

# Kernel Algorithm for Gain Function Approximation in the Feedback Particle Filter

*Sequential Monte Carlo workshop  
Uppsala Universitet, Aug 30-Sep 1, 2017*

Prashant G. Mehta<sup>†</sup>

Joint work with Amirhossein Taghvaei<sup>†</sup> and Sean Meyn<sup>+</sup>

<sup>†</sup>Coordinated Science Laboratory

Department of Mechanical Science and Engg., U. Illinois

<sup>+</sup>Department of Electrical and Computer Engg., U. Florida

Aug 30, 2017



ILLINOIS



# Numerical Solution of a BVP

And its applications to particle filters

**BVP:**

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



# Numerical Solution of a BVP

And its applications to particle filters

**BVP:**

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

**Assumptions/Notation:**

- Density  $\rho = e^{-V}$  where  $\lim_{|x| \rightarrow \infty} [-\Delta V(x) + \frac{1}{2} |\nabla V(x)|^2] = \infty$  and  $D^2 V \in L^\infty$
- Function  $h$  is given with  $h, \nabla h \in L^2(\rho; \mathbb{R}^d)$
- $\hat{h} := \int_{\mathbb{R}^d} h(x) \rho(x) dx$



# Numerical Solution of a BVP

And its applications to particle filters

**BVP:**

$$-\frac{1}{\rho(x)} \nabla \cdot (\rho(x) \underbrace{\nabla \phi}_{K}(x)) = (h(x) - \hat{h}) \quad \text{on } \mathbb{R}^d$$

**Assumptions/Notation:**

- Density  $\rho = e^{-V}$  where  $\lim_{|x| \rightarrow \infty} [-\Delta V(x) + \frac{1}{2}|\nabla V(x)|^2] = \infty$  and  $D^2V \in L^\infty$
- Function  $h$  is given with  $h, \nabla h \in L^2(\rho; \mathbb{R}^d)$
- $\hat{h} := \int_{\mathbb{R}^d} h(x)\rho(x) dx$



## BVP:

**Weighted Poisson equation:**  $-\Delta_{\rho}\phi = h - \hat{h}$ , on  $\mathbb{R}^d$

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

## Assumptions/Notation:

- Density  $\rho = e^{-V}$  where  $\lim_{|x| \rightarrow \infty} [-\Delta V(x) + \frac{1}{2}|\nabla V(x)|^2] = \infty$  and  $D^2V \in L^\infty$
- Function  $h$  is given with  $h, \nabla h \in L^2(\rho; \mathbb{R}^d)$
- $\hat{h} := \int_{\mathbb{R}^d} h(x)\rho(x) dx$

# Numerical Solution of a BVP

## And its applications to particle filters



**BVP:**

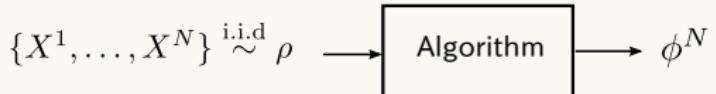
**Weighted Poisson equation:**  $-\Delta_\rho \phi = h - \hat{h}$ , on  $\mathbb{R}^d$

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

**Assumptions/Notation:**

- Density  $\rho = e^{-V}$  where  $\lim_{|x| \rightarrow \infty} [-\Delta V(x) + \frac{1}{2} |\nabla V(x)|^2] = \infty$  and  $D^2 V \in L^\infty$
- Function  $h$  is given with  $h, \nabla h \in L^2(\rho; \mathbb{R}^d)$
- $\hat{h} := \int_{\mathbb{R}^d} h(x) \rho(x) dx$

**Problem:**



Convergence and error analysis for  $\phi^N \rightarrow \phi$  as  $N \rightarrow \infty$



### Problem:

**Signal model:**  $dX_t = a(X_t) dt + dB_t \quad X_0 \sim p_0^*$

**Observation model:**  $dZ_t = h(X_t) dt + dW_t$

Posterior distribution of  $X_t$  given  $\mathcal{Z}_t := \sigma(Z_s : 0 \leq s \leq t)$ ?

# Feedback Particle Filter

A numerical algorithm for nonlinear filtering

**Problem:**

**Signal model:**  $dX_t = a(X_t) dt + dB_t \quad X_0 \sim p_0^*$

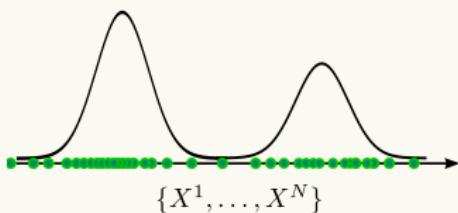
**Observation model:**  $dZ_t = h(X_t) dt + dW_t$

Posterior distribution of  $X_t$  given  $\mathcal{Z}_t := \sigma(Z_s : 0 \leq s \leq t)$ ?

**Solution: Feedback particle filter**

$P(X_t | \mathcal{Z}_t) \approx \text{empirical dist. of } \{X_t^1, \dots, X_t^N\}$

$$dX_t^i = \underbrace{a(X_t^i) dt + dB_t^i}_{\text{Propagation}} + \underbrace{\mathbf{K}_t(X_t^i) \circ (dZ_t - \frac{h(X_t^i) + \hat{h}_t}{2} dt)}_{\text{Update}}, \quad X_0^i \sim p_0^*$$





## Why it works?

### Exactness

- Fokker-Plank equation for the conditional density of  $X_t^i$ :

$$dp_t = \mathcal{L}p_t dt - \nabla \cdot (p_t \mathbf{K}_t) dZ_t + (\dots) dt, \quad p_0 = p_0^*$$

- Nonlinear filtering equation for the conditional density of  $X_t$ :

$$dp_t^* = \mathcal{L}p_t^* dt + p_t(h - \hat{h}_t)(dZ_t - \hat{h}_t dt), \quad p_0^* = p_0^*$$

The easy part

If  $\mathbf{K}_t$  satisfies the following linear pde

$$\nabla \cdot (p_t \mathbf{K}_t) = -(h - \hat{h}_t)p_t \quad \forall t > 0$$

then

$$p_t = p_t^* \quad \forall t > 0$$



## Why it works?

### Exactness

- Fokker-Plank equation for the conditional density of  $X_t^i$ :

$$dp_t = \mathcal{L}p_t dt - \nabla \cdot (p_t K_t) dZ_t + (\dots) dt, \quad p_0 = p_0^*$$

- Nonlinear filtering equation for the conditional density of  $X_t$ :

$$dp_t^* = \mathcal{L}p_t^* dt + p_t(h - \hat{h}_t)(dZ_t - \hat{h}_t dt), \quad p_0^* = p_0^*$$

### The easy part

If  $K_t$  satisfies the following linear pde

$$\nabla \cdot (p_t K_t) = -(h - \hat{h}_t)p_t \quad \forall t > 0$$

then

$$p_t = p_t^* \quad \forall t > 0$$

## Why it works?

### Exactness

- Fokker-Plank equation for the conditional density of  $X_t^i$ :

$$dp_t = \mathcal{L}p_t dt - \nabla \cdot (p_t K_t) dZ_t + (\dots) dt, \quad p_0 = p_0^*$$

- Nonlinear filtering equation for the conditional density of  $X_t$ :

$$dp_t^* = \mathcal{L}p_t^* dt + p_t(h - \hat{h}_t)(dZ_t - \hat{h}_t dt), \quad p_0^* = p_0^*$$

