I: Analysis of FPF

- Gain function approximation: Diffusion map-based algorithm and its error analysis
- Optimal transport FPF: An optimal transportation-based approach to construct FPF control law
- Error analysis of linear FPF: Convergence in linear Gaussian setting

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

Part II: Accelerated flow for probability distributions

Variational problem: $$\int_0^\infty t^3 \left(\int_{\mathbb{R}^d} \frac{1}{2} |u_t(x)|^2 \rho_t(x) \mathrm{d}x - D(\rho_t | \rho_\infty) \right) \mathrm{d}t$$

Accelerated flow: $$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \ldots, \quad \frac{\mathrm{d}Y_t}{\mathrm{d}t} = \ldots + \nabla \log(\rho_t(X_t))$$

Future work: Obtain approximation from variational formulation

Summary and future work

Part I: Analysis of FPF

- Gain function approximation: Diffusion map-based algorithm and its error analysis
- Optimal transport FPF: An optimal transportation-based approach to construct FPF control law
- Error analysis of linear FPF: Convergence in linear Gaussian setting

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

Part II: Accelerated flow for probability distributions

Variational problem: $$\int_0^\infty t^3 \left(\int_{\mathbb{R}^d} \frac{1}{2} |u_t(x)|^2 \rho_t(x) \mathrm{d}x - D(\rho_t | \rho_\infty) \right) \mathrm{d}t$$

Accelerated flow: $$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \ldots, \quad \frac{\mathrm{d}Y_t}{\mathrm{d}t} = \ldots + \nabla \log(\rho_t(X_t))$$

Future work: Obtain approximation from variational formulation

Summary and future work

Part I: Analysis of FPF

- Gain function approximation: Diffusion map-based algorithm and its error analysis
- Optimal transport FPF: An optimal transportation-based approach to construct FPF control law
- Error analysis of linear FPF: Convergence in linear Gaussian setting

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

Part II: Accelerated flow for probability distributions

Variational problem: $$\int_0^\infty t^3 \left(\int_{\mathbb{R}^d} \frac{1}{2} |u_t(x)|^2 \rho_t(x) \mathrm{d}x - D(\rho_t | \rho_\infty) \right) \mathrm{d}t$$

Accelerated flow: $$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \ldots, \quad \frac{\mathrm{d}Y_t}{\mathrm{d}t} = \ldots + \nabla \log(\rho_t(X_t))$$

Future work: Obtain approximation from variational formulation

Summary and future work

Part I: Analysis of FPF

- Gain function approximation: Diffusion map-based algorithm and its error analysis
- Optimal transport FPF: An optimal transportation-based approach to construct FPF control law
- Error analysis of linear FPF: Convergence in linear Gaussian setting

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

Part II: Accelerated flow for probability distributions

Variational problem: $$\int_0^\infty t^3 \left(\int_{\mathbb{R}^d} \frac{1}{2} |u_t(x)|^2 \rho_t(x) \mathrm{d}x - D(\rho_t | \rho_\infty) \right) \mathrm{d}t$$

Accelerated flow: $$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \ldots, \quad \frac{\mathrm{d}Y_t}{\mathrm{d}t} = \ldots + \nabla \log(\rho_t(X_t))$$

Future work: Obtain approximation from variational formulation

Acknowledgements

I like to thank:

- Committee members
- NSF, ME, Math, ECE, CSE, CSL
- Collaborators: S. P. Meyn, S. Reich, C. Zhang, J. Kim, J. de Wiljes
- My PhD advisor: Prashant Mehta
- My family and friends

Thanks for your attention!

Acknowledgements

I like to thank:

- Committee members
- NSF, ME, Math, ECE, CSE, CSL
- Collaborators: S. P. Meyn, S. Reich, C. Zhang, J. Kim, J. de Wiljes
- My PhD advisor: Prashant Mehta
- My family and friends

Thanks for your attention!

Acknowledgements

I like to thank:

- Committee members
- NSF, ME, Math, ECE, CSE, CSL
- Collaborators: S. P. Meyn, S. Reich, C. Zhang, J. Kim, J. de Wiljes
- My PhD advisor: Prashant Mehta
- My family and friends

Thanks for your attention!

Acknowledgements

I like to thank:

- Committee members
- NSF, ME, Math, ECE, CSE, CSL
- Collaborators: S. P. Meyn, S. Reich, C. Zhang, J. Kim, J. de Wiljes
- My PhD advisor: Prashant Mehta
- My family and friends

Thanks for your attention!

Acknowledgements

I like to thank:

- Committee members
- NSF, ME, Math, ECE, CSE, CSL
- Collaborators: S. P. Meyn, S. Reich, C. Zhang, J. Kim, J. de Wiljes
- My PhD advisor: Prashant Mehta
- My family and friends

Thanks for your attention!

Acknowledgements

I like to thank:

- Committee members
- NSF, ME, Math, ECE, CSE, CSL
- Collaborators: S. P. Meyn, S. Reich, C. Zhang, J. Kim, J. de Wiljes
- My PhD advisor: Prashant Mehta
- My family and friends

Thanks for your attention!

Acknowledgements

I like to thank:

- Committee members
- NSF, ME, Math, ECE, CSE, CSL
- Collaborators: S. P. Meyn, S. Reich, C. Zhang, J. Kim, J. de Wiljes
- My PhD advisor: Prashant Mehta
- My family and friends

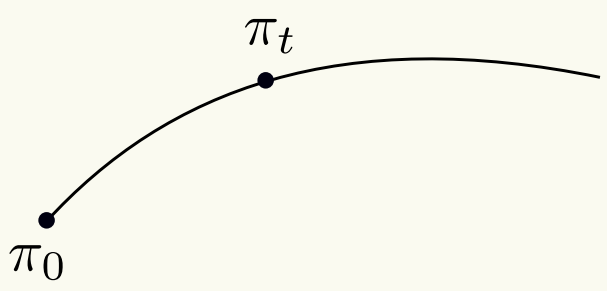
Thanks for your attention!

Optimal transport FPF
Problem overview

Control viewpoint:

$$\mathrm{d}\bar{X}_t = \text{(control law)}$$

Objective: Find the control law s.t $\bar{X}_t$ follows the posterior π_t



trajectory of the posterior distribution

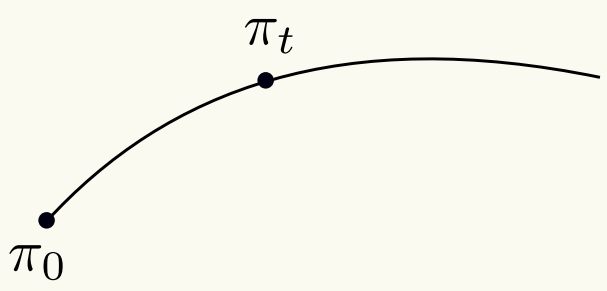
Non-uniqueness: There are infinitely many solutions

Optimal transport FPF
Problem overview

Control viewpoint:

$$\mathrm{d}\bar{X}_t = \text{(control law)}$$

Objective: Find the control law s.t $\bar{X}_t$ follows the posterior π_t



trajectory of the posterior distribution

Non-uniqueness: There are infinitely many solutions

Optimal transport FPF

Non-uniqueness example

Example:

State process: $\mathrm{d}X_t = \mathrm{d}B_t, \quad X_0 \sim N(0,1)$

Objective: compute $\pi_t(\cdot) = \mathsf{P}(X_t \in \cdot)$

Two solutions:

(I) $\mathrm{d}X_t^i = \mathrm{d}B_t^i$

(II) $\frac{\mathrm{d}}{\mathrm{d}t} X_t^i = \frac{X_t^i}{\frac{2}{N}\sum_{j=1}^N (X_t^i)^2}$

They both produce the same distribution $N(0, 1+t)$.

