Error Estimates for the Kernel Gain Function Approximation in the Feedback Particle Filter

American Control Conference, Seattle, 2017

Amirhossein Taghvaei[†]
Joint work with P. G. Mehta[†], and S. P. Meyn[*]

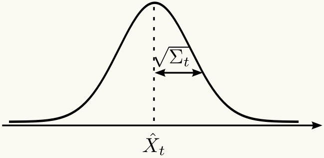
[†]Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
[*]Department of Electrical and Computer Engineering
University of Florida

May 26, 2017

ILLINOIS

Kalman Filter

Linear system
Posterior is Gaussian $N(\hat{X}_t, \Sigma_t)$



$$\mathrm{d}\hat{X}_t = \underbrace{\ldots}_{\text{Propagation}} + \underbrace{\mathsf{K}_t \,\mathrm{d}I_t}_{\text{Correction}}$$

K_t is the Kalman gain

Feedback Particle Filter

Nonlinear system
Posterior ≈ empirical dist. $\{X^1, \ldots, X^N\}$,

$$\mathrm{d}X_t^i = \underbrace{\ldots}_{\text{Propagation}} + \underbrace{\mathsf{K}_t(X_t^i) \circ \mathrm{d}I_t^i}_{\text{Correction}}$$

$\mathsf{K}_t = \nabla\phi$ from Poisson eq.

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

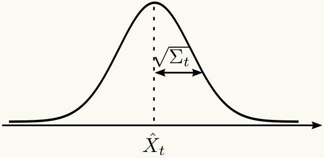
Feedback Particle Filter

Generalization of the Kalman filter

Kalman Filter

Linear system

Posterior is Gaussian $N(\hat{X}_t, \Sigma_t)$



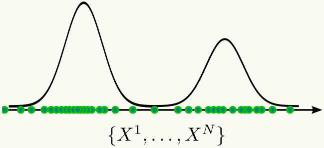
$$\mathrm{d}\hat{X}_t = \underbrace{\ldots}_{\text{Propagation}} + \underbrace{\mathsf{K}_t \,\mathrm{d}I_t}_{\text{Correction}}$$

K_t is the Kalman gain

Feedback Particle Filter

Nonlinear system

Posterior $\approx$ empirical dist. $\{X^1, \ldots, X^N\}$,



$$\mathrm{d}X^i_t = \underbrace{\ldots}_{\text{Propagation}} + \underbrace{\mathsf{K}_t(X^i_t) \circ \mathrm{d}I^i_t}_{\text{Correction}}$$

$\mathsf{K}_t = \nabla\phi$ from Poisson eq.

T. Yang, P. G. Mehta, and S. P. Meyn. feedback particle filter, *TAC*, 2013

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

Gain function: $\mathsf{K}(x) = \nabla\phi(x)$

Poisson equation: $-\frac{1}{\rho}\nabla\cdot(\rho\nabla\phi) = h - \hat{h}, \quad \text{on} \quad \mathbb{R}^d$

- $\rho : \mathbb{R}^d \to \mathbb{R}^+$ (posterior density)
- $h : \mathbb{R}^d \to \mathbb{R}$ (obs. func.), $\hat{h} := \int h(x)\rho(x)\,\mathrm{d}x$
- $\phi : \mathbb{R}^d \to \mathbb{R}$ (unknown)

Problem:

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

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

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. *SIAM*, 2015

Gain function: $\mathsf{K}(x) = \nabla \phi(x)$

Poisson equation: $-\dfrac{1}{\rho} \nabla \cdot (\rho \nabla \phi) = h - \hat{h}, \quad \text{on} \quad \mathbb{R}^d$

- $\rho : \mathbb{R}^d \to \mathbb{R}^+$ (posterior density)
- $h : \mathbb{R}^d \to \mathbb{R}$ (obs. func.), $\hat{h} := \int h(x)\rho(x)\,\mathrm{d}x$
- $\phi : \mathbb{R}^d \to \mathbb{R}$ (unknown)