### The easy part

If  $K_t$  satisfies the following linear pde

$$\nabla \cdot (p_t K_t) = -(h - \hat{h}_t)p_t \quad \forall t > 0$$

then

$$p_t = p_t^* \quad \forall t > 0$$

**The hard part:** Computing the gain function  $K_t(\cdot)$

## Why is it useful?

Analogy with the Kalman filter

**Problem:**

**Signal model:**  $dX_t = AX_t dt + dB_t, \quad X_0 \sim N(\hat{X}_0, \Sigma_0)$

**Observation model:**  $dZ_t = HX_t dt + dW_t$

**Posterior:**  $N(\hat{X}_t, \Sigma_t)$

**Solution:**

**Kalman filter:**  $d\hat{X}_t = AX_t dt + \underbrace{K_t(dZ_t - H\hat{X}_t dt)}_{\text{update}}$

## Why is it useful?

Analogy with the Kalman filter

**Problem:**

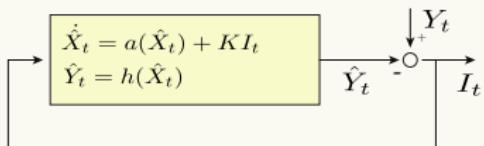
**Signal model:**  $dX_t = AX_t dt + dB_t, \quad X_0 \sim N(\hat{X}_0, \Sigma_0)$

**Observation model:**  $dZ_t = HX_t dt + dW_t$

**Posterior:**  $N(\hat{X}_t, \Sigma_t)$

**Solution:**

**Kalman filter:**  $d\hat{X}_t = A\hat{X}_t dt + \underbrace{K_t(dZ_t - H\hat{X}_t dt)}_{\text{update}}$



Kalman Filter

# Why is it useful?

Analogy with the Kalman filter

**Problem:**

$$\text{Signal model: } dX_t = AX_t dt + dB_t, \quad X_0 \sim N(\hat{X}_0, \Sigma_0)$$

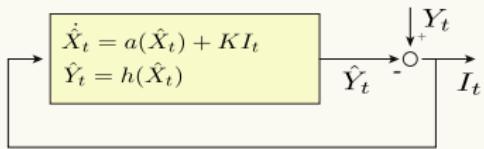
$$\text{Observation model: } dZ_t = HX_t dt + dW_t$$

$$\text{Posterior: } N(\hat{X}_t, \Sigma_t)$$

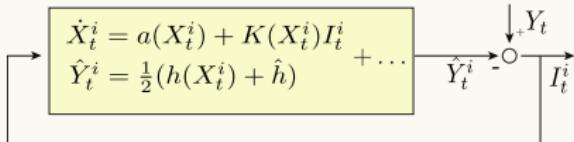
**Solution:**

$$\text{Kalman filter: } d\hat{X}_t = A\hat{X}_t dt + \underbrace{K_t(dZ_t - H\hat{X}_t dt)}_{\text{update}}$$

$$\text{FPF: } dX_t^i = a(X_t^i) dt + dB_t^i + K_t(X_t^i) \circ \underbrace{(dZ_t - \frac{h(X_t^i) + \hat{h}_t}{2} dt)}_{\text{update}}$$



Kalman Filter



Feedback Particle Filter



$$\nabla \cdot (\rho(x) \mathbf{K}(x)) = (\text{rhs}) \text{ also arises in particle flow algorithms}$$

**Continuous-time:** Crisan and Xiong (2009) Approximate McKean-Vlasov representations for a class of SPDEs.

**Ensemble Kalman filter (discrete-time):** Reich (2011) A dynamical systems framework for intermittent data assimilation; Reich (2012,2013); Bergemann and Reich (2010, 2012); Reich and Cotter (2013, 2015).

**Homotopy/Optimal transport (discrete-time):** Daum and Huang (2010- ); Moseley and Marzouk (2012); Reich (2013); Heng, Doucet and Pokern (2015) and others.

Since 2013, an invited session “*Homotopy methods for Bayesian Estimation*” is a regular fixture at the International Conference on Information Fusion.

**Applications of FPF:** Satellite tracking (Berntrop, 2015); Dredging (Stano, 2013); Motion sensing (Tilton, 2013).

**BVP:**

$$\begin{aligned} -\frac{1}{\rho(x)} \nabla \cdot (\rho(x) \nabla \phi(x)) &= (h(x) - \hat{h}) \quad \text{on } \mathbb{R}^d \\ \int_{\mathbb{R}^d} \phi(x) \rho(x) dx &= 0 \end{aligned}$$

**Problem:**

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

**Compute:**  $\{\mathbf{K}(X^1), \dots, \mathbf{K}(X^N)\}$



1 Ensemble Kalman filter +

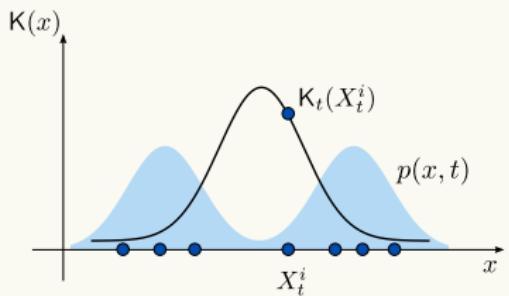
2 Kernel Algorithm

## (1) Non-Gaussian density, (2) Gaussian density

(1) Nonlinear gain function, (2) Constant gain function = Kalman gain



$$(1) \text{ FPF: } dX_t^i = a(X_t^i) dt + dB_t^i + \underbrace{K_t(X_t^i) \circ (dZ_t - \frac{h(X_t^i) + \hat{h}_t}{2} dt)}_{\text{update}}$$

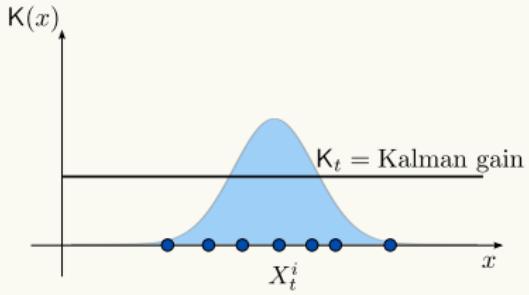
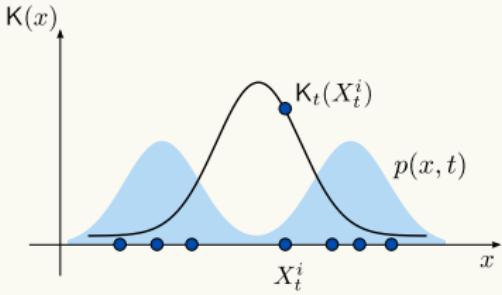


# (1) Non-Gaussian density, (2) Gaussian density

(1) Nonlinear gain function, (2) Constant gain function = Kalman gain

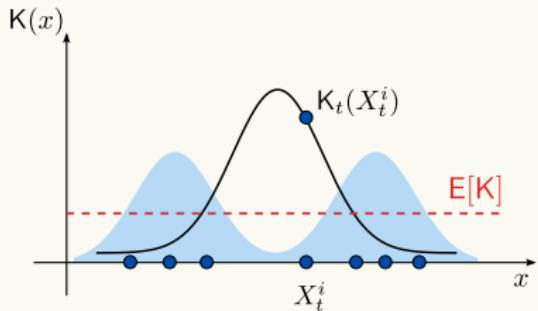
$$(1) \text{ FPF: } dX_t^i = a(X_t^i) dt + dB_t^i + \underbrace{K_t(X_t^i) \circ (dZ_t - \frac{h(X_t^i) + \hat{h}_t}{2} dt)}_{\text{update}}$$

$$(2) \text{ Linear Gaussian: } dX_t^i = AX_t^i dt + dB_t^i + K_t(dZ_t - \frac{HX_t^i + H\hat{X}_t}{2} dt) \underbrace{\qquad}_{\text{update}}$$



