Numerical Methods for Solving Poisson Equation with Applications in Filtering and Classification

4th Workshop on Cognition and Control
Gainesville, Florida, Jan 16-17, 2016

Amirhossein Taghvaei
Joint work with P. G. Mehta

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

Jan 16, 2016

ILLINOIS

Outline

1. Poisson equation
2. Poisson equation in nonlinear filtering
3. Numerical solution: Galerkin method
4. Spectral clustering
5. Numerical solution: Graph Laplacian based method
6. Poisson equation in classification

Poisson equation: A real-valued function ϕ satisfies Poisson equation if,

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

where

- ρ is a probability density function
- h is a real-valued function
- $\hat{h} = \int h\rho \,\mathrm{d}x$

In practice: ρ is not given. Only $X^1, \ldots, X^N \overset{\text{i.i.d}}{\sim} \rho$ are known.

Objective: Design an algorithm with output $\phi^{(N)}$ s.t $\phi^{(N)} \approx \phi$

Motivation:

1. Simulation and optimization theory for Markov dels [Meyn, Tweedie, 2012]
2. Nonlinear filtering [Yang, et. al. 2015]

Motivation: Nonlinear Filtering

Problem:

$$\textbf{Signal model:} \quad \mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \qquad X_0 \sim p_0(\cdot)$$

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

What is posterior distribution of X_t given $\mathcal{Z}_t := \sigma(Z_s : 0 \leq s \leq t)$, i.e

$$\mathsf{P}(X_t|\mathcal{Z}_t) = ?$$

Solution:

- Linear and Gaussian: Kalman filter
- Nonlinear and non-Gaussian: (Approximate solutions) Extended Kalman filter, particle filter, **feedback particle filter**, ...

Problem:

$$\textbf{Signal model:} \quad \mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \qquad X_0 \sim p_0(\cdot)$$
$$\textbf{Observation model:} \quad \mathrm{d}Z_t = h(X_t)\,\mathrm{d}t + \mathrm{d}W_t$$

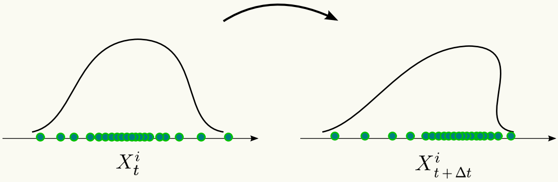
What is posterior distribution of X_t given $\mathcal{Z}_t := \sigma(Z_s : 0 \leq s \leq t)$, i.e

$$\mathsf{P}(X_t|\mathcal{Z}_t) =?$$

Solution:

- Linear and Gaussian: Kalman filter
- Nonlinear and non-Gaussian: (Approximate solutions) Extended Kalman filter, particle filter, **feedback particle filter**, ...

Feedback Particle Filter



Idea:

- Approximate $\mathsf{P}(X_t|\mathcal{Z}_t)$ with particles $\{X^1, \ldots, X^N\}$
- Update particles with a control law s.t

$$X_t^i \sim \mathsf{P}(X_t|\mathcal{Z}_t), \quad \forall t > 0$$

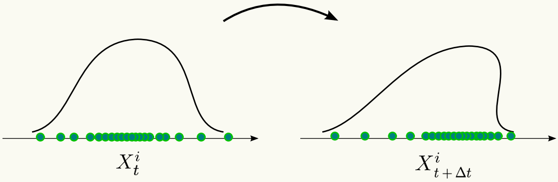
Algorithm:

$$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \mathrm{d}B_t^i + \mathsf{K}(X_t^i) \circ \left(\mathrm{d}Z_t - \frac{h(X_t^i) + \hat{h}}{2}\,\mathrm{d}t \right), \quad \text{for} \quad i = 1, \ldots, N$$

- $\hat{h} = \mathsf{E}[h(X_t)|\mathcal{Z}_t] \approx \frac{1}{N} \sum h(X_t^i)$
- $\mathsf{K}(x) = \nabla \phi(x)$, where ϕ is the solution to the Poisson equation

T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter, *Automatica*, 2015

Feedback Particle Filter



Idea:

- Approximate $\mathsf{P}(X_t|\mathcal{Z}_t)$ with particles $\{X^1, \ldots, X^N\}$
- Update particles with a control law s.t

$$X_t^i \sim \mathsf{P}(X_t|\mathcal{Z}_t), \quad \forall t > 0$$

Algorithm:

$$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \mathrm{d}B_t^i + \mathsf{K}(X_t^i) \circ \left(\mathrm{d}Z_t - \frac{h(X_t^i) + \hat{h}}{2}\,\mathrm{d}t \right), \quad \text{for} \quad i = 1, \ldots, N$$

- $\hat{h} = \mathsf{E}[h(X_t)|\mathcal{Z}_t] \approx \frac{1}{N} \sum h(X_t^i)$
- $\mathsf{K}(x) = \nabla \phi(x)$, where ϕ is the solution to the Poisson equation