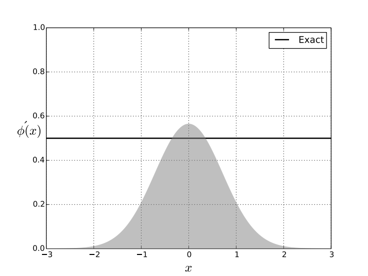
Problem:

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

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

R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. *SIAM*, 2015

Gaussian distribution
linear h



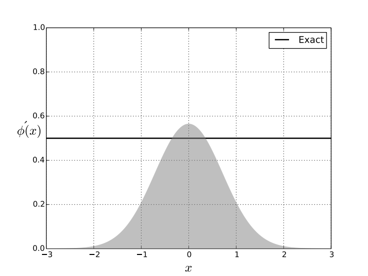
$\mathsf{K}(x) = \text{constant}$ (Kalman gain)

Bimodal distribution
linear h

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

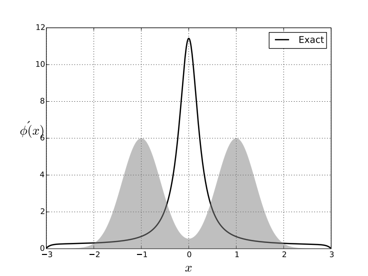
Poisson equation: examples

Gaussian distribution
linear h



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

Bimodal distribution
linear h



$K(x) = \ldots$ (Nonlinear gain)

FPF (theory and application):

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

} (Constant gain and Galerkin approximation)

Gain function approximation:

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

Also in other nonlinear filtering algorithms

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

Problem summary:

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

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

$$\{X^1, \ldots, X^N\} \overset{\text{i.i.d}}{\sim} \rho \longrightarrow \boxed{\text{Algorithm}} \longrightarrow \phi^N$$

Two solution approaches:

(I) PDE

- Theory of elliptic operators
- Weak formulation
- Approximation via projection
- Solve a system of linear equations

(Galerkin algorithm)

(II) Stochastic

- Generator of Markov process
- Fixed pt formulation using semigroup
- Approximation via kernel
- Solve the fixed pt problem iteratively

(kernel-based algorithm)

Problem summary:

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

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

$$\{X^1, \ldots, X^N\} \overset{\text{i.i.d}}{\sim} \rho \longrightarrow \boxed{\text{Algorithm}} \longrightarrow \phi^N$$

Two solution approaches:

(I) PDE

- Theory of elliptic operators
- Weak formulation
- Approximation via projection
- Solve a system of linear equations

(Galerkin algorithm)

(II) Stochastic

- Generator of Markov process
- Fixed pt formulation using semigroup
- Approximation via kernel
- Solve the fixed pt problem iteratively

(kernel-based algorithm)

Problem summary:

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

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

$$\{X^1, \ldots, X^N\} \overset{\text{i.i.d}}{\sim} \rho \longrightarrow \boxed{\text{Algorithm}} \longrightarrow \phi^N$$

Two solution approaches:

(I) PDE

- Theory of elliptic operators
- Weak formulation
- Approximation via projection
- Solve a system of linear equations

(Galerkin algorithm)

(II) Stochastic

- Generator of Markov process
- Fixed pt formulation using semigroup
- Approximation via kernel
- Solve the fixed pt problem iteratively

(kernel-based algorithm)

Problem summary:

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

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

$$\{X^1, \ldots, X^N\} \overset{\text{i.i.d}}{\sim} \rho \longrightarrow \boxed{\text{Algorithm}} \longrightarrow \phi^N$$

Two solution approaches:

(I) PDE

- Theory of elliptic operators
- Weak formulation
- Approximation via projection
- Solve a system of linear equations

(Galerkin algorithm)

