Particle Filter

Charles C. Cossette

McGill University, Department of Mechanical Engineering

March 5, 2021

Pop Quiz!

What do all the following estimation algorithms have in common?

Pop Quiz!

What do all the following estimation algorithms have in common?

- Kalman filter
- Extended Kalman filter (EKF)
- Iterated EKF
- Invariant EKF
- Rauch–Tung–Striebel Smoother
- Sliding Window Filter
- Batch estimator
- Sigma-point Kalman filter (i.e. UKF, CKF, GHKF)
- Iterated Sigma-point Kalman filter
- ESGVI [1]

Pop Quiz!

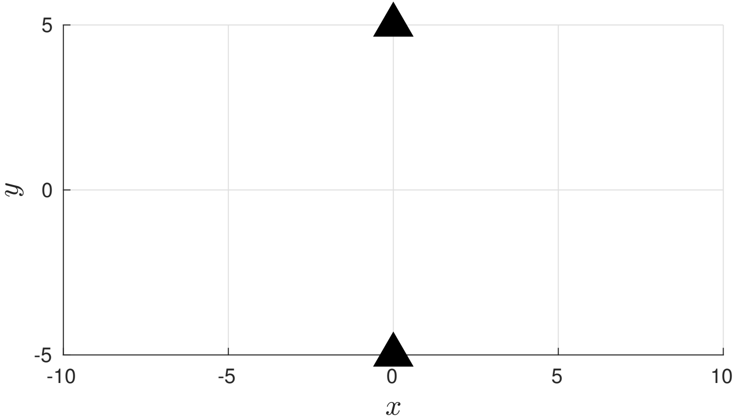
What do all the following estimation algorithms have in common?

- Kalman filter
- Extended Kalman filter (EKF)
- Iterated EKF
- Invariant EKF
- Rauch–Tung–Striebel Smoother
- Sliding Window Filter
- Batch estimator
- Sigma-point Kalman filter (i.e. UKF, CKF, GHKF)
- Iterated Sigma-point Kalman filter
- ESGVI [1]

They all assume the state distribution is Gaussian.

- This makes them Gaussian *assumed density filters*.

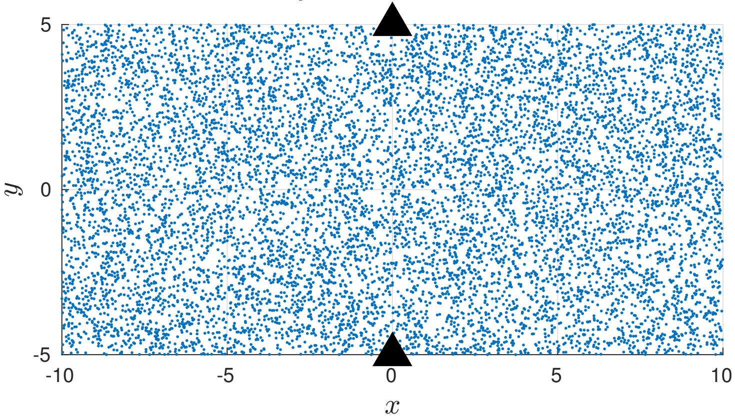
A Non-Gaussian Example



- Suppose we know a robot lies *somewhere* inside the region $\mathbf{x} = \mathbf{r}_a^{zw} \in [[-10 \ \ -5]^\mathsf{T}, [10 \ 5]^\mathsf{T}]$.
- The robot gets distance measurements to two landmarks ℓ_1, ℓ_2 (black triangles) with measurement model

$$y_j = \left\| \mathbf{r}_a^{zw} - \mathbf{r}_a^{\ell_j w} \right\| + v, \qquad v \sim \mathcal{N}(0, R) \tag{1}$$

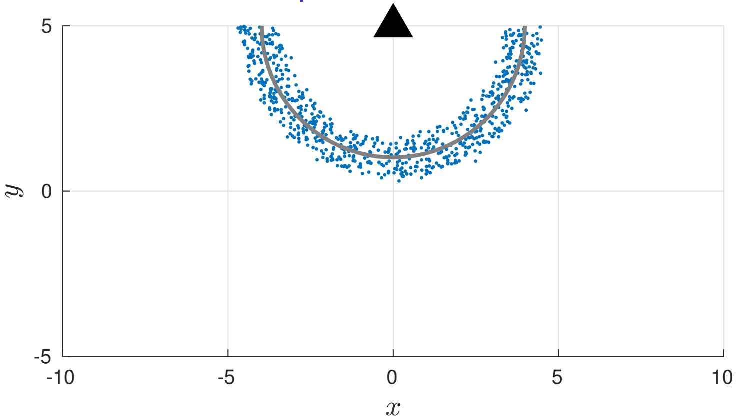
A Non-Gaussian Example



- We are already doing something impossible with Gaussian estimators, we have a *uniform prior*

