Kalman Filter and its Modern Extensions
An Interacting Particle Perspective

National Renewable Energy Laboratory (NREL)
Golden, CO, Apr 11-12, 2019

Prashant G. Mehta[†]

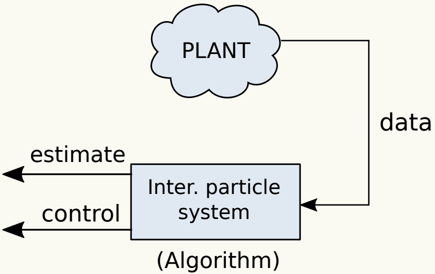
Joint work with Amirhossein Taghvaei[†] and Sean Meyn[+]

[†]Coordinated Science Laboratory
Department of Mechanical Science and Engg., U. Illinois
[+]Department of Electrical and Computer Engg., U. Florida

April 11, 2019

ILLINOIS

The What? Interacting particle system as an algorithm



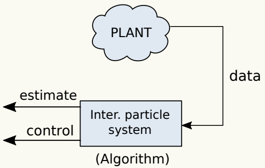
Example of an interacting particle system: Kuramoto oscillators

Example of an algorithm: particle filter

The What and the Why?

Please ask questions!

The What? Interacting particle system as an algorithm



The Why?

- Applicable to general class of models
  1. nonlinear, non-Gaussian
  2. even simulation models

- Possible benefits in high-dimensional settings

- An over-looked topic (may be?) in Control Theory (but important in related fields)

Outline

I will focus on algorithms for the estimation problem

1. Kalman filter
2. Ensemble Kalman filter $\Longleftarrow$ An interacting particle system
3. Feedback particle filter
4. Learning and optimal control $\Longleftarrow$ Only a movie!

Key takeaway

Please ask questions!

Estimation algorithm is a feedback control law:

$$[\text{control}] = [\text{gain}] \cdot [\text{error}] \qquad \text{(proportional control)}$$

Estimation algorithm is a feedback control law:

$$[\text{control}] = [\text{gain}] \cdot [\text{error}] \qquad \text{(proportional control)}$$

The question: What is the gain?

Key takeaway
Please ask questions!

Estimation algorithm is a feedback control law:

$$[\text{control}] = [\text{gain}] \cdot [\text{error}] \qquad \text{(proportional control)}$$

The question: What is the gain?

Answer: Solution to an optimization problem.

Seems appropriate given all the Optimization talent in the room!

Signal (hidden): X $\qquad X \sim \mathsf{P}(X)$ (prior)

Signal (hidden): X $\quad X \sim \mathsf{P}(X)$ (prior)

Observation (known): Y $\quad Y \sim \mathsf{P}(Y|X)$ (sensor model)

Signal (hidden): X $\quad X \sim \mathsf{P}(X)$ (prior)

Observation (known): Y $\quad Y \sim \mathsf{P}(Y|X)$ (sensor model)

Problem: What is X ?

Signal (hidden): X $\quad X \sim \mathsf{P}(X)$ (prior)

Observation (known): Y $\quad Y \sim \mathsf{P}(Y|X)$ (sensor model)

Problem: What is X ?

Solution

Bayes' rule: $$\underbrace{\mathsf{P}(X|Y)}_{\text{Posterior}} \propto \mathsf{P}(Y|X)\underbrace{\mathsf{P}(X)}_{\text{Prior}}$$

Signal (hidden): $X \qquad X \sim \mathsf{P}(X)$ (prior)

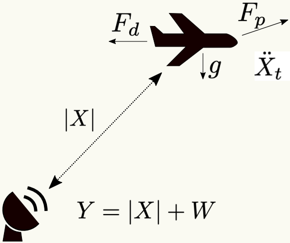
Observation (known): $Y \qquad Y \sim \mathsf{P}(Y|X)$ (sensor model)

Problem: What is X ?

Solution

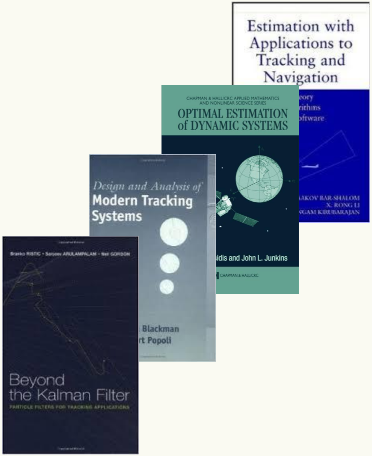
Bayes' rule: $$\underbrace{\mathsf{P}(X|Y)}_{\text{Posterior}} \propto \mathsf{P}(Y|X)\underbrace{\mathsf{P}(X)}_{\text{Prior}}$$

Key takeaway: Bayes' rule $\equiv$ proportional ([gain] $\cdot$ [error]) control!



$$\ddot{X}_t = F_P + F_d + g$$

$$Y = |X| + W$$



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

A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. Springer, 2010.

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

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

A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. Springer, 2010.

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

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

or if you prefer $Y_t := \frac{\mathrm{d}}{\mathrm{d}t} Z_t = h(X_t)$ + white noise

A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. Springer, 2010.

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

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

A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. Springer, 2010.

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

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

Problem: What is X_t? given obs. till time $t =: \mathscr{Z}_t$

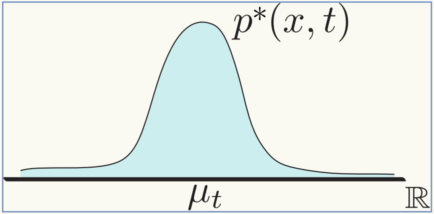
A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. Springer, 2010.

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

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

Problem: What is X_t? given obs. till time $t =: \mathscr{Z}_t$

Answer in terms of posterior: $\mathsf{P}(X_t|\mathscr{Z}_t) =: p(x,t)$.



A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. Springer, 2010.

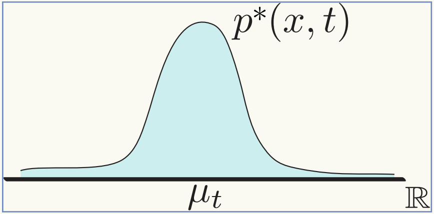
Nonlinear Filtering
Mathematical Problem

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

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

Problem: What is X_t? given obs. till time $t =: \mathscr{Z}_t$

Answer in terms of posterior: $\mathsf{P}(X_t|\mathscr{Z}_t) =: p(x,t)$.