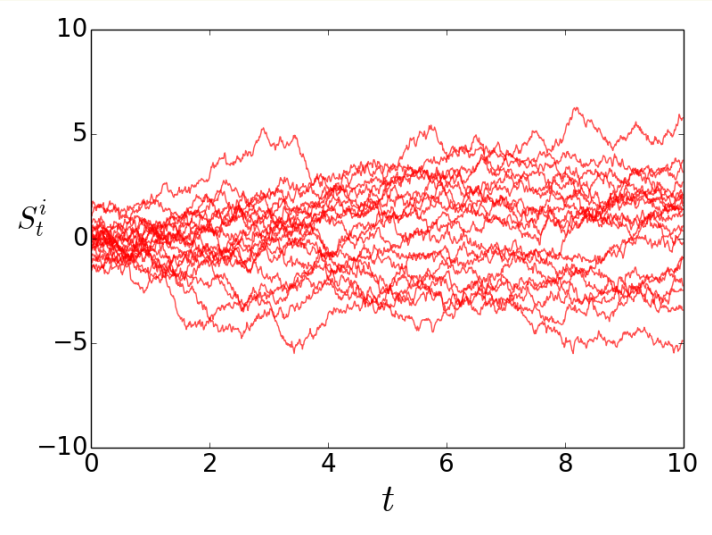
Example:

State process: $\mathrm{d}X_t = \mathrm{d}B_t, \quad X_0 \sim N(0,1)$

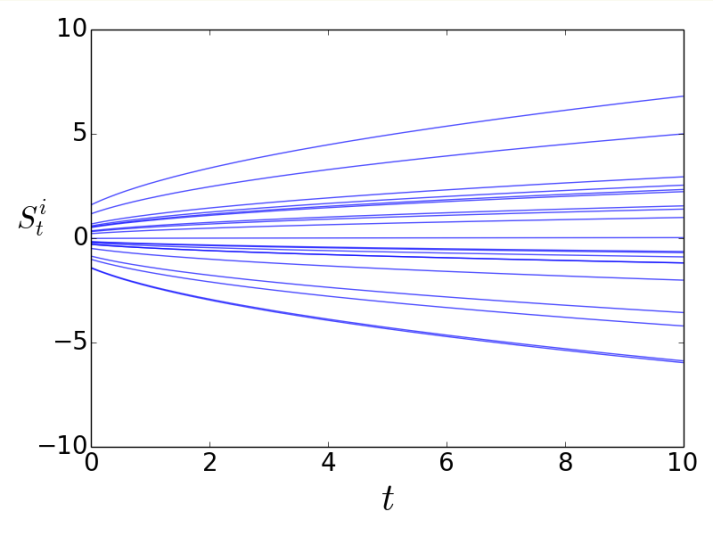
Objective: compute $\pi_t(\cdot) = \mathsf{P}(X_t \in \cdot)$

Two solutions:

(I) $\mathrm{d}X_t^i = \mathrm{d}B_t^i$



(II) $\dfrac{\mathrm{d}}{\mathrm{d}t} X_t^i = \dfrac{X_t^i}{\frac{2}{N}\sum_{j=1}^N (X_t^i)^2}$



They both produce the same distribution $N(0, 1+t)$.

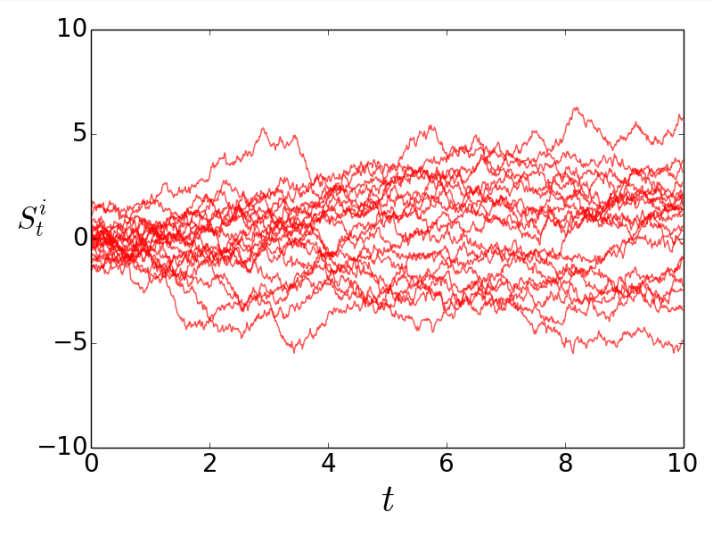
Example:

State process: $\mathrm{d}X_t = \mathrm{d}B_t, \quad X_0 \sim N(0,1)$

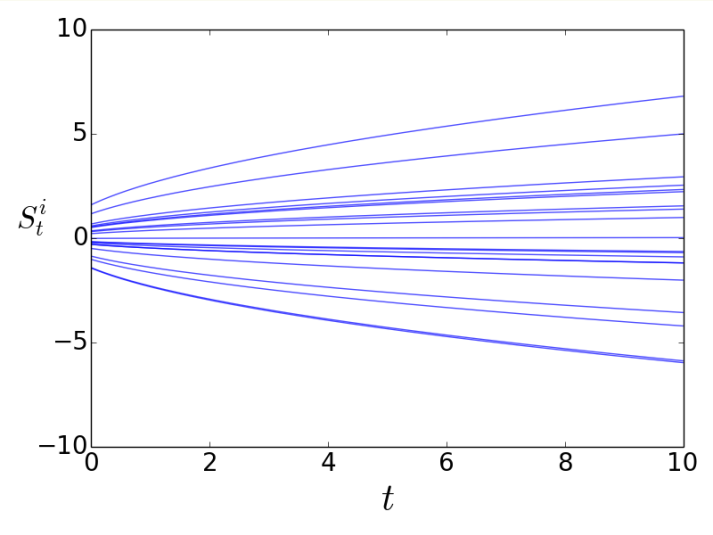
Objective: compute $\pi_t(\cdot) = \mathsf{P}(X_t \in \cdot)$

Two solutions:

(I) $\mathrm{d}X_t^i = \mathrm{d}B_t^i$



(II) $\dfrac{\mathrm{d}}{\mathrm{d}t} X_t^i = \dfrac{X_t^i}{\frac{2}{N}\sum_{j=1}^N (X_t^i)^2}$



They both produce the same distribution $N(0, 1+t)$.

Optimal transport FPF

Optimal transportation approach

- Reason for non-uniqueness: Only the marginal distributions, at each time instant, are specified
- Optimal transport maps provide a way to uniquely couple two distributions.

Proposed solution: Infinitesimal optimal transport maps

$$\bar{X}_{t+\Delta t} = T_t(\bar{X}_t),$$

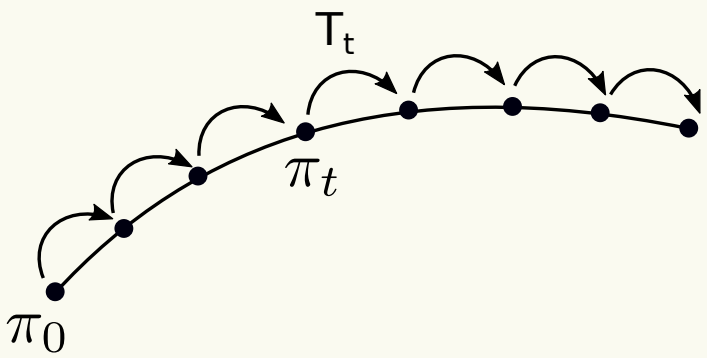
- T_t is the optimal transport map between π_t and $\pi_{t+\Delta t}$
- Take the limit as $\Delta t \to 0$

Optimal transport FPF

Optimal transportation approach

- Reason for non-uniqueness: Only the marginal distributions, at each time instant, are specified
- Optimal transport maps provide a way to uniquely couple two distributions.

Proposed solution: Infinitesimal optimal transport maps



$$\bar{X}_{t+\Delta t} = T_t(\bar{X}_t),$$

- T_t is the optimal transport map between π_t and $\pi_{t+\Delta t}$
- Take the limit as $\Delta t \to 0$

Optimal transport FPF

Linear Gaussian setting

The procedure is carried out in linear Gaussian setting.

$$\textbf{FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \mathrm{d}\bar{B}_t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t),$$

$$\textbf{Opt. FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \frac{1}{2}\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t) + \Omega_t\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t,$$

- The process noise term is replaced with a deterministic term
- Ω_t is the (skew symmetric) solution to the matrix equation:

$$\Omega_t\bar{\Sigma}_t^{-1} + \bar{\Sigma}_t^{-1}\Omega_t = A^T - A + \frac{1}{2}(\bar{\mathsf{K}}_t H - H^T\bar{\mathsf{K}}_t^T)$$

Extension: Developed an approach that does not require invertible covariance matrix

A. Taghvaei, P. G. Mehta, An optimal transport formulation for the linear feedback particle filter, (ACC) 2016

Optimal transport FPF
Linear Gaussian setting

The procedure is carried out in linear Gaussian setting.

$$\textbf{FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \mathrm{d}\bar{B}_t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t),$$

$$\textbf{Opt. FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \frac{1}{2}\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t) + \Omega_t\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t,$$

- The process noise term is replaced with a deterministic term
- Ω_t is the (skew symmetric) solution to the matrix equation:

$$\Omega_t\bar{\Sigma}_t^{-1} + \bar{\Sigma}_t^{-1}\Omega_t = A^T - A + \frac{1}{2}(\bar{\mathsf{K}}_t H - H^T\bar{\mathsf{K}}_t^T)$$

Extension: Developed an approach that does not require invertible covariance matrix

A. Taghvaei, P. G. Mehta, An optimal transport formulation for the linear feedback particle filter, (ACC) 2016

Optimal transport FPF
Linear Gaussian setting

The procedure is carried out in linear Gaussian setting.

$$\textbf{FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \mathrm{d}\bar{B}_t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t),$$

$$\textbf{Opt. FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \frac{1}{2}\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t) + \Omega_t\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t,$$