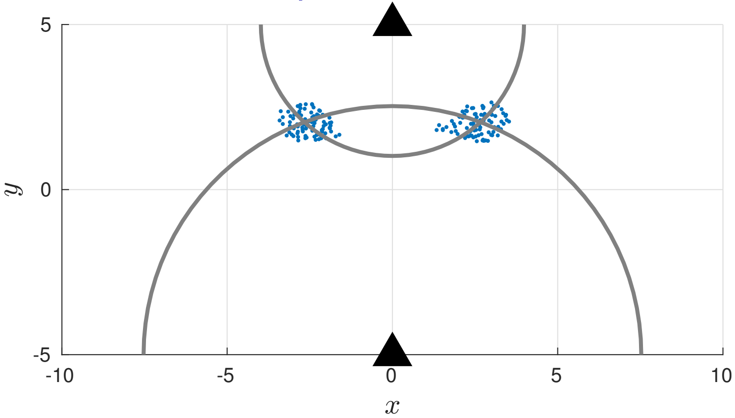
$$p(\mathbf{x}_0) = \text{Unif}\left(\begin{bmatrix} -10 \\ -5 \end{bmatrix}, \begin{bmatrix} 10 \\ 5 \end{bmatrix} \right).$$

A Non-Gaussian Example



- Obtaining a single distance measurement to the top landmark, the distribution of positions lies on a circle.
- Gaussian distributions always look like ellipses, so a Gaussian estimator would do a horrible job here.

A Non-Gaussian Example



- Obtaining a second distance measurement to the bottom landmark, we now have two possible ambiguous locations where the robot could be.
- The distribution is *multi-modal*.

Review: Probability Density Functions

Probability Density Function (PDF)

A continuous PDF is a function $p : \mathbb{R}^n \rightarrow \mathbb{R}$ that satisfies the *axiom of total probability*,

$$\int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{x}) \mathrm{d}\mathbf{x} = 1. \tag{2}$$

If the random variable $\mathbf{x} \in [\mathbf{a}, \mathbf{b}]$ is distributed according to $p(\mathbf{x})$, it is written as $\mathbf{x} \sim p(\mathbf{x})$.

Gaussian PDFs

A Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is denoted as $p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right). \tag{3}$$

The Usual Estimation Setup

- We will assume there exists a **process model** of the form

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}), \qquad \mathbf{w}_{k-1} \sim p(\mathbf{w}_{k-1}). \tag{4}$$

Markov Assumption [2, Ch. 4.1]

The current state $\mathbf{x}_k$ is independent of anything before $k-1$, if the state and input $\mathbf{x}_{k-1}, \mathbf{u}_{k-1}$ are known:

$$p(\mathbf{x}_k | \mathbf{x}_{1:k-1}, \mathbf{u}_{0:k-1}, \mathbf{y}_{0:k-1}) = p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_{k-1}). \tag{5}$$

- We will assume there is a **measurement model** of the form

$$\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k, \mathbf{v}_k), \qquad \mathbf{v}_k \sim p(\mathbf{v}_k). \tag{6}$$

Conditional Independence Assumption [2, Ch. 4.1]

The current measurement $\mathbf{y}_k$ given the current state $\mathbf{x}_k$ is conditionally independent of the measurement and state histories:

$$p(\mathbf{y}_k | \mathbf{x}_{1:k}, \mathbf{y}_{1:k-1}) = p(\mathbf{y}_k | \mathbf{x}_k) \tag{7}$$

The Task of All Estimators

All estimators seek to compute, or represent in some way, the *posterior distribution*

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}), \tag{8}$$

where

- $\mathbf{x}_{0:k} = [\mathbf{x}_0^\mathsf{T} \ \ldots \mathbf{x}_k^\mathsf{T}]^\mathsf{T} = \mathbf{x}$ is the state,
- $\mathbf{y}_{0:k} = \mathbf{y}$ are the output measurements,
- $\mathbf{u}_{0:k-1} = \mathbf{u}$ are the input measurements,
- and we also have some prior information $p(\mathbf{x}_0)$.

The Task of All Estimators

All estimators seek to compute, or represent in some way, the *posterior distribution*

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}), \tag{8}$$

where

- $\mathbf{x}_{0:k} = [\mathbf{x}_0^\mathsf{T} \ \ldots \mathbf{x}_k^\mathsf{T}]^\mathsf{T} = \mathbf{x}$ is the state,
- $\mathbf{y}_{0:k} = \mathbf{y}$ are the output measurements,
- $\mathbf{u}_{0:k-1} = \mathbf{u}$ are the input measurements,
- and we also have some prior information $p(\mathbf{x}_0)$.

When filtering, such as an EKF, the output is information about the current state $\mathbf{x}_k$ only, given all earlier measurements

$$p(\mathbf{x}_k|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}). \tag{9}$$

In general, (9) is an extremely complicated, intractable expression.