T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter, *Automatica*, 2015

Poisson Equation in Nonlinear Filtering

Poisson equation: A real-valued function ϕ satisfies Poisson equation if,

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

where

- ρ is the density of the posterior $\mathsf{P}(X_t|\mathcal{Z}_t)$,
- h is the observation function
- $\hat{h} = \int h\rho\,\mathrm{d}x$

In practice: Only $X^1,\ldots,X^N \sim \mathsf{P}(X_t|\mathcal{Z}_t)$ are given

Objective: Design an algorithm with output $\phi^{(N)}$ s.t $\phi^{(N)} \approx \phi$

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

Poisson Equation: Existence and Uniqueness of the Solution

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

Hilbert space:

$$H_0^1(\mathbb{R}^d;\rho\,\mathrm{d}x) := \left\{\phi : \mathbb{R}^d \to \mathbb{R} \mid \phi \in L^2_\rho,\ \frac{\partial\phi}{\partial x_i} \in L^2_\rho,\ \int \phi\rho\,\mathrm{d}x = 0\right\}$$

Weak form: $\phi \in H_0^1(\mathbb{R}^d;\rho\,\mathrm{d}x)$ is the weak solution of the Poisson equation if,

$$\int_{\mathbb{R}^d}(\nabla\phi\cdot\nabla\psi)\rho\,\mathrm{d}x = \int_{\mathbb{R}^d}(h-\hat{h})\psi\rho\,\mathrm{d}x, \quad \forall\psi \in H_0^1(\mathbb{R}^d;\rho\,\mathrm{d}x)$$

Existence and uniqueness: A unique weak solution in $H_0^1(\mathbb{R}^d;\rho\,\mathrm{d}x)$ exists if,

1. $h \in L^2_\rho$
2. ρ satisfies the Poincaré inequality

Poincaré inequality: $\exists\lambda > 0$ s.t

$$\int \phi^2\rho\,\mathrm{d}x \le \frac{1}{\lambda}\int |\nabla\phi|^2\rho\,\mathrm{d}x, \quad \forall\phi \in H_0^1(\mathbb{R}^d;\rho\,\mathrm{d}x)$$

Weak form:

$$\int_{\mathbb{R}^d} (\nabla \phi \cdot \nabla \psi) \rho \, \mathrm{d}x = \int_{\mathbb{R}^d} (h - \hat{h}) \psi \rho \, \mathrm{d}x, \quad \forall \psi \in H_0^1(\mathbb{R}^d; \rho \, \mathrm{d}x)$$

Galerkin Method:

1. Write ϕ as linear combination of basis functions

$$\phi = c_1 \psi_1 + \ldots + c_M \psi_M$$

2. Construct a finite dimensional approximation of the weak form

$$Ac = b$$

where

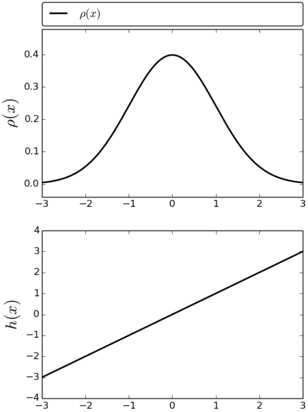
$$A_{ml} = \int_{\mathbb{R}^d} (\nabla \psi_m \cdot \nabla \psi_l) \rho \, \mathrm{d}x, \quad b_m = \int_{\mathbb{R}^d} (h - \hat{h}) \psi_m \rho \, \mathrm{d}x$$

3. Solve the system of M linear equations for $c = [c_1, \ldots, c_M]^T$

Issues:

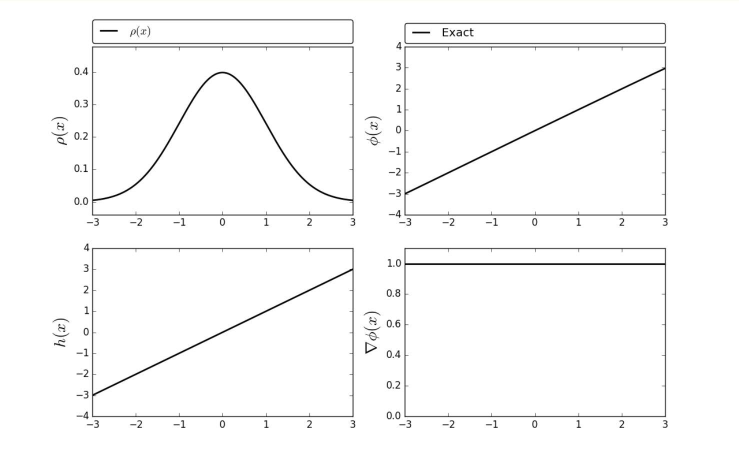
1. Choice of the basis functions
2. M grows as d grows