(II) Stochastic

- Generator of Markov process
- Fixed pt formulation using semigroup
- Approximation via kernel
- Solve the fixed pt problem iteratively

(kernel-based algorithm)

Heat equation:

$$\frac{\partial u}{\partial t} = \Delta_\rho u \quad \text{on} \quad \mathbb{R}^d$$
$$u(x,0) = f(x) \quad \text{initial condition}$$

Semigroup: The operator $e^{t\Delta_\rho}$ identifies the solution:

$$u(x,t) = e^{t\Delta_\rho} f(x)$$

Example: $\rho = 1$

$$e^{t\Delta} f(x) = \int_{\mathbb{R}^d} g_t(x,y) f(y)\,\mathrm{d}y$$

where $g_t(x,y)$ is the Gaussian kernel.

Useful identity:

$$e^{t\Delta_\rho} f = f + \int_0^t e^{s\Delta_\rho} \Delta_\rho f \;\mathrm{d}s$$

Heat equation:

$$\frac{\partial u}{\partial t} = \Delta_\rho u \quad \text{on} \quad \mathbb{R}^d$$
$$u(x,0) = f(x) \quad \text{initial condition}$$

Semigroup: The operator $e^{t\Delta_\rho}$ identifies the solution:

$$u(x,t) = e^{t\Delta_\rho} f(x)$$

Example: $\rho = 1$

$$e^{t\Delta} f(x) = \int_{\mathbb{R}^d} g_t(x,y) f(y) \, \mathrm{d}y$$

where $g_t(x,y)$ is the Gaussian kernel.

Useful identity:

$$e^{t\Delta_\rho} f = f + \int_0^t e^{s\Delta_\rho} \Delta_\rho f \; \mathrm{d}s$$

Heat equation:

$$\frac{\partial u}{\partial t} = \Delta_\rho u \quad \text{on} \quad \mathbb{R}^d$$
$$u(x,0) = f(x) \quad \text{initial condition}$$

Semigroup: The operator $e^{t\Delta_\rho}$ identifies the solution:

$$u(x,t) = e^{t\Delta_\rho} f(x)$$

Example: $\rho = 1$

$$e^{t\Delta} f(x) = \int_{\mathbb{R}^d} g_t(x,y) f(y)\,\mathrm{d}y$$

where $g_t(x,y)$ is the Gaussian kernel.

Useful identity:

$$e^{t\Delta_\rho} f = f + \int_0^t e^{s\Delta_\rho} \Delta_\rho f \;\mathrm{d}s$$

Heat equation:

$$\frac{\partial u}{\partial t} = \Delta_\rho u \quad \text{on} \quad \mathbb{R}^d$$
$$u(x,0) = f(x) \quad \text{initial condition}$$

Semigroup: The operator $e^{t\Delta_\rho}$ identifies the solution:

$$u(x,t) = e^{t\Delta_\rho} f(x)$$

Example: $\rho = 1$

$$e^{t\Delta} f(x) = \int_{\mathbb{R}^d} g_t(x,y) f(y) \, \mathrm{d}y$$

where $g_t(x,y)$ is the Gaussian kernel.

Useful identity:

$$e^{t\Delta_\rho} f = f + \int_0^t e^{s\Delta_\rho} \Delta_\rho f \; \mathrm{d}s$$

Kernel-based algorithm
Overview

Step 1: Convert to fixed point problem using the semigroup

$$\phi = e^{\epsilon\Delta_\rho}\phi + \int_0^\epsilon e^{s\Delta}(h - \hat{h})\,\mathrm{d}s$$

Step 2: Approximate the semigroup using kernel

$$e^{\epsilon\Delta_\rho}\phi(x) \approx \int_{\mathbb{R}^d} k_\epsilon(x,y)\phi(y)\rho(y)\,\mathrm{d}y =: T_\epsilon\phi(x), \quad \text{as} \quad \epsilon \downarrow 0$$

where $k_\epsilon(x,y) := \frac{1}{n_\epsilon(x)} \frac{g_\epsilon(x,y)}{\sqrt{\int g_\epsilon(y,z)\rho(z)\,\mathrm{d}z}}$.