## Non-Gaussian case

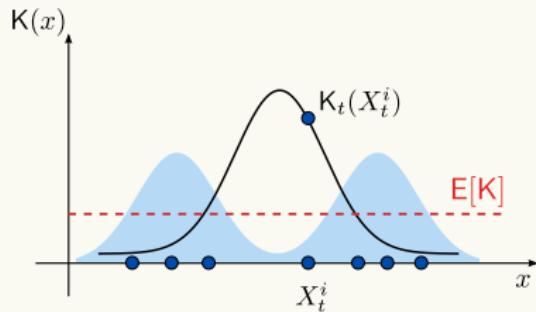
Formula for constant gain approximation



$$\mathbb{E}[K] = \int (h(x) - \hat{h})x\rho(x) dx \approx \frac{1}{N} \sum_{i=1}^N (h(X^i) - \hat{h})X^i$$

## Non-Gaussian case

Formula for constant gain approximation



$$E[K] = \int (h(x) - \hat{h})x\rho(x) dx \approx \frac{1}{N} \sum_{i=1}^N (h(X^i) - \hat{h})X^i$$

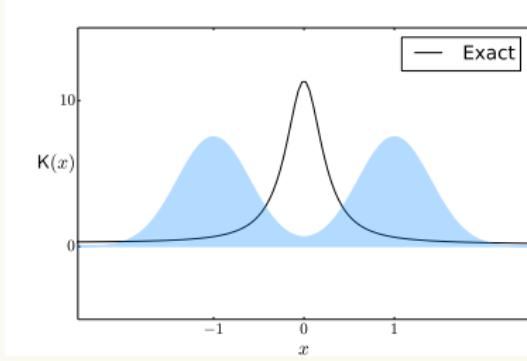
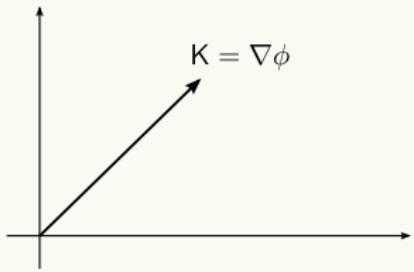
Using the constant gain approximation, linear FPF is the ensemble Kalman filter

## Non-Gaussian case

### Galerkin approximation



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

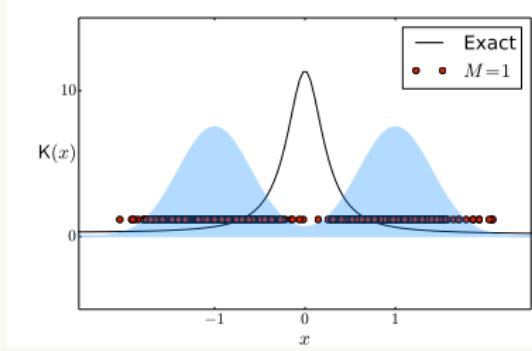
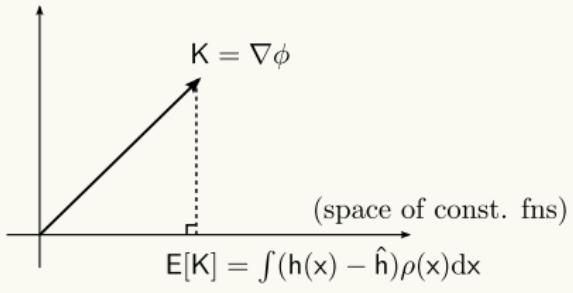


# Non-Gaussian case

## Galerkin approximation



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

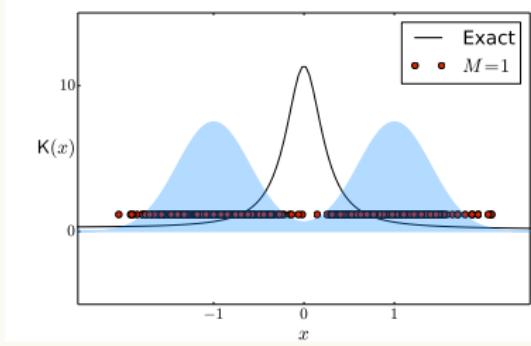
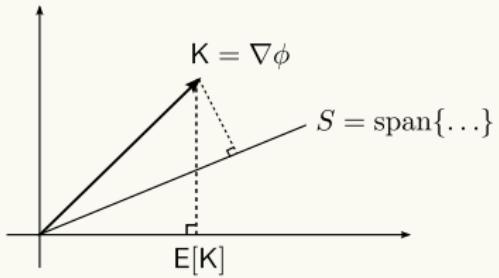


## Non-Gaussian case

### Galerkin approximation



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

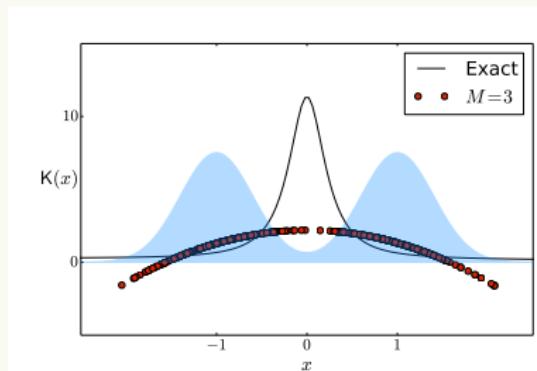
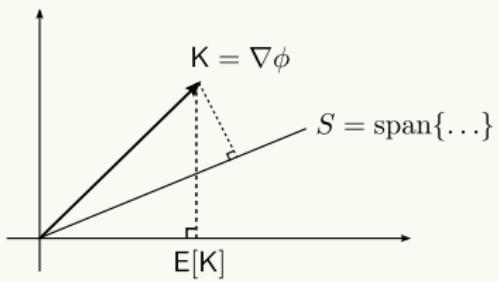


## Non-Gaussian case

### Galerkin approximation



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



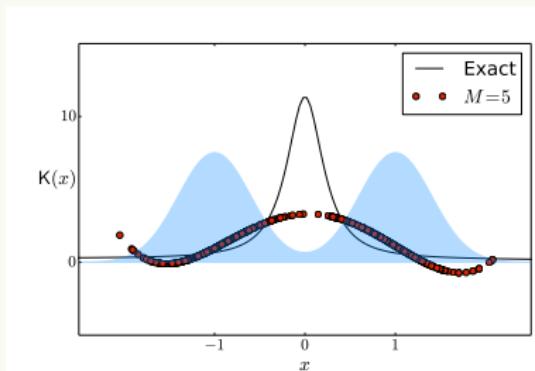
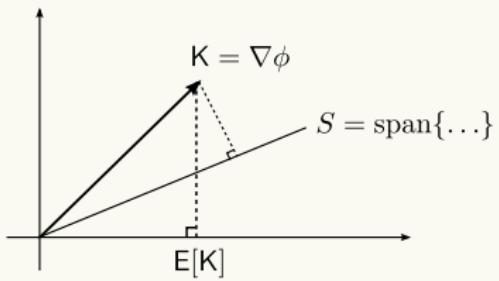
$$\psi \in \{1, x, \dots, x^M\}$$

## Non-Gaussian case

### Galerkin approximation



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



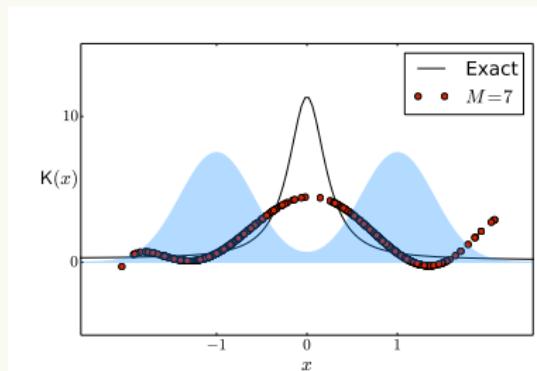
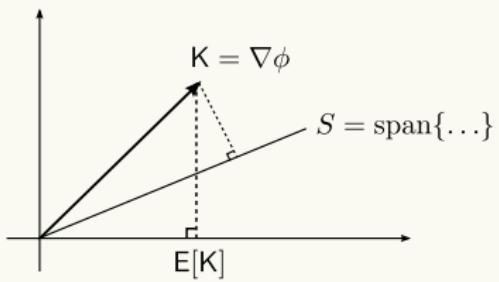
$$\psi \in \{1, x, \dots, x^M\}$$