Posterior is an information state

$$\mathsf{P}(X_t \in A|\mathscr{Z}_t) = \int_A p(x,t)\,\mathrm{d}x$$

$$\mathsf{E}(f(X_t)|\mathscr{Z}_t) = \int_{\mathbb{R}} f(x)p(x,t)\,\mathrm{d}x$$

A. Bain and D. Crisan, Fundamentals of Stochastic Filtering. Springer, 2010.

Outline

1. **Kalman filter**
2. Ensemble Kalman filter
3. Feedback particle filter

Filtering problem: Linear Gaussian setting

Model:

Signal process: $\mathrm{d}X_t = AX_t\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t$ (linear dynamics)

Observation process: $\mathrm{d}Z_t = HX_t\,\mathrm{d}t + \mathrm{d}W_t$ (linear observation)

Prior distribution: $X_0 \sim \mathcal{N}(m_0, \Sigma_0)$ (Gaussian prior)

Problem: Find conditional probability distribution, $\mathsf{P}(X_t|\mathscr{Z}_t)$

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

Filtering problem: Linear Gaussian setting

Model:

Signal process: $\mathrm{d}X_t = AX_t\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t$ (linear dynamics)

Observation process: $\mathrm{d}Z_t = HX_t\,\mathrm{d}t + \mathrm{d}W_t$ (linear observation)

Prior distribution: $X_0 \sim \mathcal{N}(m_0, \Sigma_0)$ (Gaussian prior)

Problem: Find conditional probability distribution, $\mathsf{P}(X_t|\mathscr{Z}_t)$

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

Model:

Signal process: $\mathrm{d}X_t = AX_t\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t$ (linear dynamics)

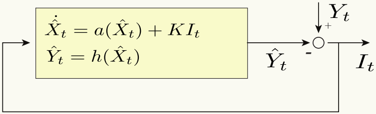
Observation process: $\mathrm{d}Z_t = HX_t\,\mathrm{d}t + \mathrm{d}W_t$ (linear observation)

Prior distribution: $X_0 \sim \mathscr{N}(m_0, \Sigma_0)$ (Gaussian prior)

Problem: Find conditional probability distribution, $\mathsf{P}(X_t|\mathscr{Z}_t)$

Solution: Kalman-Bucy filter – $P(X_t|\mathscr{Z}_t)$ is Gaussian $\mathscr{N}(\hat{X}_t, \Sigma_t)$

Update for mean: $\mathrm{d}\hat{X}_t = A\hat{X}_t\,\mathrm{d}t + \mathsf{K}_t \underbrace{(\mathrm{d}Z_t - H\hat{X}_t\,\mathrm{d}t)}_{\text{error}}$



Kalman Filter

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

Filtering problem: Linear Gaussian setting

Model:

Signal process: $\mathrm{d}X_t = AX_t\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t$ (linear dynamics)

Observation process: $\mathrm{d}Z_t = HX_t\,\mathrm{d}t + \mathrm{d}W_t$ (linear observation)

Prior distribution: $X_0 \sim \mathcal{N}(m_0, \Sigma_0)$ (Gaussian prior)

Problem: Find conditional probability distribution, $\mathsf{P}(X_t|\mathscr{Z}_t)$

Solution: Kalman-Bucy filter – $P(X_t|\mathscr{Z}_t)$ is Gaussian $\mathcal{N}(\hat{X}_t, \Sigma_t)$

Update for mean: $$\mathrm{d}\hat{X}_t = A\hat{X}_t\,\mathrm{d}t + \mathsf{K}_t\underbrace{(\mathrm{d}Z_t - H\hat{X}_t\,\mathrm{d}t)}_{\text{error}}$$

The question: What is the gain?

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

Filering problem: Linear Gaussian setting

Model:

Signal process: $\mathrm{d}X_t = AX_t\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t$ (linear dynamics)

Observation process: $\mathrm{d}Z_t = HX_t\,\mathrm{d}t + \mathrm{d}W_t$ (linear observation)

Prior distribution: $X_0 \sim \mathcal{N}(m_0, \Sigma_0)$ (Gaussian prior)

Problem: Find conditional probability distribution, $\mathsf{P}(X_t|\mathscr{Z}_t)$

Solution: Kalman-Bucy filter – $P(X_t|\mathscr{Z}_t)$ is Gaussian $\mathcal{N}(\hat{X}_t, \Sigma_t)$

Update for mean: $\mathrm{d}\hat{X}_t = A\hat{X}_t\,\mathrm{d}t + \mathsf{K}_t \underbrace{(\mathrm{d}Z_t - H\hat{X}_t\,\mathrm{d}t)}_{\text{error}}$

Update for covariance: $\frac{\mathrm{d}\Sigma_t}{\mathrm{d}t} = \mathrm{Ric}(\Sigma_t)$ (Riccati 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).

Problems and research directions

Classical settings: additional issues due to

1. uncertainties in the signal model
   - interacting multiple model (Kalman) filter [Blom and Bar-Shalom. IEEE TAC (1988).]
2. uncertainties in the measurement model
   - data association (Kalman) filter [Bar-Shalom. Automatica (1975).]
   - adaptive (Kalman) filter
3. communication constraints
   - distributed Kalman filters with consensus like terms [Olfati-Saber; others]

Analysis: Filter stability [Ocone and Pardoux SICON (1996).]

1. Requires controllability of (A, σ_B) and observability of (A, H).

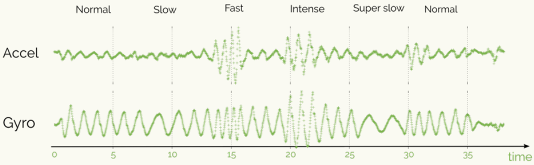
Problems and research directions

Classical settings: additional issues due to

1. uncertainties in the signal model
   - interacting multiple model (Kalman) filter [Blom and Bar-Shalom. IEEE TAC (1988).]
2. uncertainties in the measurement model
   - data association (Kalman) filter [Bar-Shalom. Automatica (1975).]
   - adaptive (Kalman) filter
3. communication constraints
   - distributed Kalman filters with consensus like terms [Olfati-Saber; others]

Modern settings: machine learning problems involving time-series data

- no good signal models!



Outline

1. Kalman filter
2. **Ensemble Kalman filter**
3. Feedback particle filter

Kalman-Bucy filter: $P(X_t|\mathscr{Z}_t)$ is Gaussian $\mathcal{N}(\hat{X}_t, \Sigma_t)$

Update for mean: $\mathrm{d}\hat{X}_t = A\hat{X}_t \,\mathrm{d}t + \mathsf{K}_t(\,\mathrm{d}Z_t - H\hat{X}_t \,\mathrm{d}t)$

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

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

Computation:

if state dimension is d $\Rightarrow$ covariance matrix is $d \times d$

$\Rightarrow$ computational complexity is $O(d^2)$

$\Rightarrow$ This becomes a problem in high-dimensional settings (e.g weather prediction)

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

Kalman-Bucy filter: $P(X_t|\mathscr{Z}_t)$ is Gaussian $\mathscr{N}(\hat{X}_t, \Sigma_t)$

Update for mean: $\mathrm{d}\hat{X}_t = A\hat{X}_t\,\mathrm{d}t + \mathsf{K}_t(\mathrm{d}Z_t - H\hat{X}_t\,\mathrm{d}t)$

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

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

Computation:

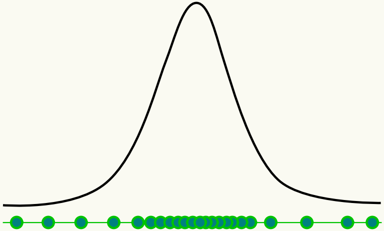
if state dimension is d $\Rightarrow$ covariance matrix is $d \times d$

$\Rightarrow$ computational complexity is $O(d^2)$

$\Rightarrow$ This becomes a problem in high-dimensional settings (e.g weather prediction)

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

Idea: approximate the posterior $P(X_t|\mathscr{Z}_t)$ using particles $\{X_t^i\}_{i=1}^N$



$$\mathsf{P}(X_t \in A|\mathscr{Z}_t) = \int_A p(x,t)\,\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}^N 1_{X_t^i \in A}$$

$$\mathsf{E}(f(X_t)|\mathscr{Z}_t) = \int_{\mathbb{R}} f(x)p(x,t)\,\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}^N f(X_t^i)$$

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).