Review: Examples of Some Known PDFs

In certain cases, we **do** have nice expressions for some PDFs.

- If we have an initial guess (a prior) of the state with mean $\check{\mathbf{x}}_0$ and covariance $\check{\mathbf{P}}_0$, then

$$p(\mathbf{x}_0) = \mathcal{N}(\check{\mathbf{x}}_0, \check{\mathbf{P}}_0) = \frac{1}{\sqrt{\det(2\pi\check{\mathbf{P}}_0)}} \exp\left(-\frac{1}{2}(\mathbf{x}_0 - \check{\mathbf{x}}_0)^{\mathsf{T}}\check{\mathbf{P}}_0^{-1}(\mathbf{x}_0 - \check{\mathbf{x}}_0)\right). \tag{10}$$

Review: Examples of Some Known PDFs

In certain cases, we **do** have nice expressions for some PDFs.

- If we have an initial guess (a prior) of the state with mean $\check{\mathbf{x}}_0$ and covariance $\check{\mathbf{P}}_0$, then

$$p(\mathbf{x}_0) = \mathcal{N}(\check{\mathbf{x}}_0, \check{\mathbf{P}}_0) = \frac{1}{\sqrt{\det(2\pi\check{\mathbf{P}}_0)}} \exp\left(-\frac{1}{2}(\mathbf{x}_0 - \check{\mathbf{x}}_0)^{\mathsf{T}}\check{\mathbf{P}}_0^{-1}(\mathbf{x}_0 - \check{\mathbf{x}}_0)\right). \quad (10)$$

- If we have a nonlinear process model with additive noise $\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1},\ \mathbf{w}_{k-1} \sim \mathcal{N}(0, \mathbf{Q}_{k-1})$ then

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) = \mathcal{N}(\mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}), \mathbf{Q}_{k-1})$$
$$= \frac{1}{\sqrt{\det(2\pi\mathbf{Q}_{k-1})}} \exp\left(-\frac{1}{2}(\mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}))^{\mathsf{T}}\mathbf{Q}_{k-1}^{-1}(\mathbf{x}_k - \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}))\right). \quad (1$$

Review: Examples of Some Known PDFs

- If we have a nonlinear measurement model with additive noise $\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k) + \mathbf{v}_k,\ \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R}_k)$ then

$$
\begin{aligned}
p(\mathbf{y}_k|\mathbf{x}_k) &= \mathcal{N}(\mathbf{g}(\mathbf{x}_k), \mathbf{R}_k) \\
&= \frac{1}{\sqrt{\det(2\pi\mathbf{R}_k)}} \exp\left(-\frac{1}{2}(\mathbf{y}_k - \mathbf{g}(\mathbf{x}_k))^{\mathsf{T}} \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{g}(\mathbf{x}_k)) \right). \qquad (12)
\end{aligned}
$$

Review: Bayes' Rule, Marginalization

Bayes' Rule

Any joint PDF $p(\mathbf{x}, \mathbf{y})$ can be written as

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y})$$

Review: Bayes' Rule, Marginalization

Bayes' Rule

Any joint PDF $p(\mathbf{x}, \mathbf{y})$ can be written as

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \implies p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{13}$$

The last equation is known as *Bayes' Rule*.

Review: Bayes' Rule, Marginalization

Bayes' Rule

Any joint PDF $p(\mathbf{x}, \mathbf{y})$ can be written as

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \implies p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (13)$$

The last equation is known as *Bayes' Rule*.

Definition (Marginalization)

Recall that *marginalization* refers to integrating a joint PDF $p(\mathbf{x}, \mathbf{y})$ with respect to some of the variables, such as $\mathbf{x}$

$$\int p(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{x} = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} = \int p(\mathbf{x}|\mathbf{y})p(\mathbf{y})\mathrm{d}\mathbf{x} = p(\mathbf{y}) \underbrace{\int p(\mathbf{x}|\mathbf{y})\mathrm{d}\mathbf{x}}_{=1} = p(\mathbf{y}). \quad (14)$$

Review: Bayes' Rule, Marginalization

Bayes' Rule

Any joint PDF $p(\mathbf{x}, \mathbf{y})$ can be written as

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \implies p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{13}$$

The last equation is known as *Bayes' Rule*.