# Non-Gaussian case

## Galerkin approximation



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



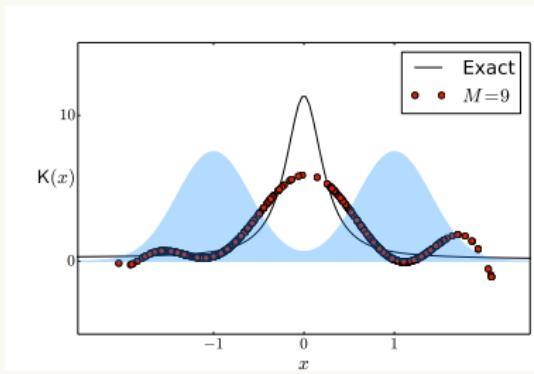
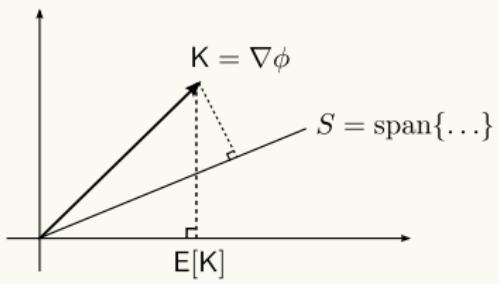
$$\psi \in \{1, x, \dots, x^M\}$$

# Non-Gaussian case

## Galerkin approximation



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



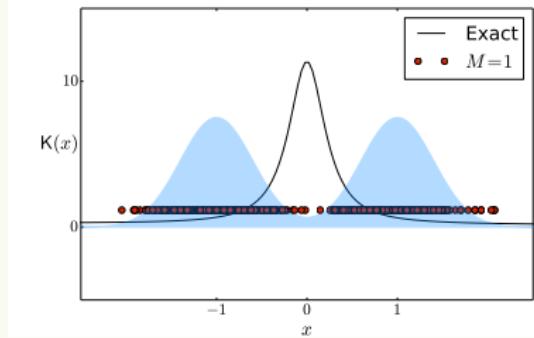
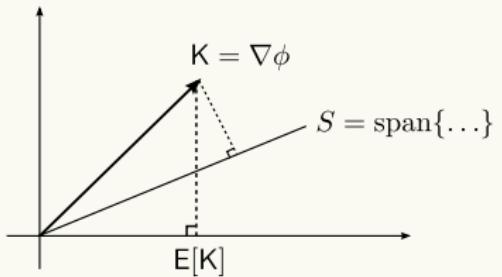
$$\psi \in \{1, x, \dots, x^M\}$$

## Non-Gaussian case

### Galerkin approximation



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



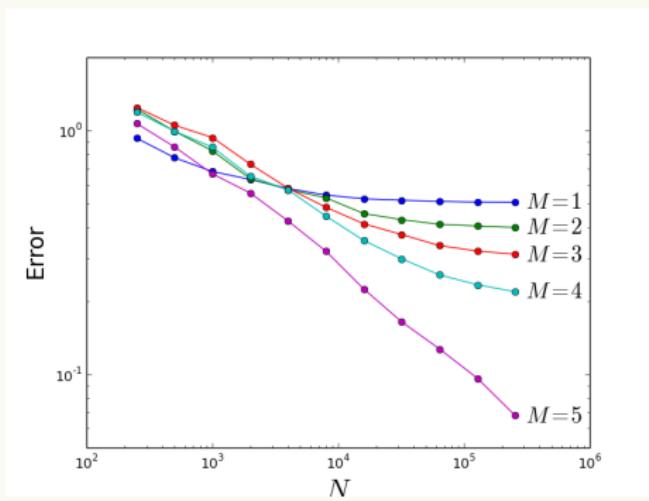
Moral of the story: basis function selection is non-trivial!

## More to the story

### Bias-variance tradeoff

**Special case:** The basis functions are the eigenfunctions of  $\Delta_\rho$

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



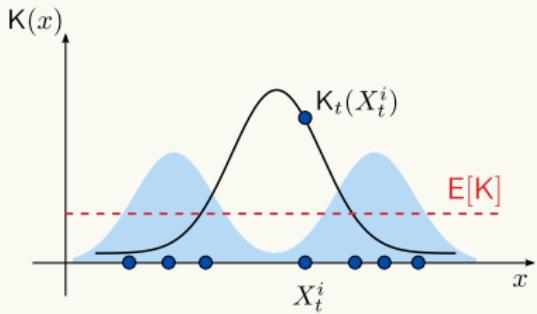


**1** Ensemble Kalman Filter +

**2** Kernel Algorithm

# What are we looking for?

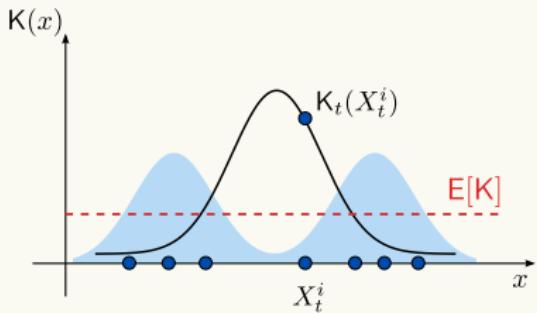
Ensemble Kalman filter +



$$E[K] = \int (h(x) - \hat{h})x \rho(x) dx \approx \frac{1}{N} \sum_{i=1}^N (h(X^i) - \hat{h}) X^i$$

# What are we looking for?

Ensemble Kalman filter +



$$E[K] = \int (h(x) - \hat{h})x\rho(x) dx \approx \frac{1}{N} \sum_{i=1}^N (h(X^i) - \hat{h})X^i$$

Question: Can we improve this approximation?