Idea: approximate the posterior $P(X_t|\mathscr{Z}_t)$ using particles $\{X_t^i\}_{i=1}^N$

$$\mathrm{d}X_t^i = \underbrace{AX_t^i\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t^i}_{\text{(simulation)}} + \underbrace{\mathsf{K}_t^{(N)}\Big(\mathrm{d}Z_t - \frac{HX_t^i\,\mathrm{d}t + N^{-1}\sum_j HX_t^j}{2}\Big)}_{\text{data assimilation step}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

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).

Idea: approximate the posterior $P(X_t|\mathscr{Z}_t)$ using particles $\{X_t^i\}_{i=1}^N$

$$\mathrm{d}X_t^i = \underbrace{AX_t^i\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t^i}_{\text{(simulation)}} + \underbrace{\mathsf{K}_t^{(N)}\Big(\mathrm{d}Z_t - \frac{HX_t^i\,\mathrm{d}t + N^{-1}\sum_j HX_t^j}{2}\Big)}_{\text{data assimilation step}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

Computations: computational complexity is $O(Nd)$ – efficient when $d >> N$

Consistency: [under additional assumptions] $m_t^{(N)} := \frac{1}{N}\sum_{i=1}^N X_t^i \overset{(N\to\infty)}{\longrightarrow} \mathsf{E}(X_t|\mathscr{Z}_t)$

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).

Idea: approximate the posterior $P(X_t|\mathscr{Z}_t)$ using particles $\{X_t^i\}_{i=1}^N$

$$\mathrm{d}X_t^i = \underbrace{AX_t^i\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t^i}_{\text{(simulation)}} + \underbrace{\mathsf{K}_t^{(N)}\Big(\mathrm{d}Z_t - \frac{HX_t^i\,\mathrm{d}t + N^{-1}\sum_j HX_t^j}{2}\Big)}_{\text{data assimilation step}}, \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

Computations: computational complexity is $O(Nd)$ – efficient when $d >> N$

Consistency: [under additional assumptions] $m_t^{(N)} := \frac{1}{N}\sum_{i=1}^N X_t^i \overset{(N\to\infty)}{\longrightarrow} \mathsf{E}(X_t|\mathscr{Z}_t)$

Computing the gain:

empirical Kalman gain: $\mathsf{K}_t^{(N)} := \Sigma_t^{(N)} H^\top$

empirical covariance: $\Sigma_t^{(N)} := \frac{1}{N-1}\sum_{i=1}^N (X_t^i - m_t^{(N)})(X_t^i - m_t^{(N)})^\top$

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).

EnKf formulation:

- EnKF based on perturbed observation (Evensen, 1994)
- The square root EnKF (Whitaker et. al. 2002)
- Continuous-time formulation (Bergemann and Reich. 2012)
- EnKF as special case of FPF (Yang et. al. 2013)
- Optimal transport formulation (Taghvaei and **M.**, 2016)

Error analysis (requires additional assumptions):

- m.s.e converges as $O(\frac{1}{N})$ for any finite time (Le Gland et. al. 2009, Mandel et. al. 2011, Kelly et. al. 2014)
- m.s.e converges as $O(\frac{1}{N})$ uniform in time (Del Moral, et. al. 2016, de Wiljes et. al. 2016, Bishop and Del Moral, 2017)

Outline

1. Kalman filter

$$[\text{control}] = K_t(\mathrm{d}Z_t - H\hat{X}_t)$$

2. Ensemble Kalman filter

$$[\text{control}] = K_t^{(N)}(\mathrm{d}Z_t - \frac{HX_t^i + N^{-1}\sum_{j=1}^N HX_t^j}{2}\,\mathrm{d}t)$$

3. **Feedback particle filter** (for nonlinear non Gaussian problems)

$$[\text{control}] = ??$$

Feedback Particle Filter
A numerical algorithm for nonlinear filtering

Problem:

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

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

Posterior distribution $P(X_t|\mathscr{Z}_t)$?

Yang, **M.**, Meyn. Feedback particle filter. *IEEE Trans. Aut. Control* (2013)
Yang, Laugesen, **M.**, Meyn. Multivariable Feedback particle filter. *Automatica* (2016)
Zhang, Taghvaei, **M**. Feedback particle filter on Riemannian manifolds and Lie groups. *IEEE Trans. Aut. Control* (2018)

Problem:

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

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

Posterior distribution $P(X_t|\mathscr{Z}_t)$?

Solution: feedback particle filter

$$\mathrm{d}X^i_t = \underbrace{a(X^i_t)\,\mathrm{d}t + \sigma(X^i_t)\,\mathrm{d}B^i_t}_{\text{simulation}} + \mathsf{K}_t(X^i_t) \circ \underbrace{\left(\mathrm{d}Z_t - \frac{h(X^i_t) + \mathsf{E}(h(X^i_t)|\mathscr{Z}_t)}{2}\,\mathrm{d}t\right)}_{\text{error}} \quad X^i_0 \sim p_0$$

Yang, **M.**, Meyn. Feedback particle filter. *IEEE Trans. Aut. Control* (2013)
Yang, Laugesen, **M.**, Meyn. Multivariable Feedback particle filter. *Automatica* (2016)
Zhang, Taghvaei, **M.** Feedback particle filter on Riemannian manifolds and Lie groups. *IEEE Trans. Aut. Control* (2018)

Problem:

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

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

Posterior distribution $P(X_t|\mathscr{Z}_t)$?

Solution: feedback particle filter

$$\mathrm{d}X_t^i = \underbrace{a(X_t^i)\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}B_t^i}_{\text{simulation}} + \mathsf{K}_t(X_t^i) \circ \underbrace{\left(\mathrm{d}Z_t - \frac{h(X_t^i) + \mathsf{E}(h(X_t^i)|\mathscr{Z}_t)}{2}\,\mathrm{d}t\right)}_{\text{error}} \qquad X_0^i \sim p_0$$

approximation:

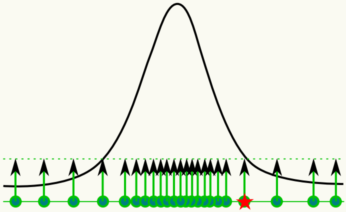
$$\mathsf{E}(f(X_t)|\mathscr{Z}_t) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f(X_t^i)$$

Yang, **M.**, Meyn. Feedback particle filter. *IEEE Trans. Aut. Control* (2013)
Yang, Laugesen, **M.**, Meyn. Multivariable Feedback particle filter. *Automatica* (2016)
Zhang, Taghvaei, **M**. Feedback particle filter on Riemannian manifolds and Lie groups. *IEEE Trans. Aut. Control* (2018)

Idea: approximate the posterior $P(X_t|\mathscr{Z}_t)$ using particles $\{X_t^i\}_{i=1}^N$

$$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}B_t^i, \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$
$$\mathrm{d}M_t^i = M_t^i h(X_t^i)\,\mathrm{d}Z_t, \quad M_0^i = 1$$

where M_t^i are referred to as the importance weights.