- The process noise term is replaced with a deterministic term
- Ω_t is the (skew symmetric) solution to the matrix equation:

$$\Omega_t\bar{\Sigma}_t^{-1} + \bar{\Sigma}_t^{-1}\Omega_t = A^T - A + \frac{1}{2}(\bar{\mathsf{K}}_t H - H^T\bar{\mathsf{K}}_t^T)$$

Extension: Developed an approach that does not require invertible covariance matrix

A. Taghvaei, P. G. Mehta, An optimal transport formulation for the linear feedback particle filter, (ACC) 2016

Optimal transport FPF
Linear Gaussian setting

The procedure is carried out in linear Gaussian setting.

$$\textbf{FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \mathrm{d}\bar{B}_t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t),$$

$$\textbf{Opt. FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \frac{1}{2}\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t) + \Omega_t\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t,$$

- The process noise term is replaced with a deterministic term
- Ω_t is the (skew symmetric) solution to the matrix equation:

$$\Omega_t\bar{\Sigma}_t^{-1} + \bar{\Sigma}_t^{-1}\Omega_t = A^T - A + \frac{1}{2}(\bar{\mathsf{K}}_t H - H^T\bar{\mathsf{K}}_t^T)$$

Extension: Developed an approach that does not require invertible covariance matrix

A. Taghvaei, P. G. Mehta, An optimal transport formulation for the linear feedback particle filter, (ACC) 2016

Optimal transport FPF
Linear Gaussian setting

The procedure is carried out in linear Gaussian setting.

$$\textbf{FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \mathrm{d}\bar{B}_t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t),$$

$$\textbf{Opt. FPF:} \quad \mathrm{d}\bar{X}_t = A\bar{X}_t\mathrm{d}t + \frac{1}{2}\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t + \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{H\bar{X}_t + H\bar{m}_t}{2}\mathrm{d}t) + \Omega_t\bar{\Sigma}_t^{-1}(\bar{X}_t - \bar{m}_t)\mathrm{d}t,$$

- The process noise term is replaced with a deterministic term
- Ω_t is the (skew symmetric) solution to the matrix equation:

$$\Omega_t\bar{\Sigma}_t^{-1} + \bar{\Sigma}_t^{-1}\Omega_t = A^T - A + \frac{1}{2}(\bar{\mathsf{K}}_t H - H^T\bar{\mathsf{K}}_t^T)$$

Extension: Developed an approach that does not require invertible covariance matrix

A. Taghvaei, P. G. Mehta, An optimal transport formulation for the linear feedback particle filter, (ACC) 2016

Linear Gaussian setting:

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

Kalman-Bucy filter: $P(X_t|Z_{[0,t]})$ is Gaussian $\mathcal{N}(m_t, \Sigma_t)$

Update for mean: $\mathrm{d}m_t = \underbrace{Am_t\mathrm{d}t}_{\text{dynamics}} + \underbrace{\mathsf{K}_t(\mathrm{d}Z_t - Hm\mathrm{d}t)}_{\text{correction}}$

Update for covariance: $\dfrac{\mathrm{d}\Sigma_t}{\mathrm{d}t} =$ (Ricatti equation)

Kalman gain: $\mathsf{K}_t := \Sigma_t H^\top$

R. E Kalman and R. S Bucy. New results in linear filtering and prediction theory, 1961

Error analysis of linear FPF

Linear Gaussian filtering problem: Kalman-Bucy Filter

Linear Gaussian setting:

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

Kalman-Bucy filter: $P(X_t|Z_{[0,t]})$ is Gaussian $\mathcal{N}(m_t, \Sigma_t)$

Update for mean: $\mathrm{d}m_t = \underbrace{Am_t\mathrm{d}t}_{\text{dynamics}} + \underbrace{\mathsf{K}_t(\mathrm{d}Z_t - Hm\mathrm{d}t)}_{\text{correction}}$

Update for covariance: $\dfrac{\mathrm{d}\Sigma_t}{\mathrm{d}t} =$ (Ricatti equation)

Kalman gain: $\mathsf{K}_t := \Sigma_t H^\top$

R. E Kalman and R. S Bucy. New results in linear filtering and prediction theory, 1961

Error analysis of linear FPF
Problem formulation

Stochastic linear FPF:

$$\mathrm{d}X_t^i = AX_t^i + \sigma_B \mathrm{d}B_t^i + \Sigma_t^{(N)} H^\top (\mathrm{d}Z_t - \frac{1}{2}H(X_t^i + m_t^{(N)})\mathrm{d}t), \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

Deterministic linear FPF:

$$\begin{aligned}\mathrm{d}X_t^i =& AX_t^i + \frac{1}{2}\Sigma_B(\Sigma^{(N)})^{-1}(X_t^i - m_t^{(N)})\mathrm{d}t \\ &+ \Sigma_t^{(N)} H^\top (\mathrm{d}Z_t - \frac{1}{2}H(X_t^i + m_t^{(N)})\mathrm{d}t), \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0\end{aligned}$$

where $m_t^{(N)}, \Sigma_t^{(N)}$ are empirical mean and covariance

Exactness: If $N = \infty$ (mean-field approximation), then $m_t^{(\infty)} = m_t$ and $\Sigma_t^{(\infty)} = \Sigma_t$

Question: What is the approximation error when $N < \infty$?

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model . . . 1994.
K. Bergemann and S. Reich. An ensemble Kalman-Bucy filter for continuous data assimilation, 2012
T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter, *Automatica*, 2016

Error analysis of linear FPF
Problem formulation

Stochastic linear FPF:

$$\mathrm{d}X_t^i = AX_t^i + \sigma_B \mathrm{d}B_t^i + \Sigma_t^{(N)} H^\top (\mathrm{d}Z_t - \frac{1}{2}H(X_t^i + m_t^{(N)})\mathrm{d}t), \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

Deterministic linear FPF:

$$\begin{aligned}\mathrm{d}X_t^i =& AX_t^i + \frac{1}{2}\Sigma_B(\Sigma^{(N)})^{-1}(X_t^i - m_t^{(N)})\mathrm{d}t \\ &+ \Sigma_t^{(N)} H^\top (\mathrm{d}Z_t - \frac{1}{2}H(X_t^i + m_t^{(N)})\mathrm{d}t), \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0\end{aligned}$$

where $m_t^{(N)}, \Sigma_t^{(N)}$ are empirical mean and covariance

Exactness: If $N = \infty$ (mean-field approximation), then $m_t^{(\infty)} = m_t$ and $\Sigma_t^{(\infty)} = \Sigma_t$

Question: What is the approximation error when $N < \infty$?

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model . . . 1994.
K. Bergemann and S. Reich. An ensemble Kalman-Bucy filter for continuous data assimilation, 2012
T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter, *Automatica*, 2016

Error analysis of linear FPF

Result: Deterministic linear FPF

Evolution of mean and covariance:

$$\mathrm{d}m_t^{(N)} = \underbrace{Am_t^{(N)}\mathrm{d}t + \mathsf{K}_t^{(N)}(\mathrm{d}Z_t - Hm_t^{(N)}\mathrm{d}t)}_{\text{Kalman filter}}$$

$$\mathrm{d}\Sigma_t^{(N)} = \underbrace{(A\Sigma_t^{(N)} + \Sigma_t^{(N)}A^\top + \sigma_B\sigma_B^\top - \Sigma_t^{(N)}H^\top H\Sigma_t^{(N)})\mathrm{d}t}_{\text{Kalman filter}}$$

Assumptions: (i) System is detectable and stabilizable; (ii) $\Sigma_0^{(N)}$ is invertible

Proposition

- Convergence of mean and variance: $\exists \lambda_0 > 0$ s.t

$$\mathsf{E}[|m_t^{(N)} - m_t|^2] \le (\text{const.})\frac{e^{-2\lambda_0 t}}{N}, \quad \mathsf{E}[\|\Sigma_t^{(N)} - \Sigma_t\|_F^2] \le (\text{const.})\frac{e^{-4\lambda_0 t}}{N}$$

- Convergence of the empirical distribution:

$$\mathsf{E}\left[\left|\frac{1}{N}\sum_{i=1}^N f(X_t^i) - \pi_t(f)\right|^2\right] \le \frac{(\text{const})}{N}$$

Evolution of mean and covariance:

$$\mathrm{d}m_t^{(N)} = \underbrace{Am_t^{(N)}\mathrm{d}t + \mathsf{K}_t^{(N)}(\mathrm{d}Z_t - Hm_t^{(N)}\mathrm{d}t)}_{\text{Kalman filter}}$$

$$\mathrm{d}\Sigma_t^{(N)} = \underbrace{(A\Sigma_t^{(N)} + \Sigma_t^{(N)}A^\top + \sigma_B\sigma_B^\top - \Sigma_t^{(N)}H^\top H\Sigma_t^{(N)})\mathrm{d}t}_{\text{Kalman filter}}$$

Assumptions: (i) System is detectable and stabilizable; (ii) $\Sigma_0^{(N)}$ is invertible