# Kernel Algorithm

First the punchline



- 1 No basis function selection!
- 2 Simple formula

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

- 3 Reduces to the constant gain in a certain limit

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$



- 1 No basis function selection!
- 2 Simple formula<sup>a</sup>

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

- 3 Reduces to the constant gain in a certain limit

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$

---

<sup>a</sup>Reminiscent of the ensemble transform

# Kernel Algorithm

First the punchline



- 1 No basis function selection!
- 2 Simple formula

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

- 3 Reduces to the constant gain in a certain limit

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$

# Kernel Algorithm

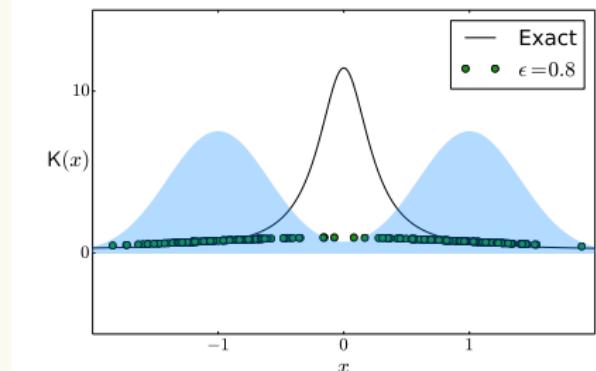
First the punchline

- 1 No basis function selection!
- 2 Simple formula

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

- 3 Reduces to the constant gain in a certain limit

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$



# Kernel Algorithm

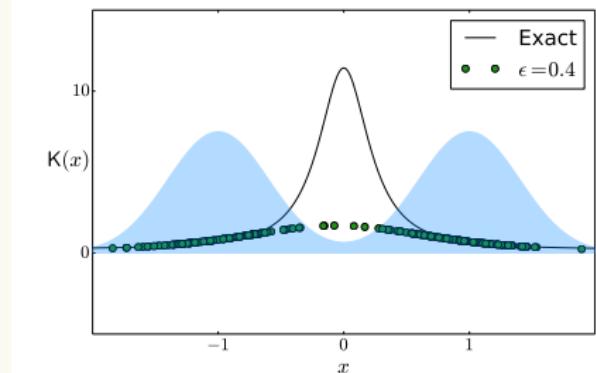
First the punchline

- 1 No basis function selection!
- 2 Simple formula

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

- 3 Reduces to the constant gain in a certain limit

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$



# Kernel Algorithm

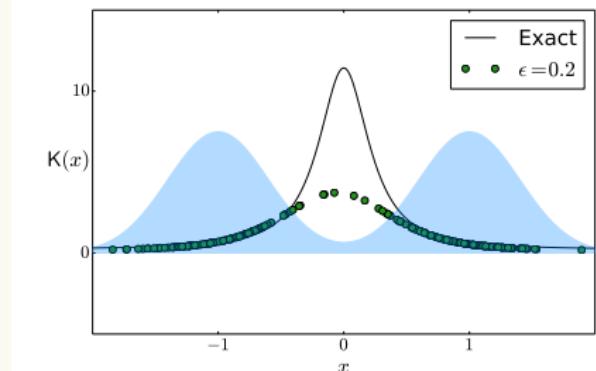
First the punchline

- 1 No basis function selection!
- 2 Simple formula

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

- 3 Reduces to the constant gain in a certain limit

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$



# Kernel Algorithm

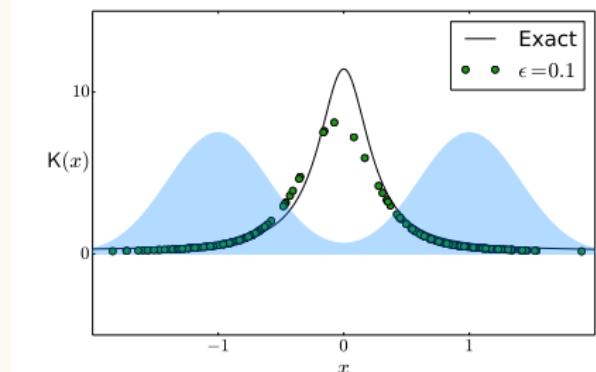
First the punchline

- 1 No basis function selection!
- 2 Simple formula

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

- 3 Reduces to the constant gain in a certain limit

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$





## 1 Ensemble Kalman Filter +

## 2 Kernel Algorithm

- Concept
- Algorithm
- Error analysis

## (2) Kernel Approximation of

These are Markov operators!



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

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

**BVP:**  $\Delta_\rho \phi = -(h - \hat{h})$

**Semigroup:**  $e^{\epsilon \Delta_\rho}$  for  $\epsilon > 0$

## (2) Kernel Approximation of

These are Markov operators!



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

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

**BVP:**  $\Delta_\rho \phi = -(h - \hat{h})$

**Semigroup:**  $e^{\epsilon \Delta_\rho}$  for  $\epsilon > 0$

## (2) Kernel Approximation of

These are Markov operators!



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

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

**BVP:**  $\Delta_\rho \phi = -(h - \hat{h})$

**Semigroup:**  $e^{\epsilon \Delta_\rho}$  for  $\epsilon > 0$

## (2) Kernel Approximation of

These are Markov operators!



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

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

**BVP:**  $\Delta_\rho \phi = -(h - \hat{h})$

**Semigroup:**  $e^{\epsilon \Delta_\rho}$  for  $\epsilon > 0$

**Examples:**

$$(\rho = 1) \quad e^{\epsilon \Delta} f(x) = \int \underbrace{g_\epsilon(x, y)}_{\text{Gaussian}} f(y) \rho(y) dy$$

## (2) Kernel Approximation of

These are Markov operators!



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

**BVP:**  $\Delta_\rho \phi = -(h - \hat{h})$

**Semigroup:**  $e^{\epsilon \Delta_\rho}$  for  $\epsilon > 0$

**Examples:**

$$(\rho = 1) \quad e^{\epsilon \Delta} f(x) = \int \underbrace{g_\epsilon(x, y)}_{\text{Gaussian}} f(y) \rho(y) dy$$

$$(\rho = \text{general}) \quad e^{\epsilon \Delta} f(x) = \int \tilde{k}_\epsilon(x, y) f(y) \rho(y) dy$$



# Kernel Algorithm

## Concept

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

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

Fixed-point equation:

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

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



# Kernel Algorithm

## Concept

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

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

**Fixed-point equation:**

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

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

# Kernel Algorithm

## Concept



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

**Semigroup identity:**  $e^{\epsilon \Delta_\rho} = I + \int_0^\epsilon e^{s \Delta_\rho} \Delta_\rho ds$

**Fixed-point equation:**

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

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

**Kernel representation:**  $\phi(x) = \underbrace{\int \tilde{k}_\epsilon(x, y) \phi(y) \rho(y) dy}_{e^{\epsilon \Delta_\rho} \phi(x)} + \tilde{h}(x)$

# Kernel Algorithm

## Concept



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

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

**Fixed-point equation:**

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

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

**Kernel representation:**  $\phi(x) = \int \tilde{k}_\epsilon(x, y) \phi(y) \rho(y) \, dy + \tilde{h}(x)$

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



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

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

**Fixed-point equation:**

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

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

**Kernel representation:**  $\phi(x) = \int \tilde{k}_\epsilon(x, y) \phi(y) \rho(y) \, dy + \tilde{h}(x)$

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

**Analysis:**  $e^{\epsilon \Delta_\rho}$  is a contraction



# Kernel Algorithm

## Concept

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

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

**Fixed-point equation:**

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

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

**Kernel representation:**  $\phi(x) = \int \tilde{k}_\epsilon(x, y) \phi(y) \rho(y) \, dy + \tilde{h}(x)$

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

**Analysis:**  $e^{\epsilon \Delta_\rho}$  is a contraction

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



## Kernel Algorithm

Heat kernel approximation of the semigroup

**Special case:**  $\rho = 1$

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

where  $g_\epsilon$  is the Gaussian kernel.

In general:

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

where  $n_\epsilon$  is the normalizing constant.

Empirical approximation:

$$e^{\epsilon \Delta_\rho} f(x) \approx \sum_{j=1}^N \frac{1}{n_\epsilon^{(N)}(x)} \frac{g_\epsilon(x, X^j)}{\sqrt{\frac{1}{N} \sum_{l=1}^N g_\epsilon(X^j, X^l)}} f(X^j) \quad (\text{for } N \uparrow \infty)$$

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

---

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,

M. Hein, J. Audibert, U. Von Luxburg, Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

# Kernel Algorithm

## Heat kernel approximation of the semigroup

**Special case:**  $\rho = 1$

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

where  $g_\epsilon$  is the Gaussian kernel.

**In general:**

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

where  $n_\epsilon$  is the normalizing constant.

**Empirical approximation:**

$$e^{\epsilon \Delta_\rho} f(x) \approx \sum_{j=1}^N \frac{1}{n_\epsilon^{(N)}(x)} \frac{g_\epsilon(x, X^j)}{\sqrt{\frac{1}{N} \sum_{l=1}^N g_\epsilon(X^j, X^l)}} f(X^j) \quad (\text{for } N \uparrow \infty)$$

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

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,

M. Hein, J. Audibert, U. Von Luxburg, Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

# Kernel Algorithm

Heat kernel approximation of the semigroup

**Special case:**  $\rho = 1$

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

where  $g_\epsilon$  is the Gaussian kernel.

**In general:**

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

where  $n_\epsilon$  is the normalizing constant.