N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation (1993).
A. Doucet and A. Johansen, A Tutorial on Particle Filtering and Smoothing: Fifteen years later (2008).
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).

Particle filter

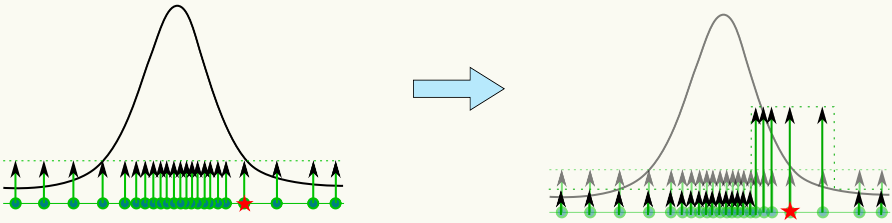
Conventional approach

Idea: approximate the posterior $P(X_t|\mathscr{Z}_t)$ using particles $\{X_t^i\}_{i=1}^N$

$$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}B_t^i, \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$

$$\mathrm{d}M_t^i = M_t^i h(X_t^i)\,\mathrm{d}Z_t, \quad M_0^i = 1$$

where M_t^i are referred to as the importance weights.



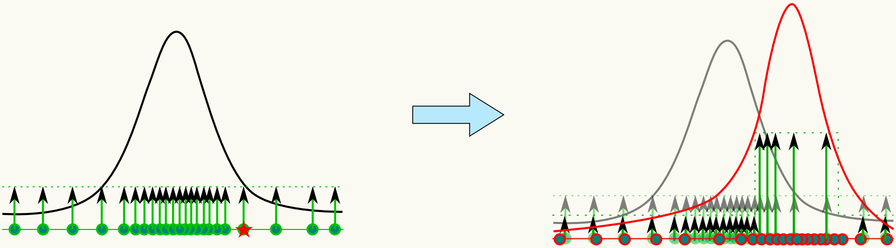
N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation (1993).
A. Doucet and A. Johansen, A Tutorial on Particle Filtering and Smoothing: Fifteen years later (2008).
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).

Particle filter
Conventional approach

Idea: approximate the posterior $P(X_t | \mathscr{Z}_t)$ using particles $\{X_t^i\}_{i=1}^N$

$$dX_t^i = a(X_t^i)\,dt + \sigma(X_t^i)\,dB_t^i, \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$
$$dM_t^i = M_t^i h(X_t^i)\,dZ_t, \quad M_0^i = 1$$

where M_t^i are referred to as the importance weights.



N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation (1993).
A. Doucet and A. Johansen, A Tutorial on Particle Filtering and Smoothing: Fifteen years later (2008).
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).

Idea: approximate the posterior $P(X_t|\mathscr{Z}_t)$ using particles $\{X_t^i\}_{i=1}^N$

$$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}B_t^i, \quad X_0^i \overset{\text{i.i.d}}{\sim} p_0$$
$$\mathrm{d}M_t^i = M_t^i h(X_t^i)\,\mathrm{d}Z_t, \quad M_0^i = 1$$

where M_t^i are referred to as the importance weights.

approximation:

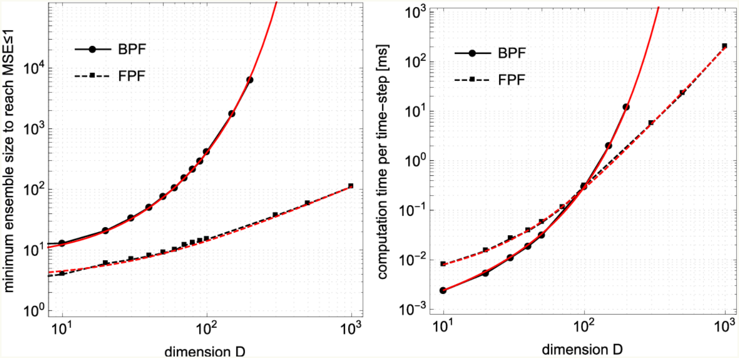
$$\mathsf{E}(f(X_t)|\mathscr{Z}_t) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N M_t^i f(X_t^i)$$

Problems:

1. High simulation variance in importance weights. This necessitates resampling.
2. Particle impoverishment for high-dimensional problems – $N \propto \exp(d)$
3. No explicit error correction structure! Where is the ensemble Kalman filter?

N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation (1993).
A. Doucet and A. Johansen, A Tutorial on Particle Filtering and Smoothing: Fifteen years later (2008).
P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions (2008).

How do these compare?
FPF vs. BPF



Reproduced from: Surace, Kutschireiter, Pfister. How to avoid the curse of dimensionality: scalability of particle filters with and without importance weights? *SIAM Review* (2019).

Additional comparisons appear in: A. K. Tilton, S. Ghiotto, and P. G. Mehta. A comparative study of nonlinear filtering techniques. In Proc. 16th Int. Conf. on Inf. Fusion, pages 1827-1834, Istanbul, Turkey, July 2013.
P. M. Stano, A. K. Tilton, and R. Babuska. Estimation of the soil-dependent time-varying parameters of the hopper sedimentation model: The FPF versus the BPF. Control Engineering Practice, 24:67-78 (2014).
K Berntorp. Feedback particle filter: Application and evaluation. In 18th Int. Conf. Infor- mation Fusion, Washington, DC, 2015.

Gain is a solution of an optimization problem:

$$\min_{\phi \in H_0^1} \int \Big(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \Big) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

$$\mathsf{K} = \nabla \phi$$

First order optimality condition (E-L equation) is the Poisson equation:

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

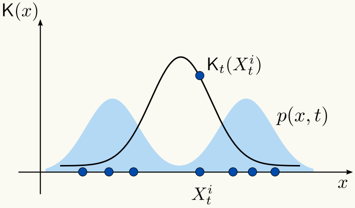
Linear Gaussian case: Solution is the Kalman gain!

Laugesen, **M.**, Meyn and Raginsky. Poisson equation in nonlinear filtering. *SIAM J. Control and Optimization* (2015).
Yang, Laugesen, **M.**, Meyn. Multivariable Feedback particle filter. *Automatica* (2016).