Example I: $X \sim N(0,1)$ and $h(x) = x$,



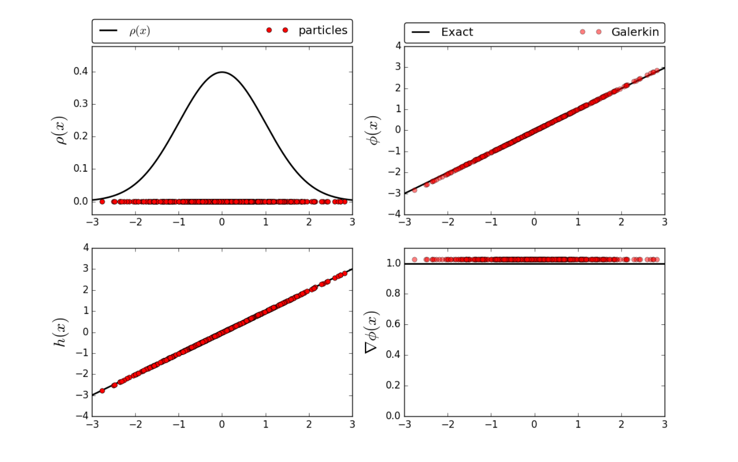
Example I: $X \sim N(0,1)$ and $h(x) = x$,

$$-e^{\frac{x^2}{2}}\frac{\mathrm{d}}{\mathrm{d}x}(e^{-\frac{x^2}{2}}\frac{\mathrm{d}\phi}{\mathrm{d}x}) = x, \quad \Rightarrow \quad \phi = x$$



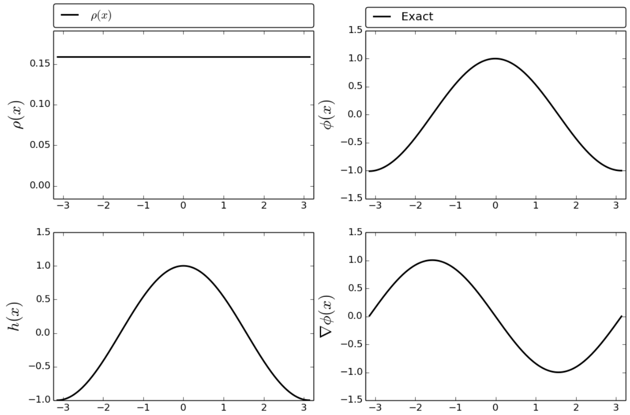
Example I: $X \sim N(0,1)$ and $h(x) = x$,

Galerkin method with basis $= \{x\}$



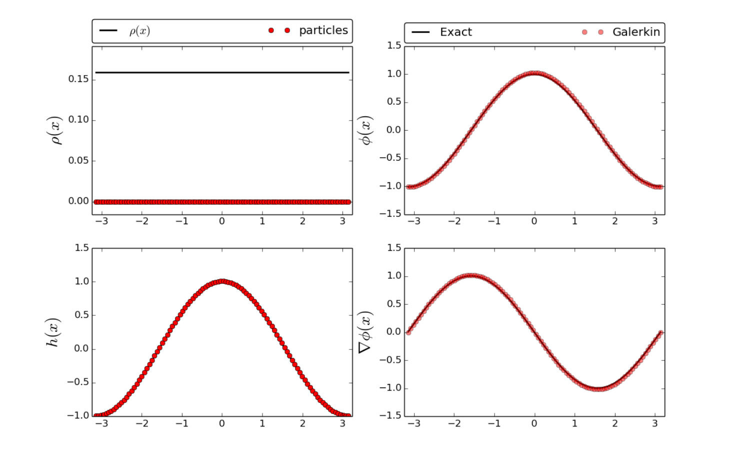
Example II: $X \sim$ unif $[-\pi, \pi]$ and $h(x) = \cos(x)$,

$$-2\pi \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2\pi}\frac{\mathrm{d}\phi}{\mathrm{d}x}\right) = \cos(x), \quad \Rightarrow \quad \phi = \cos(x)$$



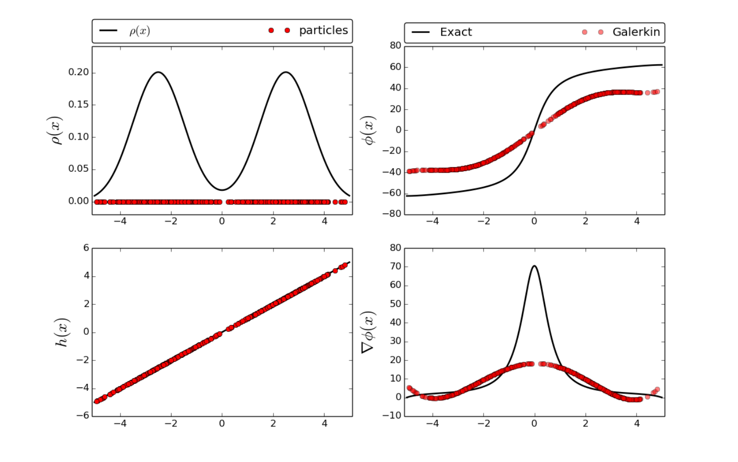
Example II: $X \sim$ unif $[-\pi, \pi]$ and $h(x) = \cos(x)$,