Step 3: Approximate the integral empirically

$$T_\epsilon\phi(x) \approx \frac{1}{N}\sum_{i=1}^{N} k_\epsilon(x,X^i)\phi(X^i) =: T_\epsilon^{(N)}\phi(x) \quad \text{as} \quad N \uparrow \infty$$

where $k_\epsilon^{(N)}(x,y) := \frac{1}{n_\epsilon^{(N)}(x)} \frac{g_\epsilon(x,y)}{\sqrt{\sum_i g_\epsilon(y,X^i)}}$.

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,
M. Hein, et. al., Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

Kernel-based algorithm
Overview

Step 1: Convert to fixed point problem using the semigroup

$$\phi = e^{\epsilon\Delta_\rho}\phi + \int_0^\epsilon e^{s\Delta}(h - \hat{h})\,\mathrm{d}s$$

Step 2: Approximate the semigroup using kernel

$$e^{\epsilon\Delta_\rho}\phi(x) \approx \int_{\mathbb{R}^d} k_\epsilon(x,y)\phi(y)\rho(y)\,\mathrm{d}y =: T_\epsilon\phi(x), \quad \text{as} \quad \epsilon \downarrow 0$$

where $k_\epsilon(x,y) := \dfrac{1}{n_\epsilon(x)} \dfrac{g_\epsilon(x,y)}{\sqrt{\int g_\epsilon(y,z)\rho(z)\,\mathrm{d}z}}$.

Step 3: Approximate the integral empirically

$$T_\epsilon\phi(x) \approx \frac{1}{N}\sum_{i=1}^N k_\epsilon(x,X^i)\phi(X^i) =: T_\epsilon^{(N)}\phi(x) \quad \text{as} \quad N \uparrow \infty$$

where $k_\epsilon^{(N)}(x,y) := \dfrac{1}{n_\epsilon^{(N)}(x)} \dfrac{g_\epsilon(x,y)}{\sqrt{\sum_i g_\epsilon(y,X^i)}}$.

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,
M. Hein, et. al., Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

Step 1: Convert to fixed point problem using the semigroup

$$\phi = e^{\epsilon\Delta_\rho}\phi + \int_0^\epsilon e^{s\Delta}(h - \hat{h})\,\mathrm{d}s$$

Step 2: Approximate the semigroup using kernel

$$e^{\epsilon\Delta_\rho}\phi(x) \approx \int_{\mathbb{R}^d} k_\epsilon(x,y)\phi(y)\rho(y)\,\mathrm{d}y =: T_\epsilon\phi(x), \quad \text{as} \quad \epsilon \downarrow 0$$

where $k_\epsilon(x,y) := \dfrac{1}{n_\epsilon(x)} \dfrac{g_\epsilon(x,y)}{\sqrt{\int g_\epsilon(y,z)\rho(z)\,\mathrm{d}z}}$.

Step 3: Approximate the integral empirically

$$T_\epsilon\phi(x) \approx \frac{1}{N}\sum_{i=1}^N k_\epsilon(x,X^i)\phi(X^i) =: T_\epsilon^{(N)}\phi(x) \quad \text{as} \quad N \uparrow \infty$$

where $k_\epsilon^{(N)}(x,y) := \dfrac{1}{n_\epsilon^{(N)}(x)} \dfrac{g_\epsilon(x,y)}{\sqrt{\sum_i g_\epsilon(y,X^i)}}$.