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

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

(1) FPF: $$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \sigma_B(X_t^i)\,\mathrm{d}B_t^i + \underbrace{\mathsf{K}_t(X_t^i)\circ\left(\mathrm{d}Z_t - \frac{h(X_t^i)+\hat{h}_t}{2}\,\mathrm{d}t\right)}_{\text{FPF}}$$

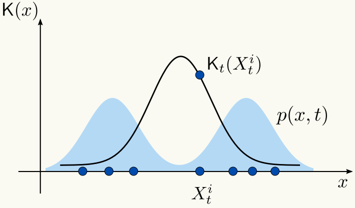


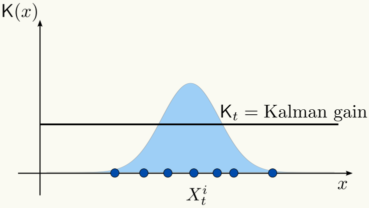
(1) Non-Gaussian density, (2) Gaussian density

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

(1) FPF: $$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \sigma_B(X_t^i)\,\mathrm{d}B_t^i + \underbrace{\mathsf{K}_t(X_t^i)\circ\left(\mathrm{d}Z_t - \frac{h(X_t^i)+\hat{h}_t}{2}\,\mathrm{d}t\right)}_{\text{FPF}}$$

(2) Linear Gaussian: $$\mathrm{d}X_t^i = AX_t^i\,\mathrm{d}t + \sigma_B\,\mathrm{d}B_t^i + \underbrace{\mathsf{K}_t\left(\mathrm{d}Z_t - \frac{HX_t^i+H\hat{X}_t}{2}\,\mathrm{d}t\right)}_{\text{EnKF}}$$





The linear Gaussian FPF is the square-root form of the EnKF. This square-root form of the EnKF was independently obtained by K. Bergemann and S. Reich. An ensemble Kalman-Bucy filter for continuous data assimilation (2012).

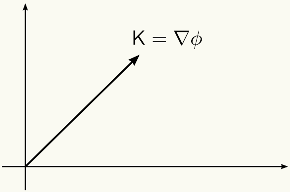
Non-Gaussian case

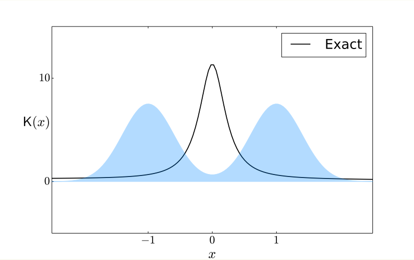
Lets get to approximation!

Gain is a solution of an optimization problem:

$$\min_{\phi \in H_0^1} \int \left(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \right) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

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

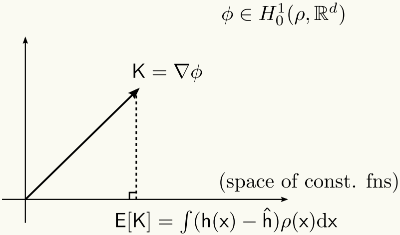


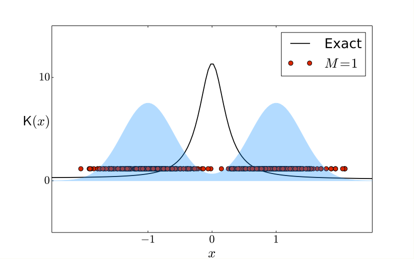


Existence uniqueness theory in: Yang, Laugesen, **M.**, Meyn. Multivariable Feedback particle filter. *Automatica* (2016)

Gain is a solution of an optimization problem:

$$\min_{\phi \in H_0^1} \int \left(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \right) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$



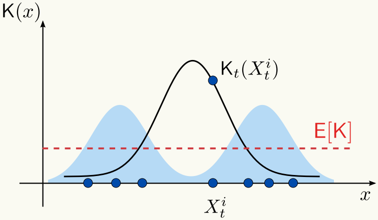


A closed-form formula:

$$\text{(best const. approximation)} = \int (h(x) - \hat{h}) x \rho(x) \, \mathrm{d}x \approx \frac{1}{N} \sum_{i=1}^{N} (h(X_t^i) - \hat{h}^{(N)}) X_t^i$$

FPF = EnKF in two limits:

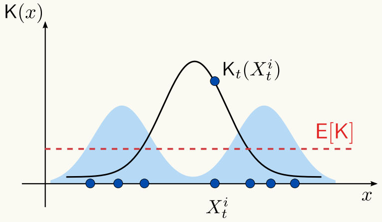
1. Linear Gaussian where gain function = Kalman gain
2. Approximation of the gain function by its average (constant) value



Taghvaei, de Wiljes, **M.**, and Reich, Kalman Filter and its Modern Extensions for the Continuous-time Nonlinear Filtering Problem, *ASME J. of Dynamic Systems, Measurement, and Control* (2018).

FPF = EnKF in two limits:

1. Linear Gaussian where gain function = Kalman gain
2. Approximation of the gain function by its average (constant) value



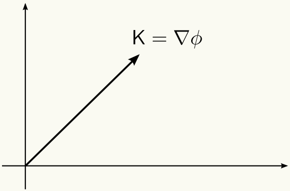
Question: Can we improve this approximation?

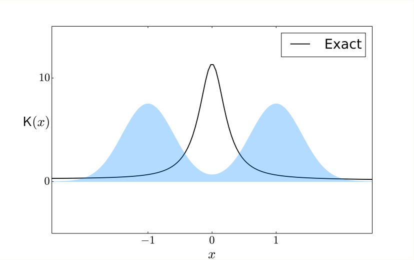
Taghvaei, de Wiljes, **M.**, and Reich, Kalman Filter and its Modern Extensions for the Continuous-time Nonlinear Filtering Problem, *ASME J. of Dynamic Systems, Measurement, and Control* (2018).

Gain is a solution of an optimization problem:

$$\min_{\phi \in S} \int \Big(|\nabla\phi|^2(x) + (h(x) - \hat{h})\phi(x) \Big) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

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

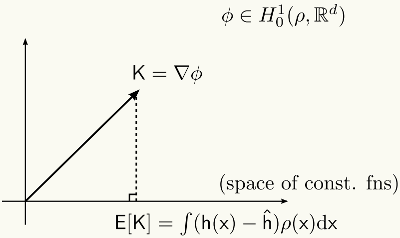


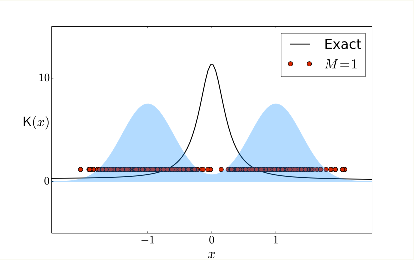


Non-Gaussian case
Galerkin approximation

Gain is a solution of an optimization problem:

$$\min_{\phi \in S} \quad \int \Big(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \Big) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$



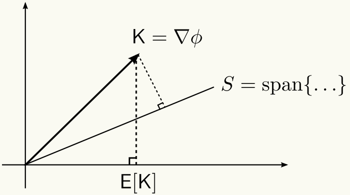


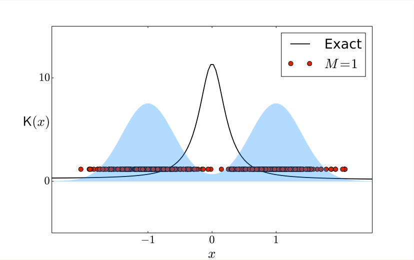
Non-Gaussian case
Galerkin approximation

Gain is a solution of an optimization problem:

$$\min_{\phi \in S} \int \left(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \right) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