Galerkin method with basis= $\{\cos(x), \sin(x)\}$



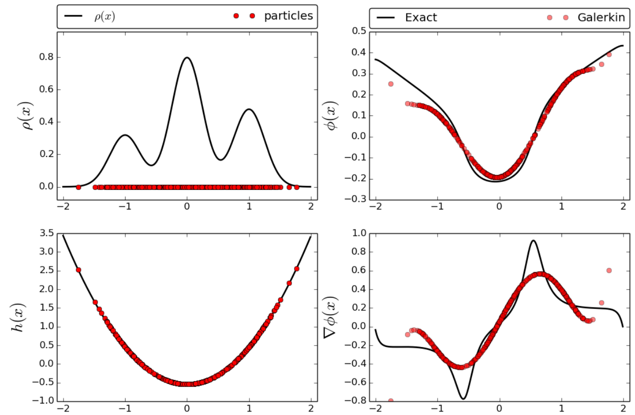
Example III: $X \sim N(-2.5,1) + N(2.5,1)$ and $h(x) = x$,

Galerkin method with basis= $\{x, x^2, \ldots, x^6\}$



Example IV: $X \sim \frac{2}{10}N(-1,\frac{1}{4}) + \frac{5}{10}N(0,\frac{1}{4}) + \frac{3}{10}N(1,\frac{1}{4})$ and $h(x) = x$,

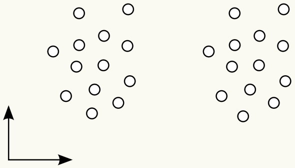
Galerkin method with basis= $\{x, x^2, \ldots, x^6\}$



Outline

1. Poisson equation
2. Poisson equation in nonlinear filtering
3. Numerical solution: Galerkin method
4. Spectral clustering
5. Numerical solution: Graph Laplacian based method
6. Poisson equation in classification

Spectral Clustering



Unlabeled data: $\{X^1, \ldots, X^N\}$

Objective: Identify groups of data that are similar

Spectral method:

1. Construct a graph where nodes are data points
2. Specify weights W_{ij} for edges
3. Construct the graph Laplacian matrix

$$L = I - D^{-1}W$$

4. Cluster according to the eigenvectors of L

U. Von Luxburg. "A tutorial on spectral clustering". Statistics and computing, 2007

Gaussian Kernel

Gaussian kernel: $K_{ij} = \exp(-\frac{|X^i - X^j|^2}{4\epsilon})$

	Weights	Limit of L as $N \to \infty$
[Belkin, 2003]	$W_{ij} = K_{ij}$	$-\Delta + O(\epsilon), \quad (*)$
[Coifman, Lafon, 2006]	$W_{ij} = \frac{K_{ij}}{(\sum_l K_{il})(\sum_l K_{jl})}$	$-\Delta + O(\epsilon)$
[Hein, 2007]	$W_{ij} = \frac{K_{ij}}{(\sum_l K_{il})^\lambda(\sum_l K_{jl})^\lambda}$	$-\frac{1}{\rho^s}\nabla \cdot (\rho^s \nabla \quad) + O(\epsilon), \quad (**)$

- $(*)$ True only if $X \sim$ unif
- $(**)$ $s = 2(\lambda - 1)$

Intresting case: $s = 1$

Weighted Laplacian: $-\frac{1}{\rho}\nabla \cdot (\rho \nabla \quad)$

Review paper on weighted Laplacian [Grigoryan, 2007]

Numerical Solution of the Poisson Equation

Graph Laplacian Based Method

Main idea:

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

Procedure:

1. Construct the weight and degree matrix

$$W_{ij} = \frac{K_{ij}}{(\sum_l K_{il})^{\frac{1}{2}}(\sum_l K_{jl})^{\frac{1}{2}}}, \quad D_{ij} = \delta_{ij}\sum_l W_{il}, \quad A := D^{-1}W$$