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,
M. Hein, et. al., Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

Fixed point problem: $\phi = e^{\epsilon \Delta_\rho} \phi + \int_0^\epsilon e^{s\Delta}(h - \hat{h})\, \mathrm{d}s$

Kernel approximation: $\phi_\epsilon = T_\epsilon \phi_\epsilon + \epsilon(h - \hat{h})$

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

- $T_\epsilon \phi(x) := \int k_\epsilon(x,y)\phi(y)\rho(y)\, \mathrm{d}y.$
- $T_\epsilon^{(N)} \phi(x) := \frac{1}{N} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i)\phi(X^i).$

Formula for approximate gain:

$$\mathsf{K}_\epsilon^{(N)} := \nabla T_\epsilon^{(N)} \phi_\epsilon^{(N)} + \epsilon \nabla T_\epsilon^{(N)} (h - \hat{h})$$

Fixed point problem: $\phi = e^{\epsilon\Delta_\rho}\phi + \int_0^\epsilon e^{s\Delta}(h-\hat{h})\,\mathrm{d}s$

Kernel approximation: $\phi_\epsilon = T_\epsilon\phi_\epsilon + \epsilon(h-\hat{h})$

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

- $T_\epsilon\phi(x) := \int k_\epsilon(x,y)\phi(y)\rho(y)\,\mathrm{d}y.$
- $T_\epsilon^{(N)}\phi(x) := \frac{1}{N}\sum_{i=1}^{N} k_\epsilon^{(N)}(x,X^i)\phi(X^i).$

Formula for approximate gain:

$$\mathsf{K}_\epsilon^{(N)} := \nabla T_\epsilon^{(N)}\phi_\epsilon^{(N)} + \epsilon\nabla T_\epsilon^{(N)}(h-\hat{h})$$

Input: $\underbrace{\epsilon}_{\text{kernel bandwidth}}$, $\{X^1, \ldots, X^N\}$, $\{h(X^1), \ldots, h(X^N)\}$

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

1. Compute the (Markov) matrix $\mathbf{T} \in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$:

$$\mathbf{T}_{ij} = k_\epsilon^{(N)}(X^i, X^j)$$

2. Compute $\Phi \in \mathbb{R}^N$ iteratively:

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

3. Compute the gain:

$$\mathsf{K}_\epsilon^{(N)}(X^i) := \sum_{j=1}^{N} s_{ij} X^j$$

where $s_{ij} = T_{ij}(\Phi_j + \epsilon \mathbf{h}_j - \sum_l T_{il}(\Phi_l + \epsilon \mathbf{h}_l))$.

$$\underbrace{\mathsf{E}\left[\|\mathsf{K}-\mathsf{K}_\epsilon^{(N)}\|_2\right]}_{\text{Total error}} \leq \underbrace{\|\mathsf{K}-\mathsf{K}_\epsilon\|_2}_{\text{Bias}} + \underbrace{\mathsf{E}\left[\|\mathsf{K}_\epsilon-\mathsf{K}_\epsilon^{(N)}\|_2\right]}_{\text{Variance}}$$

Assumptions:

1. $\rho = e^{-V}$ where $\liminf_{|x|\to\infty}[-\Delta V(x) + \frac{1}{2}|\nabla V(x)|^2] = \infty$
2. $h, \nabla h \in L^2$.

Result:

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

Proof sketch:

- **Bias:** Show the expansion $T_\epsilon f = f + \epsilon\Delta_\rho f + O(\epsilon^2)$ and show $(I-T_\epsilon)^{-1}$ is bounded.
- **Variance:** Show $T_\epsilon, T_\epsilon^{(N)}$ are collectively compact and use LLN.

$$\underbrace{\mathsf{E}\left[\|\mathsf{K}-\mathsf{K}_\epsilon^{(N)}\|_2\right]}_{\text{Total error}} \leq \underbrace{\|\mathsf{K}-\mathsf{K}_\epsilon\|_2}_{\text{Bias}} + \underbrace{\mathsf{E}\left[\|\mathsf{K}_\epsilon-\mathsf{K}_\epsilon^{(N)}\|_2\right]}_{\text{Variance}}$$

Assumptions:

1. $\rho = e^{-V}$ where $\liminf_{|x|\to\infty}[-\Delta V(x) + \frac{1}{2}|\nabla V(x)|^2] = \infty$
2. $h, \nabla h \in L^2$.