$\phi \in H^1_0(\rho, \mathbb{R}^d)$



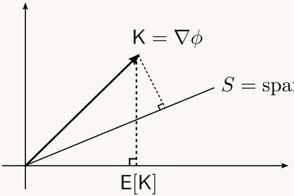


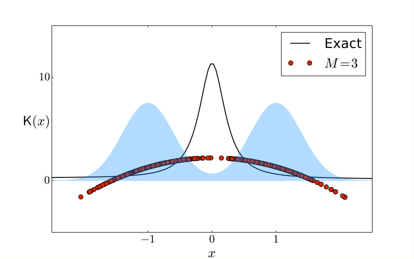
Non-Gaussian case
Galerkin approximation

Gain is a solution of an optimization problem:

$$\min_{\phi \in S} \int \left(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \right) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

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



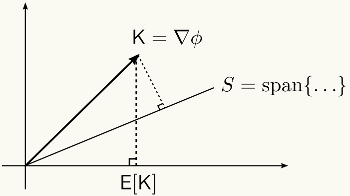


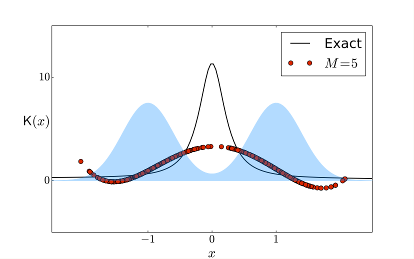
$\psi \in \{1, x, \ldots, x^M\}$

Gain is a solution of an optimization problem:

$$\min_{\phi \in S} \int \Big(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \Big) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

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





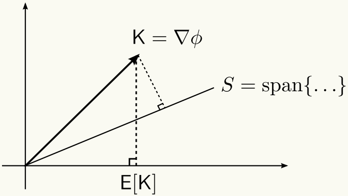
$\psi \in \{1, x, \ldots, x^M\}$

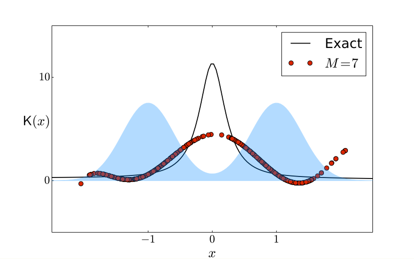
Non-Gaussian case
Galerkin approximation

Gain is a solution of an optimization problem:

$$\min_{\phi \in S} \int \left(|\nabla\phi|^2(x) + (h(x) - \hat{h})\phi(x) \right) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

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



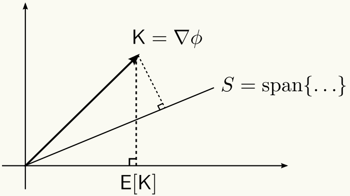


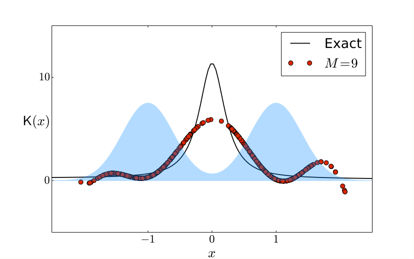
$\psi \in \{1, x, \ldots, x^M\}$

Gain is a solution of an optimization problem:

$$\min_{\phi \in S} \int \Big(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \Big) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

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



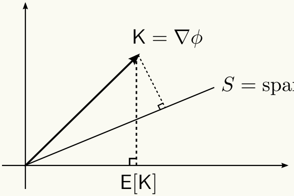


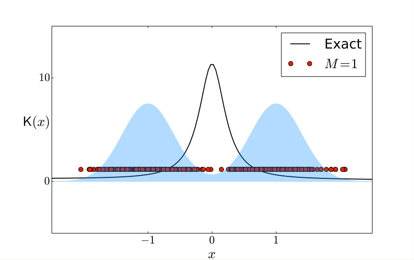
$\psi \in \{1, x, \ldots, x^M\}$

Gain is a solution of an optimization problem:

$$\min_{\phi \in S} \int \Big(|\nabla \phi|^2(x) + (h(x) - \hat{h})\phi(x) \Big) \underbrace{\rho(x)}_{\text{post.}} \, \mathrm{d}x$$

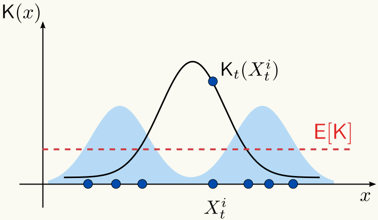
$\phi \in H_0^1(\rho, \mathbb{R}^d)$





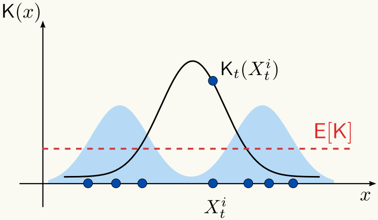
Moral of the story: basis function selection is non-trivial!

What are we looking for?
Ensemble Kalman filter +



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

What are we looking for?
Ensemble Kalman filter +



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

Question: Can we improve this approximation?

Outline

1. Kalman filter

$$K_t = \Sigma_t H$$

2. Ensemble Kalman filter

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

3. **Feedback particle filter**

$$K_t^i = ??$$

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

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

Gain function Approximation

Key idea is to use diffusion maps

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

(2) Semigroup formulation: $$(I-e^{\varepsilon\Delta_\rho})\phi = \varepsilon(h-\hat{h})\,\mathrm{d}s$$

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

Gain function Approximation

Key idea is to use diffusion maps

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

(2) Semigroup formulation: $$(I-e^{\varepsilon\Delta_\rho})\phi = \varepsilon(h-\hat{h})\,\mathrm{d}s$$

(3) Fixed-point problem: $$\phi_\varepsilon = \mathsf{T}_\varepsilon\phi_\varepsilon + \varepsilon(h-\hat{h}_\varepsilon)$$

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

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

(2) Semigroup formulation: $(I - e^{\varepsilon\Delta_\rho})\phi = \varepsilon(h-\hat{h})\,\mathrm{d}s$

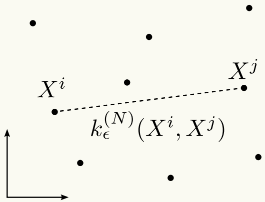
(3) Fixed-point problem: $\phi_\varepsilon = \mathsf{T}_\varepsilon \phi_\varepsilon + \varepsilon(h-\hat{h}_\varepsilon)$

(4) Empirical approximation $\phi_\varepsilon^{(N)} = \mathsf{T}_\varepsilon^{(N)} \phi_\varepsilon^{(N)} + \varepsilon(h-\hat{h}^N)$

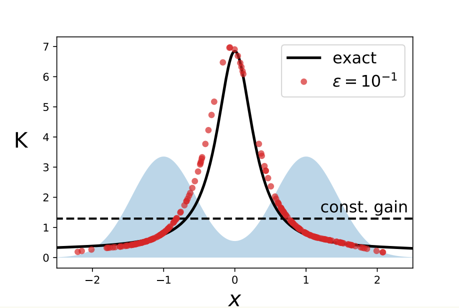
- $\mathsf{T}_\varepsilon^{(N)}$ is a $N \times N$ Markov matrix,

$$\mathsf{T}_\varepsilon^{(N)}{}_{ij} = \frac{k_\varepsilon^{(N)}(X^i, X^j)}{\sum_{l=1}^N k_\varepsilon^{(N)}(X^i, X_l)}$$