Proposition

- Convergence of mean and variance: $\exists \lambda_0 > 0$ s.t

$$\mathsf{E}[|m_t^{(N)} - m_t|^2] \le (\text{const.})\frac{e^{-2\lambda_0 t}}{N}, \quad \mathsf{E}[\|\Sigma_t^{(N)} - \Sigma_t\|_F^2] \le (\text{const.})\frac{e^{-4\lambda_0 t}}{N}$$

- Convergence of the empirical distribution:

$$\mathsf{E}\left[\left|\frac{1}{N}\sum_{i=1}^N f(X_t^i) - \pi_t(f)\right|^2\right] \le \frac{(\text{const})}{N}$$

Error analysis of linear FPF

Result: Stochastic linear FPF

Evolution of mean and covariance:

$$\mathrm{d}m_t^{(N)} = \underbrace{Am_t^{(N)}\mathrm{d}t + \mathsf{K}_t^{(N)}(\mathrm{d}Z_t - Hm_t^{(N)}\mathrm{d}t)}_{\text{Kalman filter}} + \underbrace{\frac{\sigma_B}{\sqrt{N}}\mathrm{d}\tilde{B}_t}_{\text{stochastic term}}$$

$$\mathrm{d}\Sigma_t^{(N)} = \underbrace{(A\Sigma_t^{(N)} + \Sigma_t^{(N)}A^\top + \sigma_B\sigma_B^\top - \Sigma_t^{(N)}H^\top H\Sigma_t^{(N)})\mathrm{d}t}_{\text{Kalman filter}} + \underbrace{\frac{\mathrm{d}M_t}{\sqrt{N}}}_{\text{stochastic term}}$$

Assumption: The system is stable and the observation matrix is full rank.

Proposition

- Convergence of empirical mean and covariance:

$$\mathsf{E}[|m_t - m_t^{(N)}|^2] \leq \frac{(\text{const.})}{N}, \quad \mathsf{E}[|\Sigma_t - \Sigma_t^{(N)}|^2] \leq \frac{(\text{const.})}{N}$$

- Convergence of empirical distribution: ...

P Del Moral, J Tugaut. On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters, 2018

Error analysis of linear FPF

Result: Stochastic linear FPF

Evolution of mean and covariance:

$$\mathrm{d}m_t^{(N)} = \underbrace{Am_t^{(N)}\mathrm{d}t + \mathsf{K}_t^{(N)}(\mathrm{d}Z_t - Hm_t^{(N)}\mathrm{d}t)}_{\text{Kalman filter}} + \underbrace{\frac{\sigma_B}{\sqrt{N}}\mathrm{d}\tilde{B}_t}_{\text{stochastic term}}$$

$$\mathrm{d}\Sigma_t^{(N)} = \underbrace{(A\Sigma_t^{(N)} + \Sigma_t^{(N)}A^\top + \sigma_B\sigma_B^\top - \Sigma_t^{(N)}H^\top H\Sigma_t^{(N)})\mathrm{d}t}_{\text{Kalman filter}} + \underbrace{\frac{\mathrm{d}M_t}{\sqrt{N}}}_{\text{stochastic term}}$$

Assumption: The system is stable and the observation matrix is full rank.

Proposition

- Convergence of empirical mean and covariance:

$$\mathsf{E}[|m_t - m_t^{(N)}|^2] \leq \frac{(\text{const.})}{N}, \quad \mathsf{E}[|\Sigma_t - \Sigma_t^{(N)}|^2] \leq \frac{(\text{const.})}{N}$$

- Convergence of empirical distribution: ...

P Del Moral, J Tugaut. On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters, 2018

PDE viewpoint:

$$\langle \nabla\phi, \nabla\psi \rangle = \langle \psi, h - \hat{h} \rangle$$

- T. Yang, et. al. Automatica, 2015. (Constant gain and Galerkin 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)

Stochastic viewpoint:

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

- A. Taghvaei, P. G. Mehta, CDC, 2016. (Diffusion map algorithm)

Variational viewpoint:

$$\min_{\phi} \mathsf{E}[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)h(X)]$$

- A. Radhakrishnan, A. Devraj, and S. Meyn. 2017. (RKHS method)

PDE viewpoint:

$$\langle \nabla\phi, \nabla\psi \rangle = \langle \psi, h - \hat{h} \rangle$$

- T. Yang, et. al. Automatica, 2015. (Constant gain and Galerkin 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)

Stochastic viewpoint:

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

- A. Taghvaei, P. G. Mehta, CDC, 2016. (Diffusion map algorithm)

Variational viewpoint:

$$\min_{\phi} \mathsf{E}[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)h(X)]$$

- A. Radhakrishnan, A. Devraj, and S. Meyn. 2017. (RKHS method)

PDE viewpoint:

$$\langle \nabla\phi, \nabla\psi \rangle = \langle \psi, h - \hat{h} \rangle$$

- T. Yang, et. al. Automatica, 2015. (Constant gain and Galerkin 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)

Stochastic viewpoint:

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

- A. Taghvaei, P. G. Mehta, CDC, 2016. (Diffusion map algorithm)

Variational viewpoint:

$$\min_{\phi} \mathsf{E}[\frac{1}{2}|\nabla\phi(X)|^2 - \phi(X)h(X)]$$

- A. Radhakrishnan, A. Devraj, and S. Meyn. 2017. (RKHS method)

Diffusion map approximation
Mathematical preliminaries

- **Weighted Laplacian:** $\Delta_\rho f := \frac{1}{\rho}\nabla \cdot (\rho \nabla f)$
- **Spectral representation:** $-\Delta_\rho f = \sum_{m=0}^{\infty} \lambda_m \langle \psi_m, f \rangle \psi_m$

$$\text{Eigenvalues:} \quad \lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \dots$$
$$\text{Eigenfunctions:} \quad \psi_0 = 1, \psi_1, \psi_2, , \dots$$

- **Diffusion semigroup:** $e^{t\Delta_\rho} f(x) = \sum_{m=0}^{\infty} e^{-t\lambda_m} \langle \psi_m, f \rangle \psi_m(x)$
- **Kernel representation:** $e^{t\Delta_\rho} f(x) = \int \bar{k}_t(x,y) f(y) \rho(y) \mathrm{d}y$

$$\textbf{Poincaré inequality} \quad \Rightarrow \quad \underbrace{\lambda_1 > 0}_{\textbf{Spectral gap}} \quad \Rightarrow \quad \underbrace{\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t\lambda_1} < 1}_{\textbf{Contraction}}$$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Diffusion map approximation

Mathematical preliminaries

- **Weighted Laplacian:** $\Delta_\rho f := \frac{1}{\rho}\nabla \cdot (\rho \nabla f)$
- **Spectral representation:** $-\Delta_\rho f = \sum_{m=0}^{\infty} \lambda_m \langle \psi_m, f \rangle \psi_m$

$$\text{Eigenvalues:} \quad \lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \dots$$
$$\text{Eigenfunctions:} \quad \psi_0 = 1, \psi_1, \psi_2, , \dots$$

- **Diffusion semigroup:** $e^{t\Delta_\rho} f(x) = \sum_{m=0}^{\infty} e^{-t\lambda_m} \langle \psi_m, f \rangle \psi_m(x)$
- **Kernel representation:** $e^{t\Delta_\rho} f(x) = \int \bar{k}_t(x,y) f(y) \rho(y) \mathrm{d}y$

$$\textbf{Poincaré inequality} \quad \Rightarrow \quad \underbrace{\lambda_1 > 0}_{\textbf{Spectral gap}} \quad \Rightarrow \quad \underbrace{\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t\lambda_1} < 1}_{\textbf{Contraction}}$$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Diffusion map approximation

Mathematical preliminaries

- **Weighted Laplacian:** $\Delta_\rho f := \frac{1}{\rho}\nabla \cdot (\rho \nabla f)$
- **Spectral representation:** $-\Delta_\rho f = \sum_{m=0}^{\infty} \lambda_m \langle \psi_m, f \rangle \psi_m$

$$\text{Eigenvalues:} \quad \lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \dots$$
$$\text{Eigenfunctions:} \quad \psi_0 = 1, \psi_1, \psi_2, , \dots$$

- **Diffusion semigroup:** $e^{t\Delta_\rho} f(x) = \sum_{m=0}^{\infty} e^{-t\lambda_m} \langle \psi_m, f \rangle \psi_m(x)$
- **Kernel representation:** $e^{t\Delta_\rho} f(x) = \int \bar{k}_t(x,y) f(y) \rho(y) \mathrm{d}y$

$$\text{Poincaré inequality} \quad \Rightarrow \quad \underbrace{\lambda_1 > 0}_{\text{Spectral gap}} \quad \Rightarrow \quad \underbrace{\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t\lambda_1} < 1}_{\text{Contraction}}$$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Diffusion map approximation
Mathematical preliminaries

- **Weighted Laplacian:** $\Delta_\rho f := \frac{1}{\rho}\nabla \cdot (\rho \nabla f)$
- **Spectral representation:** $-\Delta_\rho f = \sum_{m=0}^{\infty} \lambda_m \langle \psi_m, f \rangle \psi_m$

$$\text{Eigenvalues:} \quad \lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \dots$$
$$\text{Eigenfunctions:} \quad \psi_0 = 1, \psi_1, \psi_2, , \dots$$