Result:

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

Proof sketch:

- **Bias:** Show the expansion $T_\epsilon f = f + \epsilon\Delta_\rho f + O(\epsilon^2)$ and show $(I-T_\epsilon)^{-1}$ is bounded.
- **Variance:** Show $T_\epsilon, T_\epsilon^{(N)}$ are collectively compact and use LLN.

$$\underbrace{\mathsf{E}\left[\|\mathsf{K}-\mathsf{K}_\epsilon^{(N)}\|_2\right]}_{\text{Total error}} \leq \underbrace{\|\mathsf{K}-\mathsf{K}_\epsilon\|_2}_{\text{Bias}} + \underbrace{\mathsf{E}\left[\|\mathsf{K}_\epsilon-\mathsf{K}_\epsilon^{(N)}\|_2\right]}_{\text{Variance}}$$

Assumptions:

1. $\rho = e^{-V}$ where $\liminf_{|x|\to\infty}[-\Delta V(x) + \frac{1}{2}|\nabla V(x)|^2] = \infty$
2. $h, \nabla h \in L^2$.

Result:

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

Proof sketch:

- **Bias:** Show the expansion $T_\epsilon f = f + \epsilon\Delta_\rho f + O(\epsilon^2)$ and show $(I-T_\epsilon)^{-1}$ is bounded.
- **Variance:** Show $T_\epsilon, T_\epsilon^{(N)}$ are collectively compact and use LLN.

$$\underbrace{\mathsf{E}\left[\|\mathsf{K}-\mathsf{K}_\epsilon^{(N)}\|_2\right]}_{\text{Total error}} \leq \underbrace{\|\mathsf{K}-\mathsf{K}_\epsilon\|_2}_{\text{Bias}} + \underbrace{\mathsf{E}\left[\|\mathsf{K}_\epsilon-\mathsf{K}_\epsilon^{(N)}\|_2\right]}_{\text{Variance}}$$

Assumptions:

1. $\rho = e^{-V}$ where $\liminf_{|x|\to\infty}[-\Delta V(x) + \frac{1}{2}|\nabla V(x)|^2] = \infty$
2. $h, \nabla h \in L^2$.

Result:

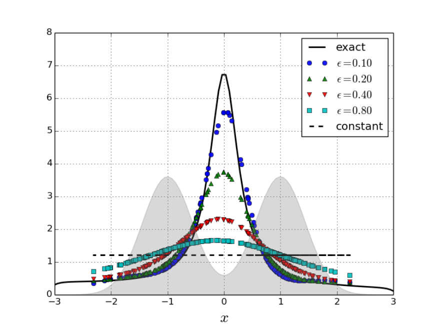
$$\mathsf{E}\left[\|\mathsf{K}-\mathsf{K}_\epsilon^{(N)}\|_2\right] \leq \underbrace{O(\epsilon)}_{\text{Bias}} + \underbrace{O(\frac{1}{\epsilon^{1+d/4}\sqrt{N}})}_{\text{Variance}}$$

Proof sketch:

- **Bias:** Show the expansion $T_\epsilon f = f + \epsilon\Delta_\rho f + O(\epsilon^2)$ and show $(I - T_\epsilon)^{-1}$ is bounded.
- **Variance:** Show $T_\epsilon, T_\epsilon^{(N)}$ are collectively compact and use LLN.