2. Construct the graph Laplacian matrix

$$L = \frac{I - D^{-1}W}{\epsilon}$$

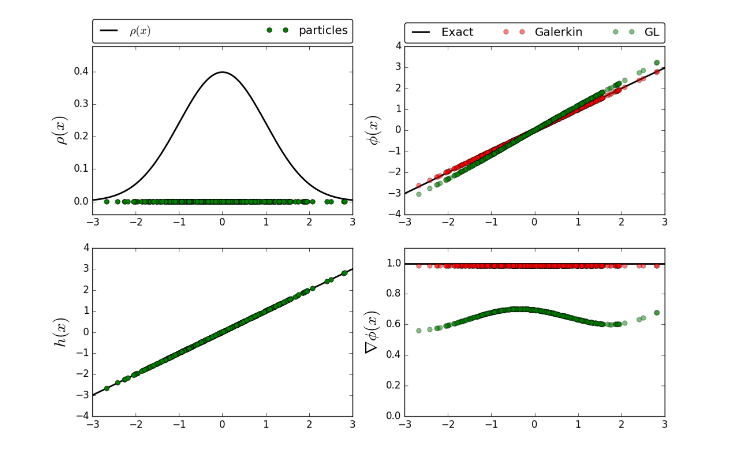
3. Solve a system of N linear equations

$$L\phi = h - \hat{h}$$

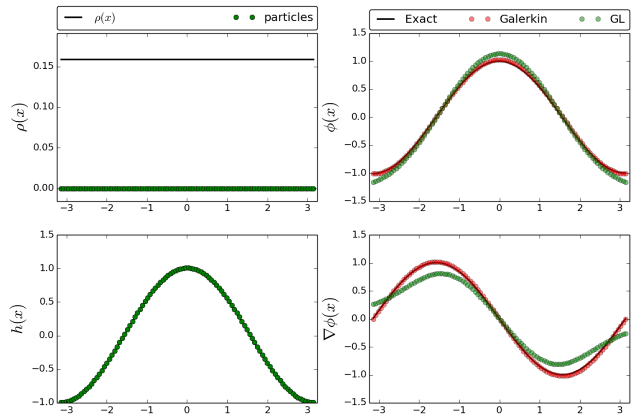
4. Approximate $\nabla\phi$

$$\frac{\partial\phi}{\partial x_i} \approx \frac{1}{2\epsilon}\left(A(x_i\phi) - A(x_i)A(\phi)\right)$$

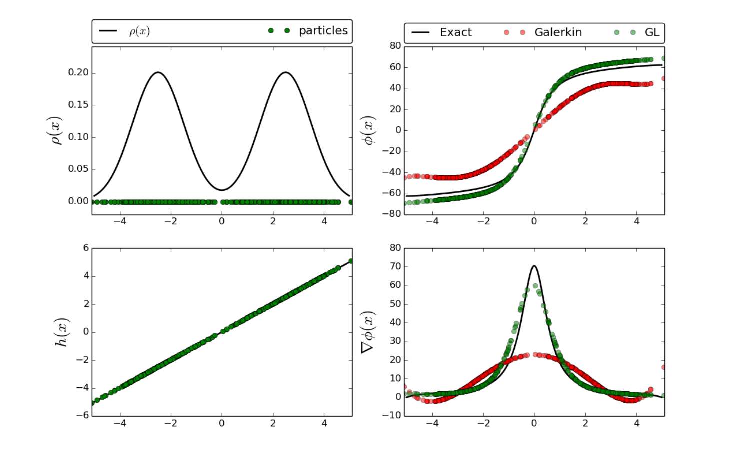
Example I: $X \sim N(0,1)$ and $h(x) = x$,



Example II: $X \sim$ unif $[-\pi, \pi]$ and $h(x) = \cos(x)$,

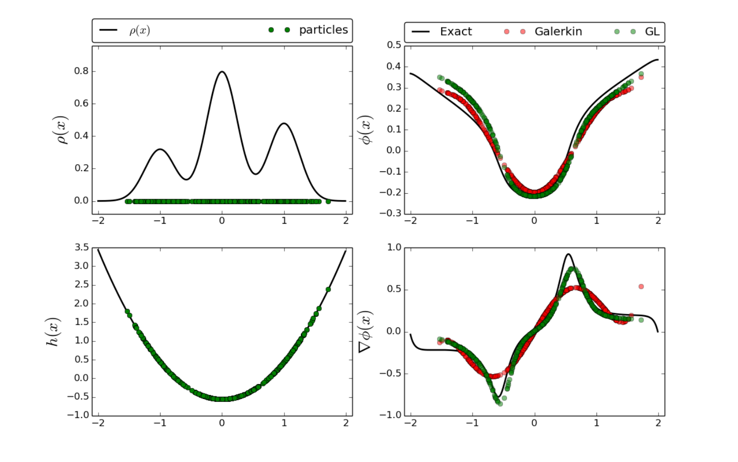


Example III: $X \sim N(-2.5, 1) + N(2.5, 1)$ and $h(x) = x$,



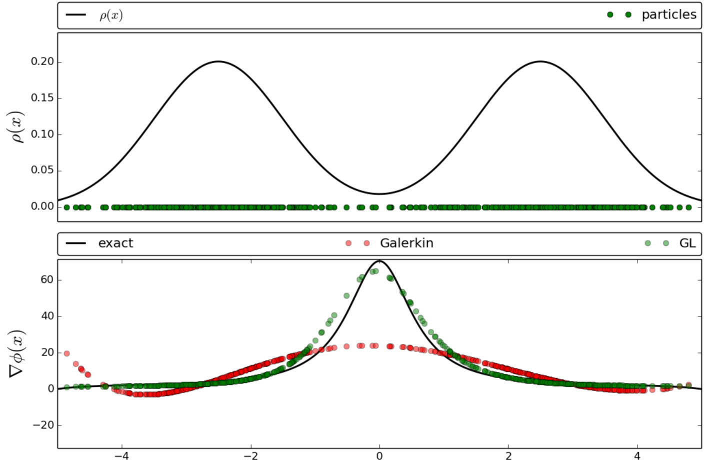
Numerics
Graph Laplacian method

Example IV: $X \sim \frac{2}{10}N(-1,\frac{1}{4}) + \frac{5}{10}N(0,\frac{1}{4}) + \frac{3}{10}N(1,\frac{1}{4})$ and $h(x) = x$,