**Empirical approximation:**

$$e^{\epsilon \Delta_\rho} f(x) \approx \sum_{j=1}^N \frac{1}{n_\epsilon^{(N)}(x)} \frac{g_\epsilon(x, X^j)}{\sqrt{\frac{1}{N} \sum_{l=1}^N g_\epsilon(X^j, X^l)}} f(X^j) \quad (\text{for } N \uparrow \infty)$$

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

R. Coifman, S. Lafon, Diffusion maps, *Applied and computational harmonic analysis*, 2006,

M. Hein, J. Audibert, U. Von Luxburg, Convergence of graph Laplacians on random neighborhood graphs, *JLMR*, 2007

# Kernel-based Algorithm

## Procedure

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

**Output:** Approximate solution  $\phi^{\epsilon, N}$

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

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

- 2 Solve for  $\Phi \in \mathbb{R}^N$ :

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

- 3 Express the approximate solution:

$$\phi^{(\epsilon, N)}(x) := \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \Phi_i + \epsilon(h(x) - \hat{h})$$

# Representation of the gain function

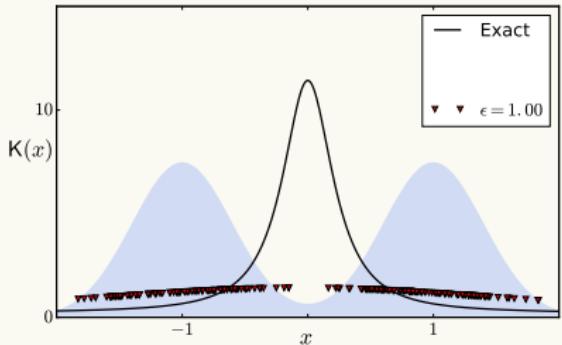


1 Simple formula:

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

2 In the ( $\epsilon = \infty$ ) limit:

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$



# Representation of the gain function

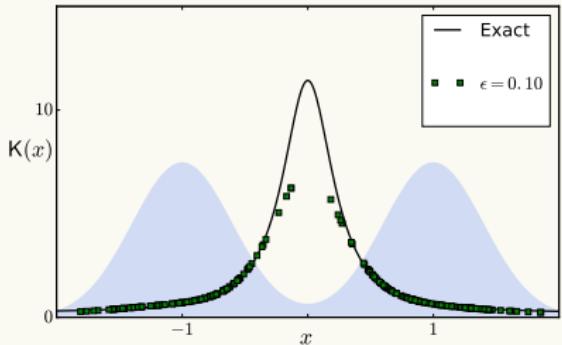


1 Simple formula:

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

2 In the ( $\epsilon = \infty$ ) limit:

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$



# Representation of the gain function

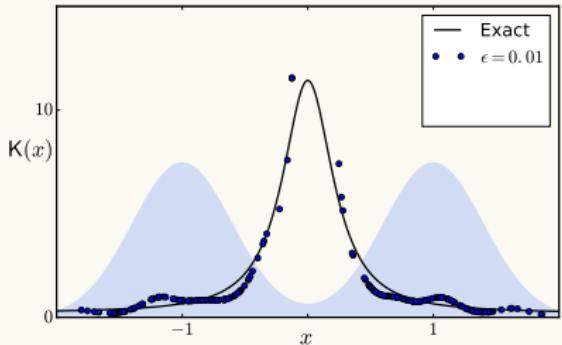


1 Simple formula:

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

2 In the ( $\epsilon = \infty$ ) limit:

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$



# Representation of the gain function

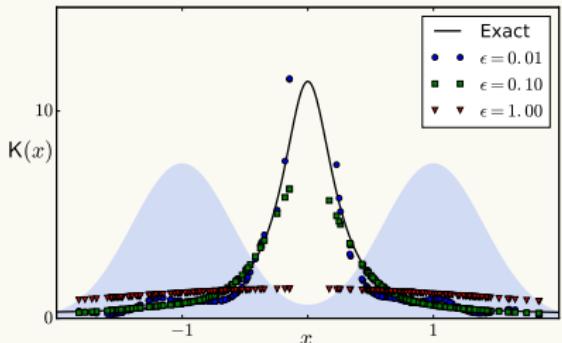


1 Simple formula:

$$K^i = \sum_{j=1}^N s_{ij} X^j$$

2 In the ( $\epsilon = \infty$ ) limit:

$$K^i = \frac{1}{N} \sum_{j=1}^N (h(X^j) - \hat{h}^{(N)}) X^j$$



# Error Analysis Metric



Exact:  $\phi(x) = -\Delta_\rho^{-1} h(x)$

Kernel approx.:  $\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$

Empirical approx.:  $\phi_\epsilon^{(N)}(x) = \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i) + \epsilon h(x)$



Exact:  $\phi(x) = -\Delta_\rho^{-1} h(x)$

Kernel approx.:  $\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$

Empirical approx.:  $\phi_\epsilon^{(N)}(x) = \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i) + \epsilon h(x)$



Exact:  $\phi(x) = -\Delta_\rho^{-1} h(x)$

Kernel approx.:  $\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$

Empirical approx.:  $\phi_\epsilon^{(N)}(x) = \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i) + \epsilon h(x)$



Exact:  $\phi(x) = -\Delta_\rho^{-1} h(x)$

Kernel approx.:  $\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$

Empirical approx.:  $\phi_\epsilon^{(N)}(x) = \frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i) + \epsilon h(x)$

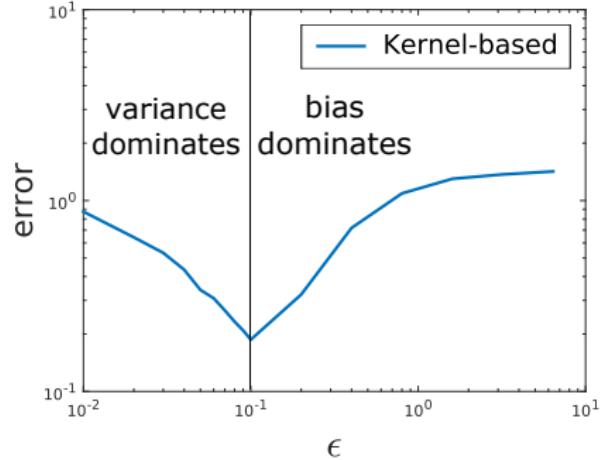
Error metric:

$$\mathbb{E}[\|\phi_\epsilon^{(N)} - \phi\|_{L^2(\rho)}] \leq \underbrace{\mathbb{E}[\|\phi_\epsilon^{(N)} - \phi_\epsilon\|_{L^2(\rho)}]}_{\text{variance}} + \underbrace{\|\phi_\epsilon - \phi\|_{L^2(\rho)}}_{\text{bias}}$$

# Main Result



$$(\text{error}) \leq \underbrace{O\left(\frac{1}{\sqrt{N} \epsilon^{1+d/4}}\right)}_{\text{Variance}} + \underbrace{O(\epsilon)}_{\text{Bias}}$$





- A. Taghvaei, P. Mehta and S. Meyn, [Error Estimates for the Gain Function Approximation in the Feedback Particle Filter](#), In the Procs. of *American Control Conference*, Seattle, May 2017.
- A. Taghvaei and P. Mehta, [Gain Function Approximation for the Feedback Particle Filter](#), In the Procs. of *IEEE Conference on Decision and Control*, Las Vegas, December 2016.
- T. Yang, P. Mehta and S. Meyn. [Feedback particle filter](#). *IEEE Trans. Automat. Control* 58(10):2465-2480 (2013).
- T. Yang, R. Laugesen, P. Mehta and S. Meyn. Multivariable feedback particle filter. *Automatica* 71:10-23 (2016).
- R. Laugesen, P. Mehta, S. Meyn and M. Raginsky. [Poisson's equation in nonlinear filtering](#). *SIAM J. Control Optimiz.* 53(1):501-525 (2015).
- A. Taghvaei and P. Mehta, [An Optimal Transport Formulation of the Linear Feedback Particle Filter](#). In the *Proceedings of American Control Conference*, Boston, July 2016.