- **Diffusion semigroup:** $e^{t\Delta_\rho} f(x) = \sum_{m=0}^{\infty} e^{-t\lambda_m} \langle \psi_m, f \rangle \psi_m(x)$
- **Kernel representation:** $e^{t\Delta_\rho} f(x) = \int \bar{k}_t(x,y) f(y) \rho(y) \mathrm{d}y$

Poincaré inequality $\Rightarrow$ $\underbrace{\lambda_1 > 0}_{\text{Spectral gap}}$ $\Rightarrow$ $\underbrace{\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t\lambda_1} < 1}_{\text{Contraction}}$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Diffusion map approximation
Mathematical preliminaries

- **Weighted Laplacian:** $\Delta_\rho f := \frac{1}{\rho}\nabla \cdot (\rho \nabla f)$
- **Spectral representation:** $-\Delta_\rho f = \sum_{m=0}^{\infty} \lambda_m \langle \psi_m, f \rangle \psi_m$

$$\text{Eigenvalues:} \quad \lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \dots$$
$$\text{Eigenfunctions:} \quad \psi_0 = 1, \psi_1, \psi_2, , \dots$$

- **Diffusion semigroup:** $e^{t\Delta_\rho} f(x) = \sum_{m=0}^{\infty} e^{-t\lambda_m} \langle \psi_m, f \rangle \psi_m(x)$
- **Kernel representation:** $e^{t\Delta_\rho} f(x) = \int \bar{k}_t(x,y) f(y) \rho(y) \mathrm{d}y$

$$\textbf{Poincaré inequality} \quad \Rightarrow \quad \underbrace{\lambda_1 > 0}_{\textbf{Spectral gap}} \quad \Rightarrow \quad \underbrace{\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t\lambda_1} < 1}_{\textbf{Contraction}}$$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Diffusion map approximation
Gaussian example

Assume ρ is Gaussian density $N(0,1)$

- **Weighted Laplacian**:

$$\Delta_\rho f(x) = e^{\frac{x^2}{2}} \frac{\mathrm{d}}{\mathrm{d}x}(e^{-\frac{x^2}{2}} \frac{\mathrm{d}f}{\mathrm{d}x}(x)) = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x) - x\frac{\mathrm{d}f}{\mathrm{d}x}(x)$$

- **Spectral representation:**

$$\text{Eigenvalues:} \quad \lambda_m = m \quad \text{for} \quad m = 0, 1, 2, \ldots$$
$$\text{Eigenfunctions:} \quad \text{Hermite polynomials}$$

- **Diffusion semigroup:**

$$e^{t\Delta_\rho} f(x) = \int_{\mathbb{R}^d} \underbrace{\frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp\left(-\frac{e^{-2t}x^2 + e^{-2t}y^2 + e^{-t}xy}{2(1-e^{-2t})}\right)}_{\bar{k}_t(x,y)} f(y)\rho(y)\mathrm{d}y$$

Spectral gap: $\lambda_1 = 1$, **Contraction:** $\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t} < 1$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Diffusion map approximation

Gaussian example

Assume ρ is Gaussian density $N(0,1)$

- **Weighted Laplacian**:

$$\Delta_\rho f(x) = e^{\frac{x^2}{2}} \frac{\mathrm{d}}{\mathrm{d}x}(e^{-\frac{x^2}{2}} \frac{\mathrm{d}f}{\mathrm{d}x}(x)) = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x) - x\frac{\mathrm{d}f}{\mathrm{d}x}(x)$$

- **Spectral representation:**

$$\begin{aligned} \text{Eigenvalues:} \quad & \lambda_m = m \quad \text{for} \quad m = 0, 1, 2, \ldots \\ \text{Eigenfunctions:} \quad & \text{Hermite polynomials} \end{aligned}$$

- **Diffusion semigroup:**

$$e^{t\Delta_\rho} f(x) = \int_{\mathbb{R}^d} \underbrace{\frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp\left(-\frac{e^{-2t}x^2 + e^{-2t}y^2 + e^{-t}xy}{2(1-e^{-2t})}\right)}_{\bar{k}_t(x,y)} f(y)\rho(y)\mathrm{d}y$$

Spectral gap: $\lambda_1 = 1$, **Contraction:** $\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t} < 1$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Diffusion map approximation

Gaussian example

Assume ρ is Gaussian density $N(0,1)$

- **Weighted Laplacian**:

$$\Delta_\rho f(x) = e^{\frac{x^2}{2}} \frac{\mathrm{d}}{\mathrm{d}x}(e^{-\frac{x^2}{2}} \frac{\mathrm{d}f}{\mathrm{d}x}(x)) = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x) - x\frac{\mathrm{d}f}{\mathrm{d}x}(x)$$

- **Spectral representation:**

$$\begin{aligned} \text{Eigenvalues:} \quad & \lambda_m = m \quad \text{for} \quad m = 0, 1, 2, \ldots \\ \text{Eigenfunctions:} \quad & \text{Hermite polynomials} \end{aligned}$$

- **Diffusion semigroup:**

$$e^{t\Delta_\rho} f(x) = \int_{\mathbb{R}^d} \underbrace{\frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp\left(-\frac{e^{-2t}x^2 + e^{-2t}y^2 + e^{-t}xy}{2(1-e^{-2t})}\right)}_{\bar{k}_t(x,y)} f(y)\rho(y)\mathrm{d}y$$

Spectral gap: $\lambda_1 = 1$, **Contraction:** $\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t} < 1$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Diffusion map approximation

Gaussian example

Assume ρ is Gaussian density $N(0,1)$

- **Weighted Laplacian**:

$$\Delta_\rho f(x) = e^{\frac{x^2}{2}} \frac{\mathrm{d}}{\mathrm{d}x}(e^{-\frac{x^2}{2}} \frac{\mathrm{d}f}{\mathrm{d}x}(x)) = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x) - x\frac{\mathrm{d}f}{\mathrm{d}x}(x)$$

- **Spectral representation:**

$$\begin{aligned} \text{Eigenvalues:} \quad & \lambda_m = m \quad \text{for} \quad m = 0, 1, 2, \ldots \\ \text{Eigenfunctions:} \quad & \text{Hermite polynomials} \end{aligned}$$

- **Diffusion semigroup:**

$$e^{t\Delta_\rho} f(x) = \int_{\mathbb{R}^d} \underbrace{\frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp\left(-\frac{e^{-2t}x^2 + e^{-2t}y^2 + e^{-t}xy}{2(1-e^{-2t})}\right)}_{\bar{k}_t(x,y)} f(y)\rho(y)\mathrm{d}y$$

Spectral gap: $\lambda_1 = 1$, **Contraction:** $\|e^{t\Delta_\rho}\|_{L^2_0(\rho)} = e^{-t} < 1$

D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, 2013.

Poisson equation to semigroup formulation

Step $(1) \Leftrightarrow (2)$

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

(2) Semigroup formulation: $$\phi = e^{\epsilon \Delta_\rho}\phi + \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\mathrm{d}s$$

- $(1) \Leftrightarrow (2)$ follows from the semigroup identity

$$\frac{\partial}{\partial t} P_t \phi = P_t \Delta_\rho \phi \quad \Rightarrow \quad P_\epsilon \phi = \phi + \int_0^\epsilon P_s \Delta_\rho \phi \, \mathrm{d}s$$

- P_ϵ is a contraction $\Rightarrow$ A unique solution $\phi \in L_0^2(\rho)$ exists

Proposition

Under Assumptions (A1)-(A2) the unique solution to (1) is the unique solution to (2)

Poisson equation to semigroup formulation

Step $(1) \Leftrightarrow (2)$

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

(2) Semigroup formulation: $$\phi = e^{\epsilon \Delta_\rho}\phi + \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\mathrm{d}s$$

- $(1) \Leftrightarrow (2)$ follows from the semigroup identity

$$\frac{\partial}{\partial t} P_t \phi = P_t \Delta_\rho \phi \quad \Rightarrow \quad P_\epsilon \phi = \phi + \int_0^\epsilon P_s \Delta_\rho \phi \, \mathrm{d}s$$

- P_ϵ is a contraction $\Rightarrow$ A unique solution $\phi \in L_0^2(\rho)$ exists

Proposition

Under Assumptions (A1)-(A2) the unique solution to (1) is the unique solution to (2)

Poisson equation to semigroup formulation

Step $(1) \Leftrightarrow (2)$

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

(2) Semigroup formulation: $$\phi = e^{\epsilon \Delta_\rho}\phi + \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\mathrm{d}s$$

- $(1) \Leftrightarrow (2)$ follows from the semigroup identity

$$\frac{\partial}{\partial t} P_t \phi = P_t \Delta_\rho \phi \quad \Rightarrow \quad P_\epsilon \phi = \phi + \int_0^\epsilon P_s \Delta_\rho \phi \, \mathrm{d}s$$