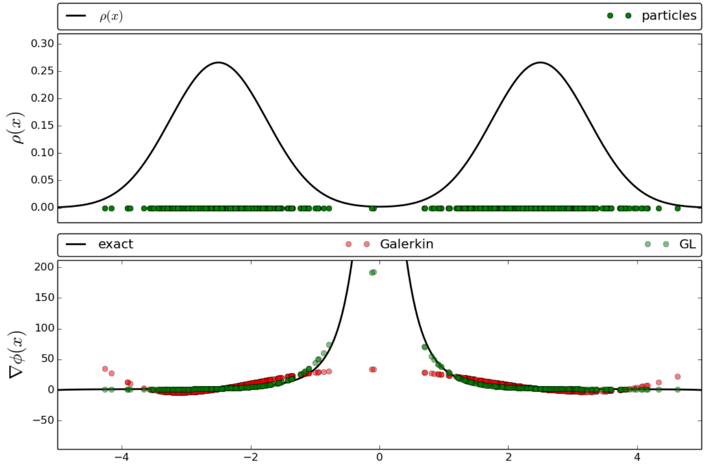
Numerics

Bimodal distribution, Varying Variance

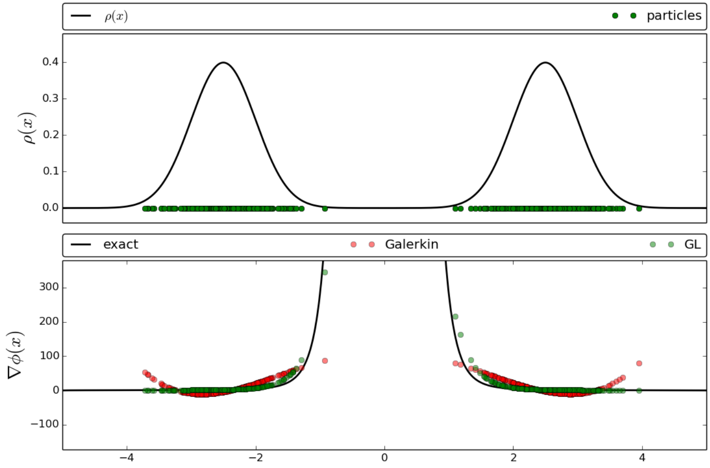


Numerics

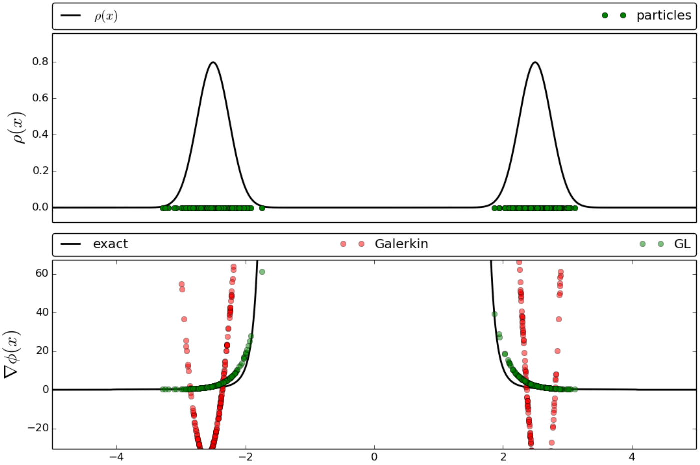
Bimodal distribution, Varying Variance



Numerics
Bimodal distribution, Varying Variance

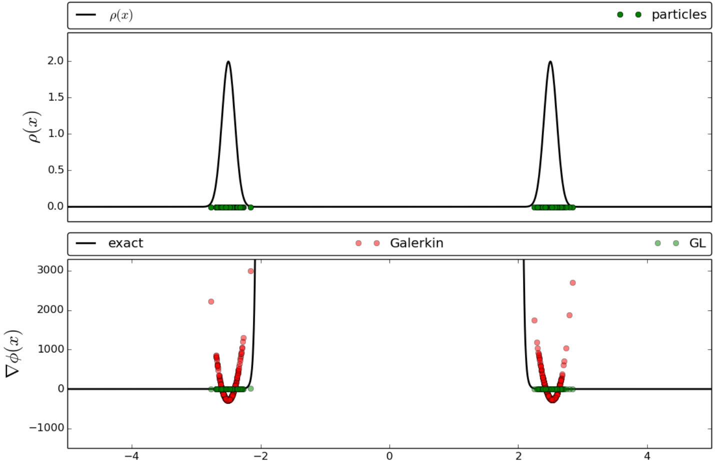


Numerics
Bimodal distribution, Varying Variance

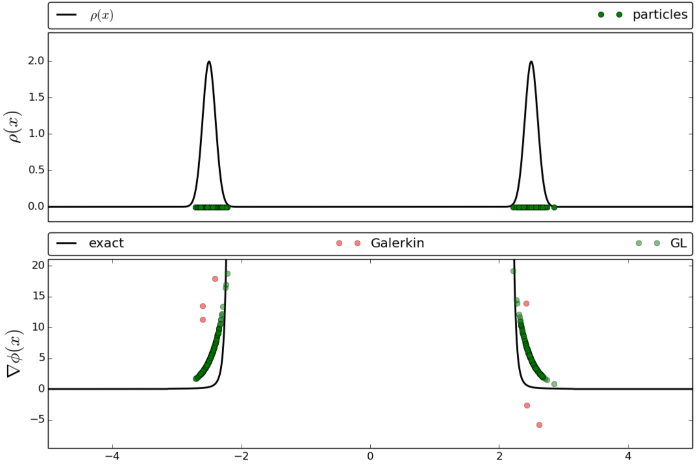


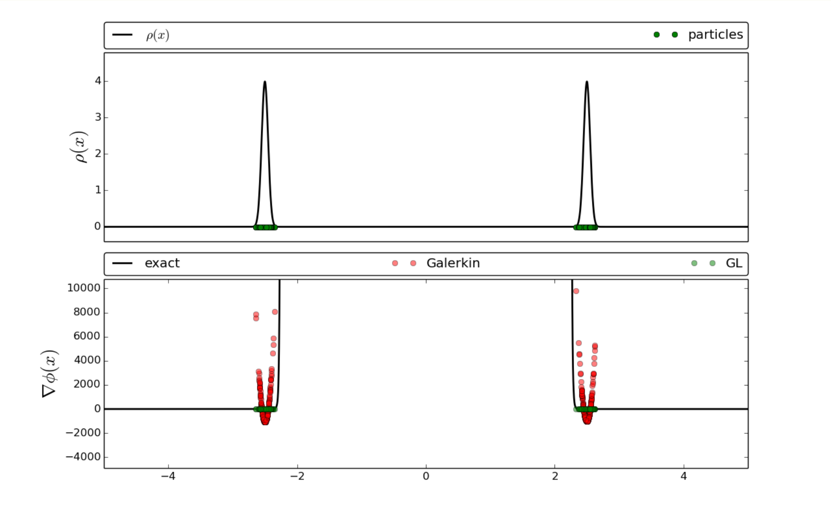
Numerics

Bimodal distribution, Varying Variance



Bimodal distribution, Varying Variance

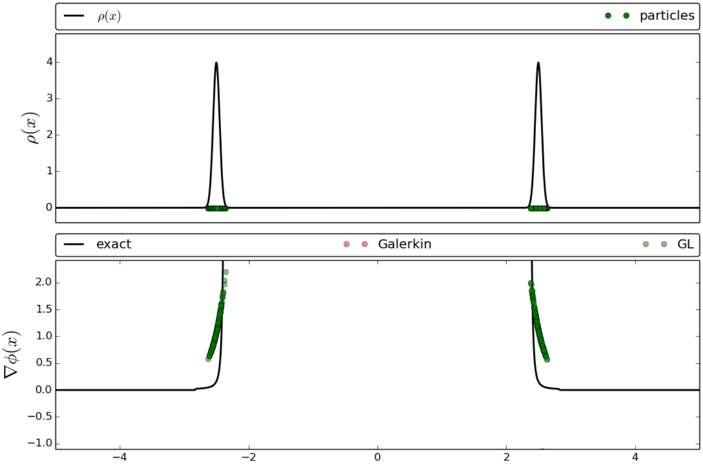




$\rho(x)$
particles
4
3
2
1
0
$\rho(x)$
exact
Galerkin
GL
10000
8000
6000
4000
2000
0
−2000
−4000
$\nabla\phi(x)$
−4
−2
0
2
4

Numerics

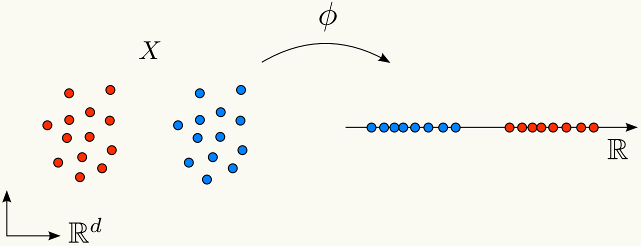
Bimodal distribution, Varying Variance



Outline

1. Poisson equation
2. Poisson equation in nonlinear filtering
3. Numerical solution: Galerkin method
4. Spectral clustering
5. Numerical solution: Graph Laplacian based method
6. Poisson equation in classification

Classification Problem



Feature vector: $X \in \mathbb{R}^d$

Label: $Y \in \{-1, 1\}$

Training data: $\{(X_1, Y_1), \ldots, (X_N, Y_N)\}$

Classifier: $\phi(x) = ?$

Poisson Equation in Classification

Optimization: A classifier is obtained through optimization

$$\phi^{(N)} = \arg\min_{\phi \in \Phi^{(N)}} \underbrace{\sum_{i=1}^{N} \frac{1}{2} |\nabla \phi(X_i)|^2}_{\text{Complexity penalty}} - \underbrace{\sum_{i=1}^{N} \phi(X_i)(Y_i - \hat{Y})}_{\text{Linear loss}}$$

In the limit as $N \to \infty$:

$$\phi = \arg\min_{\phi \in \Phi} \left(\int \frac{1}{2} |\nabla \phi|^2 \rho \, \mathrm{d}x - \int \phi(h - \hat{h}) \rho \, \mathrm{d}x \right)$$