- $k_\varepsilon^{(N)}(x,y)$ is the diffusion map kernel



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



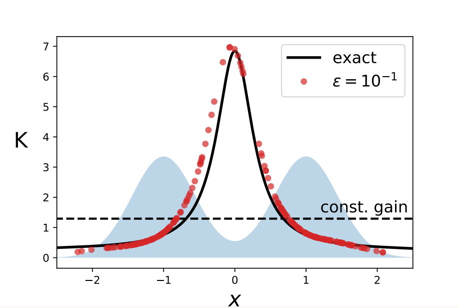
1. No basis function selection!
2. Simple formula

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

3. Reduces to the constant gain in the limit as $\varepsilon \to \infty$

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

Taghvaei and **M.**, Gain Function Approximation for the Feedback Particle Filter, *IEEE Conference on Decision and Control*, (2016).
Taghvaei, **M.**, and Meyn, Error Estimates for the Kernel Gain Function Approximation in the Feedback Particle Filter, *American Control Conference*, (2017).



1. No basis function selection!
2. Simple formula[a]

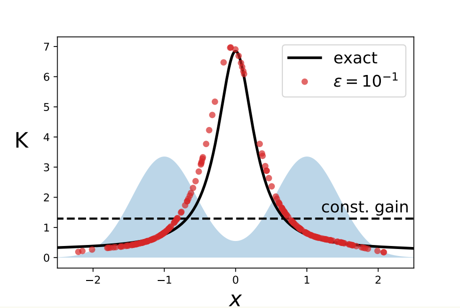
$$\mathsf{K}^i = \sum_{j=1}^{N} s_{ij} X^j$$

3. Reduces to the constant gain in the limit as $\varepsilon \to \infty$

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

[a] Reminiscent of the ensemble transform (Reich, A non-parametric ensemble transform method for Bayesian inference, *SIAM J. Sci. Comput.*, (2013))

Taghvaei and **M.**, Gain Function Approximation for the Feedback Particle Filter, *IEEE Conference on Decision and Control*, (2016).
Taghvaei, **M.**, and Meyn, Error Estimates for the Kernel Gain Function Approximation in the Feedback Particle Filter, *American Control Conference*, (2017).



1. No basis function selection!
2. Simple formula

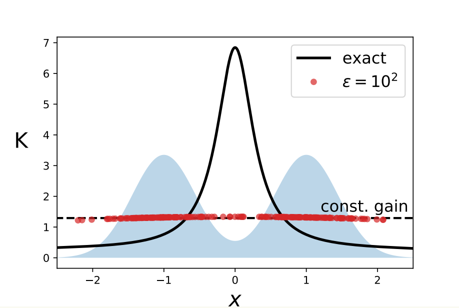
$$\mathsf{K}^i = \sum_{j=1}^{N} s_{ij} X^j$$

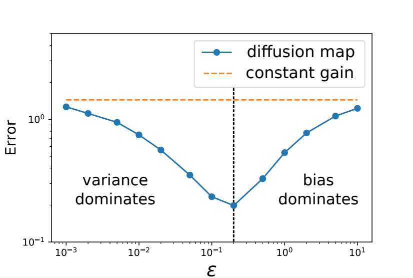
3. Reduces to the constant gain in the limit as $\varepsilon \to \infty$

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

Taghvaei and **M.**, Gain Function Approximation for the Feedback Particle Filter, *IEEE Conference on Decision and Control*, (2016).
Taghvaei, **M.**, and Meyn, Error Estimates for the Kernel Gain Function Approximation in the Feedback Particle Filter, *American Control Conference*, (2017).

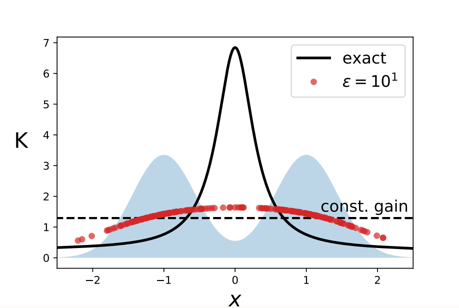
Error estimates: $\text{r.m.s.e} = \underbrace{O(\varepsilon)}_{\text{bias}} + \underbrace{O\left(\frac{1}{\varepsilon^{1+d/2}N^{1/2}}\right)}_{\text{variance}}$

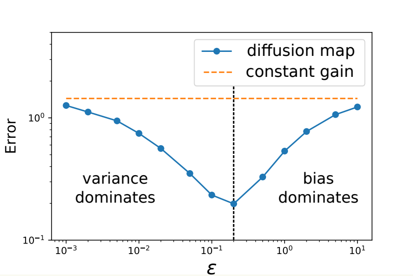




(Bias-variance tradeoff)

Error estimates: $\text{r.m.s.e} = \underbrace{O(\varepsilon)}_{\text{bias}} + \underbrace{O\left(\frac{1}{\varepsilon^{1+d/2}N^{1/2}}\right)}_{\text{variance}}$

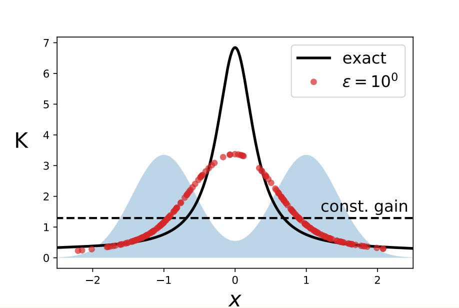


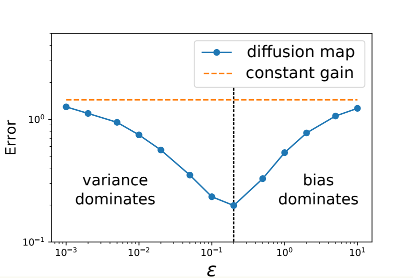


(Bias-variance tradeoff)

Convergence analysis: $\phi_\varepsilon^{(N)} \xrightarrow[\text{variance}]{N\uparrow\infty} \phi_\varepsilon \xrightarrow[\text{bias}]{\varepsilon\downarrow 0} \phi$

Error estimates: $$\text{r.m.s.e} = \underbrace{O(\varepsilon)}_{\text{bias}} + \underbrace{O\Big(\frac{1}{\varepsilon^{1+d/2} N^{1/2}}\Big)}_{\text{variance}}$$