- P_ϵ is a contraction $\Rightarrow$ A unique solution $\phi \in L_0^2(\rho)$ exists

Proposition

Under Assumptions (A1)-(A2) the unique solution to (1) is the unique solution to (2)

Poisson equation to semigroup formulation

Step $(1) \Leftrightarrow (2)$

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

(2) Semigroup formulation: $$\phi = e^{\epsilon \Delta_\rho}\phi + \int_0^\epsilon e^{s\Delta_\rho}(h - \hat{h})\mathrm{d}s$$

- $(1) \Leftrightarrow (2)$ follows from the semigroup identity

$$\frac{\partial}{\partial t} P_t \phi = P_t \Delta_\rho \phi \quad \Rightarrow \quad P_\epsilon \phi = \phi + \int_0^\epsilon P_s \Delta_\rho \phi \, \mathrm{d}s$$

- P_ϵ is a contraction $\Rightarrow$ A unique solution $\phi \in L_0^2(\rho)$ exists

Proposition

Under Assumptions (A1)-(A2) the unique solution to (1) is the unique solution to (2)

Empirical approximation to Finit-dim. eq

Step $(4) \Leftrightarrow (5)$

(4) Empirical approx: $$\phi_\epsilon^{(N)} = T_\epsilon^{(N)} \phi_\epsilon^{(N)} + \epsilon(h - \pi(h))$$

(5) Finite-dim. eq: $$\Phi = \mathsf{T}\Phi + \epsilon(\mathsf{h} - \pi(h))$$

where

$$\mathsf{T}_{ij} := \frac{g_\epsilon(X^i, X^j)}{n_\epsilon^{(N)}(X^i)\sqrt{\sum_{l=1}^N g_\epsilon(X^j, X^l)}}$$

$$\mathsf{h} := (h(X^1), \ldots, h(X^N))$$

- T is a reversible $N \times N$ Markov matrix with invariant dist π
- $\pi(h) = \sum_{i=1}^N \pi_i \mathsf{h}_i$

A unique solution $\Phi \in \mathbb{R}_0^N$ exists

Empirical approximation to Finit-dim. eq

Step (4) $\Leftrightarrow$ (5)

(4) Empirical approx: $$\phi_\epsilon^{(N)} = T_\epsilon^{(N)} \phi_\epsilon^{(N)} + \epsilon(h - \pi(h))$$

(5) Finite-dim. eq: $$\Phi = \mathsf{T}\Phi + \epsilon(\mathsf{h} - \pi(h))$$

where

$$\mathsf{T}_{ij} := \frac{g_\epsilon(X^i, X^j)}{n_\epsilon^{(N)}(X^i)\sqrt{\sum_{l=1}^N g_\epsilon(X^j, X^l)}}$$

$$\mathsf{h} := (h(X^1), \ldots, h(X^N))$$

- T is a reversible $N \times N$ Markov matrix with invariant dist π
- $\pi(h) = \sum_{i=1}^N \pi_i \mathsf{h}_i$

A unique solution $\Phi \in \mathbb{R}_0^N$ exists

Empirical approximation to Finit-dim. eq

Step (4) $\Leftrightarrow$ (5)

(4) Empirical approx: $$\phi_\epsilon^{(N)} = T_\epsilon^{(N)} \phi_\epsilon^{(N)} + \epsilon(h - \pi(h))$$

(5) Finite-dim. eq: $$\Phi = \mathsf{T}\Phi + \epsilon(\mathsf{h} - \pi(h))$$

where

$$\mathsf{T}_{ij} := \frac{g_\epsilon(X^i, X^j)}{n_\epsilon^{(N)}(X^i)\sqrt{\sum_{l=1}^N g_\epsilon(X^j, X^l)}}$$

$$\mathsf{h} := (h(X^1), \ldots, h(X^N))$$

- T is a reversible $N \times N$ Markov matrix with invariant dist π
- $\pi(h) = \sum_{i=1}^N \pi_i \mathsf{h}_i$

A unique solution $\Phi \in \mathbb{R}_0^N$ exists

Empirical approximation to Finit-dim. eq

Step (4) $\Leftrightarrow$ (5)

(4) Empirical approx: $$\phi_\epsilon^{(N)} = T_\epsilon^{(N)} \phi_\epsilon^{(N)} + \epsilon(h - \pi(h))$$

(5) Finite-dim. eq: $$\Phi = \mathsf{T}\Phi + \epsilon(\mathsf{h} - \pi(h))$$

where

$$\mathsf{T}_{ij} := \frac{g_\epsilon(X^i, X^j)}{n_\epsilon^{(N)}(X^i)\sqrt{\sum_{l=1}^N g_\epsilon(X^j, X^l)}}$$

$$\mathsf{h} := (h(X^1), \ldots, h(X^N))$$

- T is a reversible $N \times N$ Markov matrix with invariant dist π
- $\pi(h) = \sum_{i=1}^N \pi_i \mathsf{h}_i$

$$\boxed{\text{A unique solution } \Phi \in \mathbb{R}_0^N \text{ exists}}$$

Error analysis: Bias

Study convergence $T^n_{\frac{t}{n}} \to P_t$ as $n \to \infty$

Proposition

- Feynman-Kac representation of the semigroup

$$P_t f(x) = e^{U(x)} \mathsf{E}[e^{-\int_0^t W(B^x_{2s})\mathrm{d}s} e^{-U(B^x_{2t})} f(B^x_{2t})]$$

  where B^x_t is the Brownian motion with $B^x_0 = x$
- Stochastic representation of the Diffusion map

$$T^n_\epsilon f(x) = e^{U_\epsilon(x)} \mathsf{E}[e^{-\epsilon \sum_{k=0}^{n-1} W_\epsilon(B^x_{2k\epsilon})} e^{-U_\epsilon(B^x_{2t})} f(B^x_{2t})]$$

  where $U_\epsilon = U + O(\epsilon)$ and $W_\epsilon = W + O(\epsilon)$
- Convergence: For all functions $f, \nabla f \in L^4(\rho)$

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

Define $U(x) = \frac{1}{2}\log(\rho(x))$ and $W(x) = |\nabla U(x)|^2 - \Delta U(x)$

Error analysis: Bias

Study convergence $T^n_{\frac{t}{n}} \to P_t$ as $n \to \infty$

Proposition

- Feynman-Kac representation of the semigroup

$$P_t f(x) = e^{U(x)} \mathsf{E}[e^{-\int_0^t W(B^x_{2s})\mathrm{d}s} e^{-U(B^x_{2t})} f(B^x_{2t})]$$

  where B^x_t is the Brownian motion with $B^x_0 = x$

- Stochastic representation of the Diffusion map

$$T^n_\epsilon f(x) = e^{U_\epsilon(x)} \mathsf{E}[e^{-\epsilon \sum_{k=0}^{n-1} W_\epsilon(B^x_{2k\epsilon})} e^{-U_\epsilon(B^x_{2t})} f(B^x_{2t})]$$

  where $U_\epsilon = U + O(\epsilon)$ and $W_\epsilon = W + O(\epsilon)$

- Convergence: For all functions $f, \nabla f \in L^4(\rho)$

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

Define $U(x) = \dfrac{1}{2} \log(\rho(x))$ and $W(x) = |\nabla U(x)|^2 - \Delta U(x)$

Error analysis: Bias

Study convergence $T^n_{\frac{t}{n}} \to P_t$ as $n \to \infty$

Proposition

- Feynman-Kac representation of the semigroup

$$P_t f(x) = e^{U(x)} \mathsf{E}[e^{-\int_0^t W(B^x_{2s})\mathrm{d}s} e^{-U(B^x_{2t})} f(B^x_{2t})]$$

  where B^x_t is the Brownian motion with $B^x_0 = x$

- Stochastic representation of the Diffusion map

$$T^n_\epsilon f(x) = e^{U_\epsilon(x)} \mathsf{E}[e^{-\epsilon \sum_{k=0}^{n-1} W_\epsilon(B^x_{2k\epsilon})} e^{-U_\epsilon(B^x_{2t})} f(B^x_{2t})]$$

  where $U_\epsilon = U + O(\epsilon)$ and $W_\epsilon = W + O(\epsilon)$

- Convergence: For all functions $f, \nabla f \in L^4(\rho)$

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

Define $U(x) = \dfrac{1}{2} \log(\rho(x))$ and $W(x) = |\nabla U(x)|^2 - \Delta U(x)$