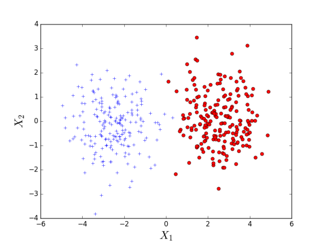
where

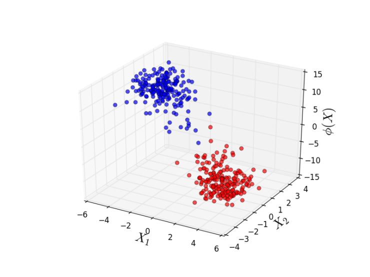
- ρ is the pdf generating data
- $h = \mathsf{E}[Y|X]$ is the underlying labeling function

First order optimality condition:

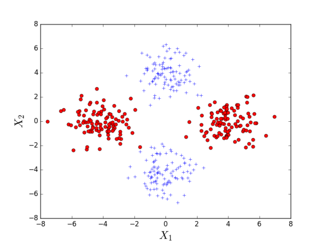
$$-\frac{1}{\rho} \nabla \cdot (\rho \nabla \phi) = h - \hat{h},$$

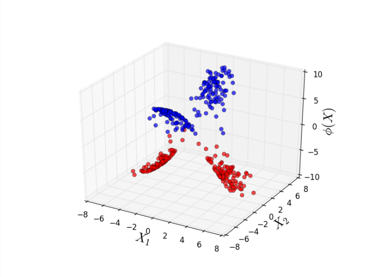
Example I



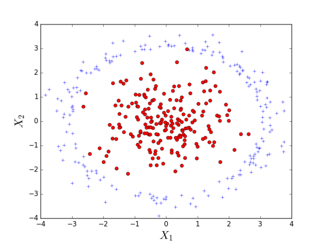


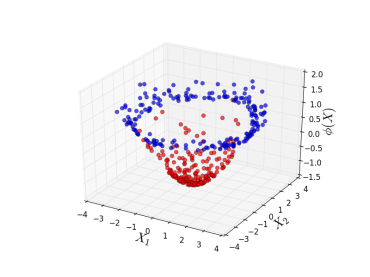
Example II



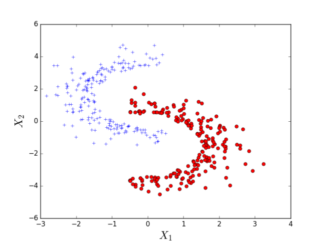


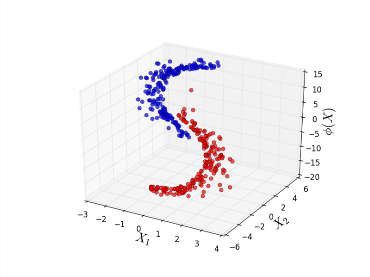
Example III





Example IV





Thank you!

References

- S. Meyn and R. Tweedie. Markov chains and stochastic stability. Springer Science & Business Media, 2012.
- T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn. Multivariable feedback particle filter, *Automatica*, 2015
- R. S. Laugesen, P. G. Mehta, S. P. Meyn, and M. Raginsky. Poisson Equation in Nonlinear Filtering. SICON, 2015
- U. Von Luxburg. "A tutorial on spectral clustering. Statistics and computing", 2007
- M. Belkin. "Problems of learning on manifolds", 2003.
- R. Coifman, and S. Lafon. Diffusion maps.Applied and computational harmonic analysis, 2006
- M. Hein, J. Audibert, and U. Von Luxburg. "Graph Laplacians and their convergence on random neighborhood graphs.", 2006.
- A. Grigoryan, "Heat kernels on weighted manifolds and applications." Cont. Math , 2006.
- Grace Wahba. Spline models for observational data, Siam, 1990.

- Heat kernel:

$$k_t(x,y) = \frac{1}{\sqrt{4\pi t}} \exp(-\frac{|x-y|^2}{4t}$$

$$K_\epsilon(\phi)(x) := \int k_\epsilon(x,y)\phi(y)\,\mathrm{d}y = \phi(x) + \epsilon\Delta\phi(x) + O(\epsilon^2)$$

- Data dependent kernel:

$$W_\epsilon(x,y) = \frac{k_\epsilon(x,y)}{\sqrt{K_\epsilon(\rho)(x)}\sqrt{K_\epsilon(\rho)(y)}}$$

$$A_\epsilon(\phi)(x) := \frac{\int W_\epsilon(x,y)\phi(y)\rho(y)\,\mathrm{d}y}{\int W_\epsilon(x,y)\rho(y)\,\mathrm{d}y} = \phi(x) + \epsilon\frac{1}{\rho}\nabla\cdot(\rho\nabla\phi)(x) + O(\epsilon^2)$$

Proof idea:

$$\frac{\partial k_\epsilon}{\partial \epsilon} = \Delta k_\epsilon, \quad \lim_{\epsilon\to 0} k_\epsilon = \delta$$