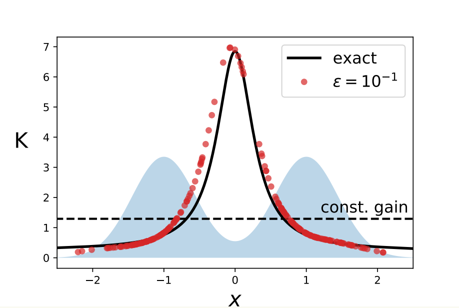


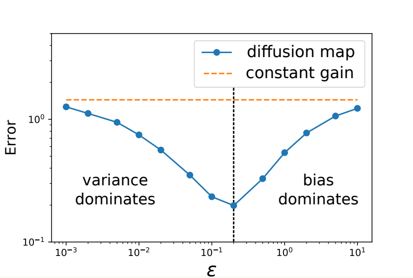


(Bias-variance tradeoff)

Convergence analysis: $\phi_\varepsilon^{(N)} \xrightarrow[\text{variance}]{N\uparrow\infty} \phi_\varepsilon \xrightarrow[\text{bias}]{\varepsilon\downarrow 0} \phi$

Error estimates: $$\text{r.m.s.e} = \underbrace{O(\varepsilon)}_{\text{bias}} + \underbrace{O\left(\frac{1}{\varepsilon^{1+d/2}N^{1/2}}\right)}_{\text{variance}}$$

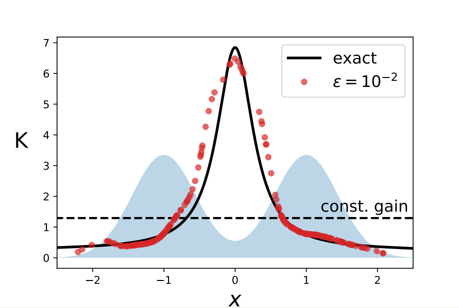


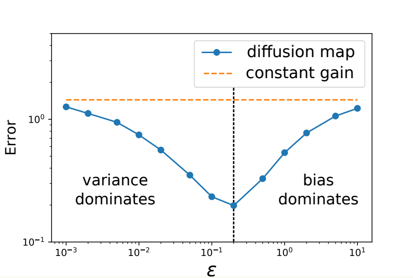


(Bias-variance tradeoff)

Convergence analysis: $\phi_\varepsilon^{(N)} \xrightarrow[\text{variance}]{N\uparrow\infty} \phi_\varepsilon \xrightarrow[\text{bias}]{\varepsilon\downarrow 0} \phi$

Error estimates: $$\text{r.m.s.e} = \underbrace{O(\varepsilon)}_{\text{bias}} + \underbrace{O\left(\frac{1}{\varepsilon^{1+d/2}N^{1/2}}\right)}_{\text{variance}}$$

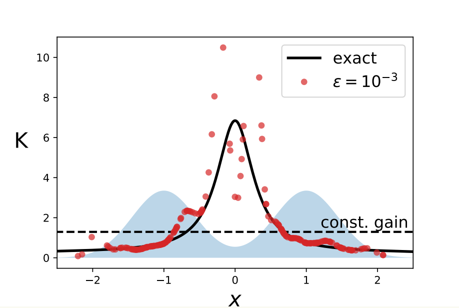


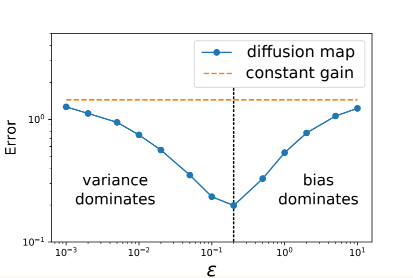


(Bias-variance tradeoff)

Convergence analysis: $\phi_\varepsilon^{(N)} \xrightarrow[\text{variance}]{N\uparrow\infty} \phi_\varepsilon \xrightarrow[\text{bias}]{\varepsilon\downarrow 0} \phi$

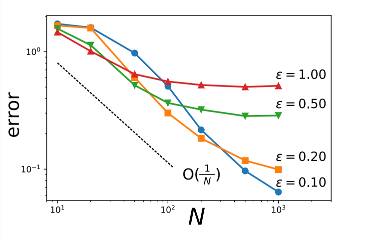
Error estimates: $$\text{r.m.s.e} = \underbrace{O(\varepsilon)}_{\text{bias}} + \underbrace{O\left(\frac{1}{\varepsilon^{1+d/2}N^{1/2}}\right)}_{\text{variance}}$$

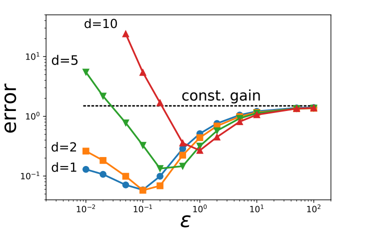


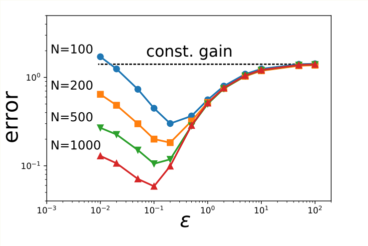


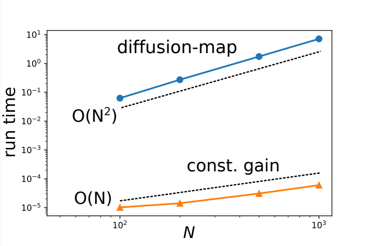
(Bias-variance tradeoff)

Error analysis
Numerical experiments

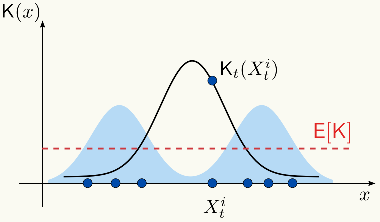








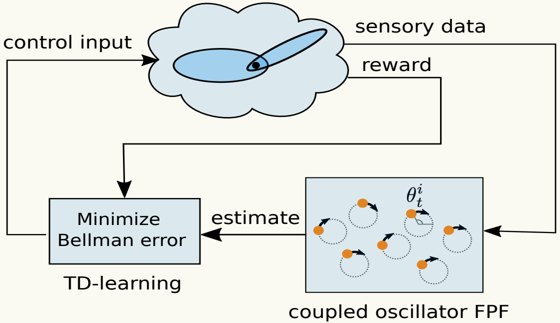
$$\mathrm{d}X_t^i = \underbrace{a(X_t^i)\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}B_t^i}_{\text{simulation}} + \mathsf{K}_t(X_t^i) \circ \underbrace{\left(\mathrm{d}Z_t - \frac{h(X_t^i) + N^{-1}\sum_j h(X_t^j)}{2}\,\mathrm{d}t\right)}_{\text{error}} \qquad X_0^i \sim p_0$$



ENKF: $$\mathsf{K}_t(X_t^i) = \frac{1}{N}\sum_{j=1}^{N}\left(h(X_t^j) - \hat{h}_t^{(N)}\right)X_t^j$$

FPF: $$\mathsf{K}_t(X_t^i) = \sum_{j=1}^{N} s_{ij} X_t^j$$

Interacting particle systems for estimation, learning and optimal control



[Click to play the movie]

T. Wang, A. Taghvaei, P. Mehta, Q-learning for POMDP: An application to learning locomotion gaits, CDC (2019)
A. Taghvaei, S. Hutchinson and P. G. Mehta. A coupled-oscillators-based control architecture for locomotory gaits. CDC (2014).

Backup!