Kernel based gain approximation

Large ϵ values



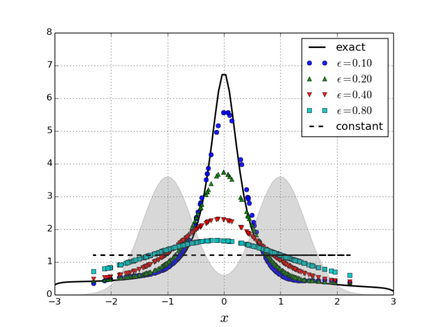
Small ϵ values

Result: Convergence to constant gain approximation

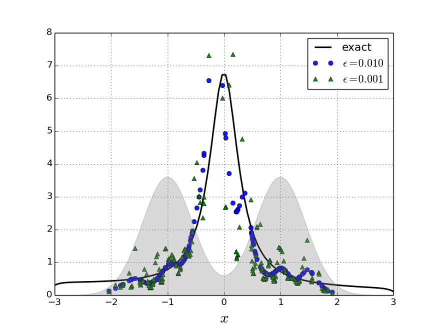
$$\mathsf{K}_\epsilon^{(N)} \to \mathsf{K}_c \quad \text{as} \quad \epsilon \to \infty$$

Kernel based gain approximation

Large ϵ values



Small ϵ values

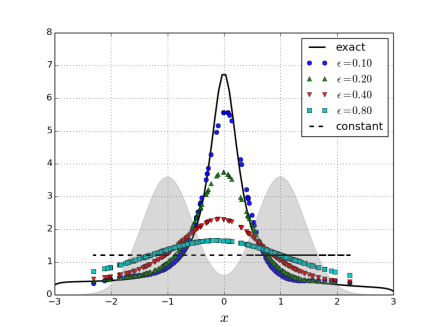


Result: Convergence to constant gain approximation

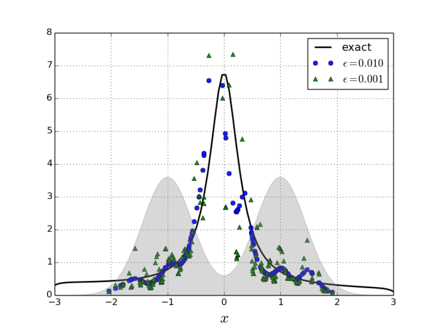
$$\mathsf{K}_\epsilon^{(N)} \to \mathsf{K}_c \quad \text{as} \quad \epsilon \to \infty$$

Kernel based gain approximation

Large ϵ values



Small ϵ values

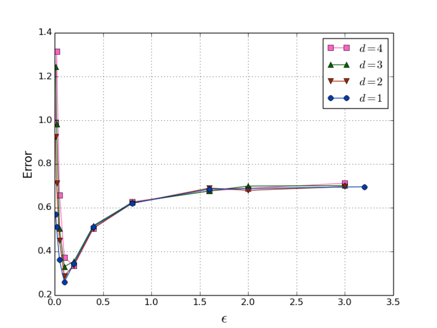


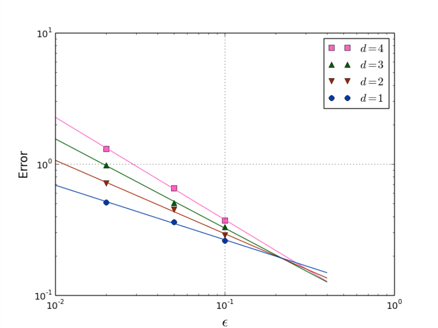
Result: Convergence to constant gain approximation

$$\mathsf{K}_\epsilon^{(N)} \to \mathsf{K}_c \quad \text{as} \quad \epsilon \to \infty$$

Example: Bimodal distribution

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





Conclusion

Properties of kernel-based approximation

1. Numerical stability
2. Easy extension to Manifolds [C. Zhang, et. al. CDC 2015]
3. Provable error bounds
4. Computational cost $O(N^2)$

Future work:

1. Error analysis of the overall filtering algorithm
2. Improve the computational efficiency
3. Distributed implementation

Acknowledgement: Support from CSE fellowship award is gratefully acknowledged

Thank you for your attention!