# Error Analysis

## Bias

$$\begin{aligned}
 & \text{(exact)} & \phi(x) = -\Delta_\rho^{-1} h(x) \\
 & \text{(kernel approx.)} & \phi_\epsilon(x) = \underbrace{\frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy}_{T_\epsilon \phi_\epsilon(x)} + \epsilon h(x) \\
 & & \phi_\epsilon = (I - T_\epsilon)^{-1} \epsilon h
 \end{aligned}$$

Proof steps:

- 1  $T_\epsilon$  is a Markov operator with finite invariant measure
- 2  $T_\epsilon$  satisfies the (geometric ergodic) Lyapunov criteria

$$\|T_\epsilon\|_{L^2(\rho)} \leq 1 - \epsilon \lambda + O(\epsilon^2)$$

- 3  $T_\epsilon$  has the Taylor expansion

$$T_\epsilon f = f + \epsilon \Delta_\rho f + O(\epsilon^2)$$

$$\therefore \|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq \underbrace{\|(I - T_\epsilon)^{-1}\|_{L^2(\rho)}}_{O(\frac{1}{\epsilon})} \underbrace{\|(T_\epsilon - I - \epsilon \Delta_\rho)\phi\|_{L^2(\rho)}}_{O(\epsilon^2)} \leq O(\epsilon)$$



$$\begin{aligned}
 & \text{(exact)} & \phi(x) = -\Delta_\rho^{-1} h(x) \\
 & \text{(kernel approx.)} & \phi_\epsilon(x) = \underbrace{\frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy}_{T_\epsilon \phi_\epsilon(x)} + \epsilon h(x) \\
 & & \phi_\epsilon = (I - T_\epsilon)^{-1} \epsilon h
 \end{aligned}$$

Proof steps:

- 1  $T_\epsilon$  is a Markov operator with finite invariant measure
- 2  $T_\epsilon$  satisfies the (geometric ergodic) Lyapunov criteria

$$\|T_\epsilon\|_{L^2(\rho)} \leq 1 - \epsilon \lambda + O(\epsilon^2)$$

- 3  $T_\epsilon$  has the Taylor expansion

$$T_\epsilon f = f + \epsilon \Delta_\rho f + O(\epsilon^2)$$

$$\therefore \|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq \underbrace{\|(I - T_\epsilon)^{-1}\|_{L^2(\rho)}}_{O(\frac{1}{\epsilon})} \underbrace{\|(T_\epsilon - I - \epsilon \Delta_\rho)\phi\|_{L^2(\rho)}}_{O(\epsilon^2)} \leq O(\epsilon)$$



$$\begin{aligned}
 & \text{(exact)} & \phi(x) = -\Delta_\rho^{-1} h(x) \\
 & \text{(kernel approx.)} & \phi_\epsilon(x) = \underbrace{\frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy}_{T_\epsilon \phi_\epsilon(x)} + \epsilon h(x) \\
 & & \phi_\epsilon = (I - T_\epsilon)^{-1} \epsilon h
 \end{aligned}$$

Proof steps:

- 1  $T_\epsilon$  is a Markov operator with finite invariant measure
- 2  $T_\epsilon$  satisfies the (geometric ergodic) Lyapunov criteria

$$\|T_\epsilon\|_{L^2(\rho)} \leq 1 - \epsilon \lambda + O(\epsilon^2)$$

- 3  $T_\epsilon$  has the Taylor expansion

$$T_\epsilon f = f + \epsilon \Delta_\rho f + O(\epsilon^2)$$

$$\therefore \|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq \underbrace{\|(I - T_\epsilon)^{-1}\|_{L^2(\rho)}}_{O(\frac{1}{\epsilon})} \underbrace{\|(T_\epsilon - I - \epsilon \Delta_\rho)\phi\|_{L^2(\rho)}}_{O(\epsilon^2)} \leq O(\epsilon)$$



# Error Analysis

## Bias

$$\begin{aligned}
 & \text{(exact)} & \phi(x) = -\Delta_\rho^{-1} h(x) \\
 & \text{(kernel approx.)} & \phi_\epsilon(x) = \underbrace{\frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy}_{T_\epsilon \phi_\epsilon(x)} + \epsilon h(x) \\
 & & \phi_\epsilon = (I - T_\epsilon)^{-1} \epsilon h
 \end{aligned}$$

Proof steps:

- 1  $T_\epsilon$  is a Markov operator with finite invariant measure
- 2  $T_\epsilon$  satisfies the (geometric ergodic) Lyapunov criteria

$$\|T_\epsilon\|_{L^2(\rho)} \leq 1 - \epsilon \lambda + O(\epsilon^2)$$

- 3  $T_\epsilon$  has the Taylor expansion

$$T_\epsilon f = f + \epsilon \Delta_\rho f + O(\epsilon^2)$$

$$\therefore \|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq \underbrace{\|(I - T_\epsilon)^{-1}\|_{L^2(\rho)}}_{O(\frac{1}{\epsilon})} \underbrace{\|(T_\epsilon - I - \epsilon \Delta_\rho)\phi\|_{L^2(\rho)}}_{O(\epsilon^2)} \leq O(\epsilon)$$



$$\begin{aligned}
 & \text{(exact)} & \phi(x) = -\Delta_\rho^{-1} h(x) \\
 & \text{(kernel approx.)} & \phi_\epsilon(x) = \underbrace{\frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy}_{T_\epsilon \phi_\epsilon(x)} + \epsilon h(x) \\
 & & \phi_\epsilon = (I - T_\epsilon)^{-1} \epsilon h
 \end{aligned}$$

Proof steps:

- 1  $T_\epsilon$  is a Markov operator with finite invariant measure
- 2  $T_\epsilon$  satisfies the (geometric ergodic) Lyapunov criteria

$$\|T_\epsilon\|_{L^2(\rho)} \leq 1 - \epsilon \lambda + O(\epsilon^2)$$

- 3  $T_\epsilon$  has the Taylor expansion

$$T_\epsilon f = f + \epsilon \Delta_\rho f + O(\epsilon^2)$$

$$\therefore \|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq \underbrace{\|(I - T_\epsilon)^{-1}\|_{L^2(\rho)}}_{O(\frac{1}{\epsilon})} \underbrace{\|(T_\epsilon - I - \epsilon \Delta_\rho)\phi\|_{L^2(\rho)}}_{O(\epsilon^2)} \leq O(\epsilon)$$



# Error Analysis

## Bias

$$\begin{aligned}
 & \text{(exact)} \quad \phi(x) = -\Delta_\rho^{-1} h(x) \\
 & \text{(kernel approx.)} \quad \phi_\epsilon(x) = \underbrace{\frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy}_{T_\epsilon \phi_\epsilon(x)} + \epsilon h(x) \\
 & \quad \phi_\epsilon = (I - T_\epsilon)^{-1} \epsilon h
 \end{aligned}$$

Proof steps:

- 1  $T_\epsilon$  is a Markov operator with finite invariant measure
- 2  $T_\epsilon$  satisfies the (geometric ergodic) Lyapunov criteria

$$\|T_\epsilon\|_{L^2(\rho)} \leq 1 - \epsilon \lambda + O(\epsilon^2)$$

- 3  $T_\epsilon$  has the Taylor expansion

$$T_\epsilon f = f + \epsilon \Delta_\rho f + O(\epsilon^2)$$

$$\therefore \|\phi_\epsilon - \phi\|_{L^2(\rho)} \leq \underbrace{\|(I - T_\epsilon)^{-1}\|_{L^2(\rho)}}_{O(\frac{1}{\epsilon})} \underbrace{\|(T_\epsilon - I - \epsilon \Delta_\rho)\phi\|_{L^2(\rho)}}_{O(\epsilon^2)} \leq O(\epsilon)$$