Error analysis: Bias

Study convergence $T^n_{\frac{t}{n}} \to P_t$ as $n \to \infty$

Proposition

- Feynman-Kac representation of the semigroup

$$P_t f(x) = e^{U(x)} \mathsf{E}[e^{-\int_0^t W(B^x_{2s})\mathrm{d}s} e^{-U(B^x_{2t})} f(B^x_{2t})]$$

  where B^x_t is the Brownian motion with $B^x_0 = x$

- Stochastic representation of the Diffusion map

$$T^n_\epsilon f(x) = e^{U_\epsilon(x)} \mathsf{E}[e^{-\epsilon \sum_{k=0}^{n-1} W_\epsilon(B^x_{2k\epsilon})} e^{-U_\epsilon(B^x_{2t})} f(B^x_{2t})]$$

  where $U_\epsilon = U + O(\epsilon)$ and $W_\epsilon = W + O(\epsilon)$

- Convergence: For all functions $f, \nabla f \in L^4(\rho)$

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

Define $U(x) = \frac{1}{2}\log(\rho(x))$ and $W(x) = |\nabla U(x)|^2 - \Delta U(x)$

Error analysis: Bias

Spectral gap and error bound

$$\phi_\epsilon = T_\epsilon \phi_\epsilon + \epsilon(h - \hat{h}_\epsilon) \quad \Rightarrow \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \epsilon \sum_{k=0}^{n-1} T_\epsilon^k (h - \hat{h}_\epsilon)$$

Proposition

- Foster-Lyapunov condition DV(3): $\exists a, b, R, \delta > 0$ and prob. measure ν such that

$$\log(e^{-U_\epsilon} T_\epsilon^n e^{U_\epsilon}) \leq -atU_\epsilon + bt\mathbf{1}_{[|x| \leq R]}$$
$$T_\epsilon^n \mathbf{1}_{[A]} \geq \delta \nu(A) \quad \forall |x| \leq R$$

- T_ϵ^n admits a spectral gap on $L^2(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \leq \frac{t\|h\|_{L^2(\rho_\epsilon)}}{\lambda}$$

- In the asymptotic limit as $\epsilon \to 0$

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

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.
G. Roberts, and J. Rosenthal. Geometric ergodicity and hybrid Markov chains. 1997
I. Kontoyiannis, and S. P. Meyn, Geometric ergodicity and the spectral gap of non-reversible Markov chains. 2012

Error analysis: Bias

Spectral gap and error bound

$$\phi_\epsilon = T_\epsilon \phi_\epsilon + \epsilon(h - \hat{h}_\epsilon) \quad \Rightarrow \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \epsilon \sum_{k=0}^{n-1} T_\epsilon^k (h - \hat{h}_\epsilon)$$

Proposition

- Foster-Lyapunov condition DV(3): $\exists a, b, R, \delta > 0$ and prob. measure ν such that

$$\log(e^{-U_\epsilon} T_\epsilon^n e^{U_\epsilon}) \leq -atU_\epsilon + bt\mathbf{1}_{[|x| \leq R]}$$
$$T_\epsilon^n \mathbf{1}_{[A]} \geq \delta \nu(A) \quad \forall |x| \leq R$$

- T_ϵ^n admits a spectral gap on $L^2(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \leq \frac{t\|h\|_{L^2(\rho_\epsilon)}}{\lambda}$$

- In the asymptotic limit as $\epsilon \to 0$

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

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.
G. Roberts, and J. Rosenthal. Geometric ergodicity and hybrid Markov chains. 1997
I. Kontoyiannis, and S. P. Meyn, Geometric ergodicity and the spectral gap of non-reversible Markov chains. 2012

Error analysis: Bias

Spectral gap and error bound

$$\phi_\epsilon = T_\epsilon \phi_\epsilon + \epsilon(h - \hat{h}_\epsilon) \quad \Rightarrow \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \epsilon \sum_{k=0}^{n-1} T_\epsilon^k (h - \hat{h}_\epsilon)$$

Proposition

- Foster-Lyapunov condition DV(3): $\exists a, b, R, \delta > 0$ and prob. measure ν such that

$$\log(e^{-U_\epsilon} T_\epsilon^n e^{U_\epsilon}) \leq -atU_\epsilon + bt\mathbf{1}_{[|x| \leq R]}$$
$$T_\epsilon^n \mathbf{1}_{[A]} \geq \delta \nu(A) \quad \forall |x| \leq R$$

- T_ϵ^n admits a spectral gap on $L^2(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \leq \frac{t\|h\|_{L^2(\rho_\epsilon)}}{\lambda}$$

- In the asymptotic limit as $\epsilon \to 0$

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

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.
G. Roberts, and J. Rosenthal. Geometric ergodicity and hybrid Markov chains. 1997
I. Kontoyiannis, and S. P. Meyn, Geometric ergodicity and the spectral gap of non-reversible Markov chains. 2012

Error analysis: Bias

Spectral gap and error bound

$$\phi_\epsilon = T_\epsilon \phi_\epsilon + \epsilon(h - \hat{h}_\epsilon) \quad \Rightarrow \quad \phi_\epsilon = T_\epsilon^n \phi_\epsilon + \epsilon \sum_{k=0}^{n-1} T_\epsilon^k (h - \hat{h}_\epsilon)$$

Proposition

- Foster-Lyapunov condition DV(3): $\exists a, b, R, \delta > 0$ and prob. measure ν such that

$$\log(e^{-U_\epsilon} T_\epsilon^n e^{U_\epsilon}) \leq -atU_\epsilon + bt\mathbf{1}_{[|x| \leq R]}$$
$$T_\epsilon^n \mathbf{1}_{[A]} \geq \delta \nu(A) \quad \forall |x| \leq R$$

- T_ϵ^n admits a spectral gap on $L^2(\rho_\epsilon)$ and

$$\|\phi_\epsilon\|_{L^2(\rho_\epsilon)} \leq \frac{t\|h\|_{L^2(\rho_\epsilon)}}{\lambda}$$

- In the asymptotic limit as $\epsilon \to 0$

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

S. P. Meyn and R. L. Tweedie. Markov chains and stochastic stability. 2012.
G. Roberts, and J. Rosenthal. Geometric ergodicity and hybrid Markov chains. 1997
I. Kontoyiannis, and S. P. Meyn, Geometric ergodicity and the spectral gap of non-reversible Markov chains. 2012

Optimal transport FPF
Singular covariance case

Idea: Modification of the procedure to construct sequence of stochastic maps

Result:

$$\begin{aligned}\mathrm{d}\bar{X}_t =& A\bar{X}_t + \frac{1}{2}(\sigma_B + e_t)u_t^\top(\bar{X}_t - \bar{m}_t)\mathrm{d}t + e_t\mathrm{d}\bar{B}_t \\ &+ \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{HX_t^i + H\bar{m}_t}{2}\mathrm{d}t) + \text{(extra terms)}\end{aligned}$$

- $u_t = \mathsf{Proj}(\sigma_B; \mathsf{Range}(\bar{\Sigma}_t))$
- $e_t = \sigma_B - \bar{\Sigma}_t u_t$
- It simplifies to the previous sde if $\bar{\Sigma}_t$ is invertible
- Important for finite-N implementation when $N < d$ and $\bar{\Sigma}_t$ is singular

Optimal transport FPF

Singular covariance case

Idea: Modification of the procedure to construct sequence of stochastic maps

Result:

$$\begin{aligned}\mathrm{d}\bar{X}_t =& A\bar{X}_t + \frac{1}{2}(\sigma_B + e_t)u_t^\top(\bar{X}_t - \bar{m}_t)\mathrm{d}t + e_t\mathrm{d}\bar{B}_t \\ &+ \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{HX_t^i + H\bar{m}_t}{2}\mathrm{d}t) + (\text{extra terms})\end{aligned}$$

- $u_t = \mathsf{Proj}(\sigma_B; \mathsf{Range}(\bar{\Sigma}_t))$
- $e_t = \sigma_B - \bar{\Sigma}_t u_t$
- It simplifies to the previous sde if $\bar{\Sigma}_t$ is invertible
- Important for finite-N implementation when $N < d$ and $\bar{\Sigma}_t$ is singular

Optimal transport FPF

Singular covariance case

Idea: Modification of the procedure to construct sequence of stochastic maps

Result:

$$\begin{aligned} \mathrm{d}\bar{X}_t =& A\bar{X}_t + \frac{1}{2}(\sigma_B + e_t)u_t^\top(\bar{X}_t - \bar{m}_t)\mathrm{d}t + e_t\mathrm{d}\bar{B}_t \\ &+ \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{HX_t^i + H\bar{m}_t}{2}\mathrm{d}t) + \text{(extra terms)} \end{aligned}$$

- $u_t = \mathsf{Proj}(\sigma_B; \mathsf{Range}(\bar{\Sigma}_t))$
- $e_t = \sigma_B - \bar{\Sigma}_t u_t$
- It simplifies to the previous sde if $\bar{\Sigma}_t$ is invertible
- Important for finite-N implementation when $N < d$ and $\bar{\Sigma}_t$ is singular

Optimal transport FPF

Singular covariance case

Idea: Modification of the procedure to construct sequence of stochastic maps

Result:

$$\begin{aligned} \mathrm{d}\bar{X}_t =& A\bar{X}_t + \frac{1}{2}(\sigma_B + e_t)u_t^\top(\bar{X}_t - \bar{m}_t)\mathrm{d}t + e_t\mathrm{d}\bar{B}_t \\ &+ \bar{\mathsf{K}}_t(\mathrm{d}Z_t - \frac{HX_t^i + H\bar{m}_t}{2}\mathrm{d}t) + (\text{extra terms}) \end{aligned}$$

- $u_t = \mathsf{Proj}(\sigma_B; \mathsf{Range}(\bar{\Sigma}_t))$
- $e_t = \sigma_B - \bar{\Sigma}_t u_t$
- It simplifies to the previous sde if $\bar{\Sigma}_t$ is invertible
- Important for finite-N implementation when $N < d$ and $\bar{\Sigma}_t$ is singular

Error analysis of finite-N system
Problem formulation

Finite-N system:

$$\mathrm{d}X_t^i = AX_t^i + \sigma_B \mathrm{d}B_t^i + \mathsf{K}_t^{(N)}(\mathrm{d}Z_t - \frac{1}{2}H(X_t^i + m_t^{(N)})\mathrm{d}t), \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

$$\mathsf{K}_t^{(N)} = \Sigma_t^{(N)} H^\top$$

with empirical mean $m_t^{(N)}$ and covariance $\Sigma_t^{(N)}$

Mean-field limit:

$$\mathrm{d}\bar{X}_t = (\text{linear dynamics}) + \bar{\mathsf{K}}_t(\mathrm{d}\bar{Z}_t - \frac{1}{2}H(X_t^i + \bar{m}_t)\mathrm{d}t), \quad \bar{X}_0 \sim p_0$$

$$\bar{\mathsf{K}}_t = \bar{\Sigma}_t H^\top$$

with mean-field mean $\bar{m}_t = \mathsf{E}[\bar{X}_t|\mathcal{Z}_t]$ and covariance $\bar{\Sigma}_t = \mathsf{Cov}(\bar{X}_t|\mathcal{Z}_t)$

Error analysis:

1. Analysis of the mean-field system
2. Analysis of the converegnce of the finite-N system to the mean-field limit

$$\text{Finite-}N\text{ system} \overset{(2)}{\approx} \text{mean-field system} \overset{(1)}{=} \text{Kalman filter}$$

Error analysis of finite-N system
Problem formulation

Finite-N system:

$$\mathrm{d}X_t^i = AX_t^i + \sigma_B \mathrm{d}B_t^i + \mathsf{K}_t^{(N)}(\mathrm{d}Z_t - \frac{1}{2}H(X_t^i + m_t^{(N)})\mathrm{d}t), \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

$$\mathsf{K}_t^{(N)} = \Sigma_t^{(N)} H^\top$$

with empirical mean $m_t^{(N)}$ and covariance $\Sigma_t^{(N)}$

Mean-field limit:

$$\mathrm{d}\bar{X}_t = (\text{linear dynamics}) + \bar{\mathsf{K}}_t(\mathrm{d}\bar{Z}_t - \frac{1}{2}H(X_t^i + \bar{m}_t)\mathrm{d}t), \quad \bar{X}_0 \sim p_0$$

$$\bar{\mathsf{K}}_t = \bar{\Sigma}_t H^\top$$

with mean-field mean $\bar{m}_t = \mathsf{E}[\bar{X}_t | \mathcal{Z}_t]$ and covariance $\bar{\Sigma}_t = \mathsf{Cov}(\bar{X}_t | \mathcal{Z}_t)$

Error analysis:

1. Analysis of the mean-field system
2. Analysis of the converegnce of the finite-N system to the mean-field limit

$$\text{Finite-}N\text{ system} \overset{(2)}{\approx} \text{mean-field system} \overset{(1)}{=} \text{Kalman filter}$$

Error analysis of finite-N system
Problem formulation

Finite-N system:

$$\mathrm{d}X_t^i = AX_t^i + \sigma_B \mathrm{d}B_t^i + \mathsf{K}_t^{(N)}(\mathrm{d}Z_t - \frac{1}{2}H(X_t^i + m_t^{(N)})\mathrm{d}t), \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

$$\mathsf{K}_t^{(N)} = \Sigma_t^{(N)} H^\top$$

with empirical mean $m_t^{(N)}$ and covariance $\Sigma_t^{(N)}$

Mean-field limit:

$$\mathrm{d}\bar{X}_t = (\text{linear dynamics}) + \bar{\mathsf{K}}_t(\mathrm{d}\bar{Z}_t - \frac{1}{2}H(X_t^i + \bar{m}_t)\mathrm{d}t), \quad \bar{X}_0 \sim p_0$$

$$\bar{\mathsf{K}}_t = \bar{\Sigma}_t H^\top$$

with mean-field mean $\bar{m}_t = \mathsf{E}[\bar{X}_t | \mathcal{Z}_t]$ and covariance $\bar{\Sigma}_t = \mathsf{Cov}(\bar{X}_t | \mathcal{Z}_t)$

Error analysis:

1. Analysis of the mean-field system
2. Analysis of the converegnce of the finite-N system to the mean-field limit

$$\text{Finite-}N\text{ system} \overset{(2)}{\approx} \text{mean-field system} \overset{(1)}{=} \text{Kalman filter}$$

Background: Wasserstein gradient

- Objective functional:

$$\mathsf{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

- Wasserstein gradient: $\nabla_W \mathsf{F}(\rho) : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{F}(\rho_t)\Big|_{t=0} = \langle \nabla_W \mathsf{F}(\rho), u \rangle_{L^2(\rho)},$$

for all path $\{\rho_t\}$ s.t $\quad \dfrac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u) = 0$

- Example:

$$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty) \quad \text{(relative entropy)}$$

$$\Longrightarrow \quad \nabla_W \mathsf{F}(\rho)(x) = \nabla \log(\rho(x)) + \nabla f(x)$$

where $f = -\log(\rho_\infty)$

Background: Wasserstein gradient

- Objective functional:

$$\mathsf{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

- Wasserstein gradient: $\nabla_W \mathsf{F}(\rho) : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{F}(\rho_t)\big|_{t=0} = \langle \nabla_W \mathsf{F}(\rho), u \rangle_{L^2(\rho)},$$

$$\text{for all path } \{\rho_t\} \text{ s.t } \quad \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u) = 0$$

- Example:

$$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty) \quad \text{(relative entropy)}$$

$$\Longrightarrow \quad \nabla_W \mathsf{F}(\rho)(x) = \nabla \log(\rho(x)) + \nabla f(x)$$

where $f = -\log(\rho_\infty)$

Background: Wasserstein gradient

- Objective functional:

$$\mathsf{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

- Wasserstein gradient: $\nabla_W \mathsf{F}(\rho) : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{F}(\rho_t)\big|_{t=0} = \langle \nabla_W \mathsf{F}(\rho), u\rangle_{L^2(\rho)},$$

$$\text{for all path } \{\rho_t\} \text{ s.t } \quad \frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t u) = 0$$

- Example:

$$\mathsf{F}(\rho) = \mathsf{D}(\rho \| \rho_\infty) \quad \text{(relative entropy)}$$

$$\Longrightarrow \quad \nabla_W \mathsf{F}(\rho)(x) = \nabla \log(\rho(x)) + \nabla f(x)$$

where $f = -\log(\rho_\infty)$

Comparison to Hamiltonian MCMC

- Proposed accelerated flow:

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}$$

- Continuous-time limit of Hamiltonian MCMC (under-damped Langevin eq.)

$$\begin{aligned} \mathrm{d}X_t &= v_t \mathrm{d}t \\ \mathrm{d}v_t &= -\gamma v_t \mathrm{d}t - \nabla f(X_t)\mathrm{d}t + \underbrace{\sqrt{2}\mathrm{d}B_t}_{\text{stochastic term}} \end{aligned}$$

- Trade-off between computational efficiency and accuracy

Comparison to Hamiltonian MCMC

- Proposed accelerated flow:

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}$$

- Continuous-time limit of Hamiltonian MCMC (under-damped Langevin eq.)

$$\begin{aligned} \mathrm{d}X_t &= v_t \mathrm{d}t \\ \mathrm{d}v_t &= -\gamma v_t \mathrm{d}t - \nabla f(X_t)\mathrm{d}t + \underbrace{\sqrt{2}\mathrm{d}B_t}_{\text{stochastic term}} \end{aligned}$$

- Trade-off between computational efficiency and accuracy

Comparison to Hamiltonian MCMC

- Proposed accelerated flow:

$$\ddot{X}_t = -\frac{3}{t}\dot{X}_t - \nabla f(X_t) - \underbrace{\nabla \log(\rho_t(X_t))}_{\text{mean-field term}}$$

- Continuous-time limit of Hamiltonian MCMC (under-damped Langevin eq.)

$$\begin{aligned} \mathrm{d}X_t &= v_t \mathrm{d}t \\ \mathrm{d}v_t &= -\gamma v_t \mathrm{d}t - \nabla f(X_t)\mathrm{d}t + \underbrace{\sqrt{2}\mathrm{d}B_t}_{\text{stochastic term}} \end{aligned}$$

- Trade-off between computational efficiency and accuracy

Numerical example
comparison with MCMC and HMCMC