# Error Analysis

## Variance



(kernel approx.)

$$\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$$

(empirical approx.)

$$\phi_\epsilon^{(N)}(x) = \underbrace{\frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i)}_{T_\epsilon^{(N)} \phi_\epsilon(x)} + \epsilon h(x)$$

Proof steps:

1  $T_\epsilon : L^2(\rho) \rightarrow L^2(\rho)$  is a compact operator

2  $T_\epsilon^{(N)}$  converges pointwise to  $T_\epsilon$

$$\lim_{N \rightarrow \infty} \mathbb{E} \|T_\epsilon^{(N)} f - T_\epsilon f\|_{L^2(\rho)} = 0, \quad \forall f \in L^2(\rho)$$

3 The sequence of operators  $\{T_\epsilon^{(N)}\}_{N=1}^\infty$  are collectively compact

$$\therefore \lim_{N \rightarrow \infty} \mathbb{E} \left\| \left( \frac{I - T_\epsilon^{(N)}}{\epsilon} \right)^{-1} h - \left( \frac{I - T_\epsilon}{\epsilon} \right)^{-1} h \right\|_{L^2(\rho)} = 0, \quad \forall h \in L^2(\rho)$$

# Error Analysis

## Variance



(kernel approx.)

$$\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$$

(empirical approx.)

$$\phi_\epsilon^{(N)}(x) = \underbrace{\frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i)}_{T_\epsilon^{(N)} \phi_\epsilon(x)} + \epsilon h(x)$$

Proof steps:

1  $T_\epsilon : L^2(\rho) \rightarrow L^2(\rho)$  is a compact operator

2  $T_\epsilon^{(N)}$  converges pointwise to  $T_\epsilon$

$$\lim_{N \rightarrow \infty} \mathbb{E} \|T_\epsilon^{(N)} f - T_\epsilon f\|_{L^2(\rho)} = 0, \quad \forall f \in L^2(\rho)$$

3 The sequence of operators  $\{T_\epsilon^{(N)}\}_{N=1}^\infty$  are collectively compact

$$\therefore \lim_{N \rightarrow \infty} \mathbb{E} \left\| \left( \frac{I - T_\epsilon^{(N)}}{\epsilon} \right)^{-1} h - \left( \frac{I - T_\epsilon}{\epsilon} \right)^{-1} h \right\|_{L^2(\rho)} = 0, \quad \forall h \in L^2(\rho)$$

# Error Analysis

## Variance



(kernel approx.)

$$\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$$

(empirical approx.)

$$\phi_\epsilon^{(N)}(x) = \underbrace{\frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i)}_{T_\epsilon^{(N)} \phi_\epsilon(x)} + \epsilon h(x)$$

Proof steps:

1  $T_\epsilon : L^2(\rho) \rightarrow L^2(\rho)$  is a compact operator

2  $T_\epsilon^{(N)}$  converges pointwise to  $T_\epsilon$

$$\lim_{N \rightarrow \infty} \mathbb{E} \|T_\epsilon^{(N)} f - T_\epsilon f\|_{L^2(\rho)} = 0, \quad \forall f \in L^2(\rho)$$

3 The sequence of operators  $\{T_\epsilon^{(N)}\}_{N=1}^\infty$  are collectively compact

$$\therefore \lim_{N \rightarrow \infty} \mathbb{E} \left\| \left( \frac{I - T_\epsilon^{(N)}}{\epsilon} \right)^{-1} h - \left( \frac{I - T_\epsilon}{\epsilon} \right)^{-1} h \right\|_{L^2(\rho)} = 0, \quad \forall h \in L^2(\rho)$$



(kernel approx.)

$$\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$$

(empirical approx.)

$$\phi_\epsilon^{(N)}(x) = \underbrace{\frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i)}_{T_\epsilon^{(N)} \phi_\epsilon(x)} + \epsilon h(x)$$

Proof steps:

- 1  $T_\epsilon : L^2(\rho) \rightarrow L^2(\rho)$  is a compact operator
- 2  $T_\epsilon^{(N)}$  converges pointwise to  $T_\epsilon$

$$\lim_{N \rightarrow \infty} \mathbb{E} \|T_\epsilon^{(N)} f - T_\epsilon f\|_{L^2(\rho)} = 0, \quad \forall f \in L^2(\rho)$$

- 3 The sequence of operators  $\{T_\epsilon^{(N)}\}_{N=1}^\infty$  are collectively compact

$$\therefore \lim_{N \rightarrow \infty} \mathbb{E} \left\| \left( \frac{I - T_\epsilon^{(N)}}{\epsilon} \right)^{-1} h - \left( \frac{I - T_\epsilon}{\epsilon} \right)^{-1} h \right\|_{L^2(\rho)} = 0, \quad \forall h \in L^2(\rho)$$

# Error Analysis

## Variance



(kernel approx.)

$$\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$$

(empirical approx.)

$$\phi_\epsilon^{(N)}(x) = \underbrace{\frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i)}_{T_\epsilon^{(N)} \phi_\epsilon(x)} + \epsilon h(x)$$

Proof steps:

1  $T_\epsilon : L^2(\rho) \rightarrow L^2(\rho)$  is a compact operator

2  $T_\epsilon^{(N)}$  converges pointwise to  $T_\epsilon$

$$\lim_{N \rightarrow \infty} \mathbb{E} \|T_\epsilon^{(N)} f - T_\epsilon f\|_{L^2(\rho)} = 0, \quad \forall f \in L^2(\rho)$$

3 The sequence of operators  $\{T_\epsilon^{(N)}\}_{N=1}^\infty$  are collectively compact

$$\therefore \lim_{N \rightarrow \infty} \mathbb{E} \left\| \left( \frac{I - T_\epsilon^{(N)}}{\epsilon} \right)^{-1} h - \left( \frac{I - T_\epsilon}{\epsilon} \right)^{-1} h \right\|_{L^2(\rho)} = 0, \quad \forall h \in L^2(\rho)$$

# Error Analysis

## Variance



(kernel approx.)

$$\phi_\epsilon(x) = \frac{1}{n_\epsilon(x)} \int k_\epsilon(x, y) \phi_\epsilon(y) \rho(y) dy + \epsilon h(x)$$

(empirical approx.)

$$\phi_\epsilon^{(N)}(x) = \underbrace{\frac{1}{n_\epsilon^{(N)}(x)} \sum_{i=1}^N k_\epsilon^{(N)}(x, X^i) \phi_\epsilon(X^i)}_{T_\epsilon^{(N)} \phi_\epsilon(x)} + \epsilon h(x)$$

Proof steps:

- 1  $T_\epsilon : L^2(\rho) \rightarrow L^2(\rho)$  is a compact operator
- 2  $T_\epsilon^{(N)}$  converges pointwise to  $T_\epsilon$

$$\lim_{N \rightarrow \infty} \mathbb{E} \|T_\epsilon^{(N)} f - T_\epsilon f\|_{L^2(\rho)} = 0, \quad \forall f \in L^2(\rho)$$

- 3 The sequence of operators  $\{T_\epsilon^{(N)}\}_{N=1}^\infty$  are collectively compact

$$\therefore \lim_{N \rightarrow \infty} \mathbb{E} \left\| \left( \frac{I - T_\epsilon^{(N)}}{\epsilon} \right)^{-1} h - \left( \frac{I - T_\epsilon}{\epsilon} \right)^{-1} h \right\|_{L^2(\rho)} = 0, \quad \forall h \in L^2(\rho)$$