Definition (Marginalization)

Recall that *marginalization* refers to integrating a joint PDF $p(\mathbf{x}, \mathbf{y})$ with respect to some of the variables, such as $\mathbf{x}$

$$\int p(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{x} = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} = \int p(\mathbf{x}|\mathbf{y})p(\mathbf{y})\mathrm{d}\mathbf{x} = p(\mathbf{y}) \underbrace{\int p(\mathbf{x}|\mathbf{y})\mathrm{d}\mathbf{x}}_{=1} = p(\mathbf{y}). \tag{14}$$

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x}} \triangleq \eta p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{15}$$

Review: Bayes' Filter

- Back to our goal of determining the posterior distribution $p(\mathbf{x}_k|\mathbf{y}, \mathbf{u})$, we can use Bayes' rule to write

$$p(\mathbf{x}_k|\mathbf{y}, \mathbf{u}) = \eta p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{u}, \mathbf{y}_{0:k-1}). \tag{16}$$

Review: Bayes' Filter

- Back to our goal of determining the posterior distribution $p(\mathbf{x}_k|\mathbf{y},\mathbf{u})$, we can use Bayes' rule to write

$$p(\mathbf{x}_k|\mathbf{y},\mathbf{u}) = \eta p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{u},\mathbf{y}_{0:k-1}). \tag{16}$$

- For the second term, we can insert a dependence on $\mathbf{x}_{k-1}$ through marginalization,

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{u},\mathbf{y}_{0:k-1}) &= \int p(\mathbf{x}_k,\mathbf{x}_{k-1}|\mathbf{u},\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1} && (17) \\ &= \int p(\mathbf{x}_k|\mathbf{u},\mathbf{y}_{0:k-1},\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{u},\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1} \\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{u}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{u},\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1}. && (18) \end{aligned}$$

Review: Bayes' Filter

- Back to our goal of determining the posterior distribution $p(\mathbf{x}_k|\mathbf{y},\mathbf{u})$, we can use Bayes' rule to write

$$p(\mathbf{x}_k|\mathbf{y},\mathbf{u}) = \eta p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{u},\mathbf{y}_{0:k-1}). \tag{16}$$

- For the second term, we can insert a dependence on $\mathbf{x}_{k-1}$ through marginalization,

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{u},\mathbf{y}_{0:k-1}) &= \int p(\mathbf{x}_k,\mathbf{x}_{k-1}|\mathbf{u},\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1} && (17)\\ &= \int p(\mathbf{x}_k|\mathbf{u},\mathbf{y}_{0:k-1},\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{u},\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1}\\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{u}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{u},\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1}. && (18)\end{aligned}$$

Bayes' Filter

Substituting (18) into (16) gives Bayes' filter,

$$p(\mathbf{x}_k|\mathbf{y},\mathbf{u}) = \eta p(\mathbf{y}_k|\mathbf{x}_k)\int p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{u}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{u},\mathbf{y}_{0:k-1})\mathrm{d}\mathbf{x}_{k-1}. \tag{19}$$

Monte Carlo Integration

- Clearly, we need a method to evaluate generic integrals of the form

$$E[\mathbf{h}(\mathbf{x})] = \int \mathbf{h}(\mathbf{x}) p(\mathbf{x}|\mathbf{y}) \mathrm{d}\mathbf{x}. \tag{20}$$

Monte Carlo Integration

- Clearly, we need a method to evaluate generic integrals of the form

$$E[\mathbf{h}(\mathbf{x})] = \int \mathbf{h}(\mathbf{x}) p(\mathbf{x}|\mathbf{y}) \mathrm{d}\mathbf{x}. \tag{20}$$

- In an ideal Monte Carlo approximation, we can draw samples $\mathbf{x}^{(i)} \sim p(\mathbf{x}|\mathbf{y}),\ i = 1, \ldots, N$ and approximate the integral with

$$E[\mathbf{h}(\mathbf{x})] \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{h}(\mathbf{x}^{(i)}). \tag{21}$$

Monte Carlo Integration Example

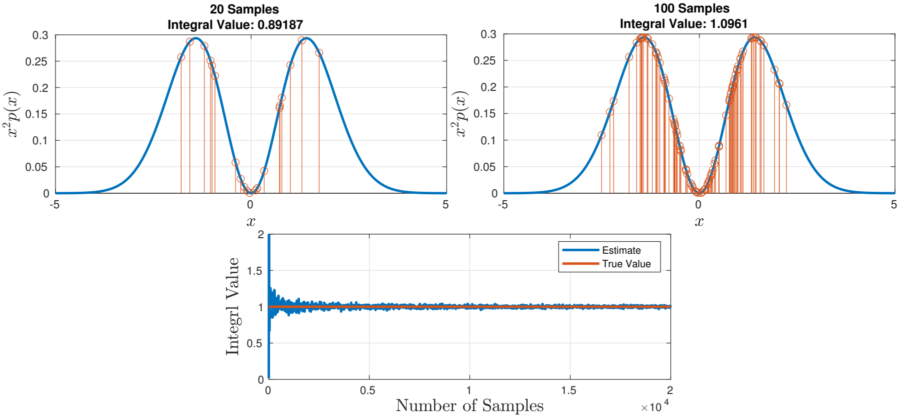


Figure 1: Computation of the integral $\int x^2 p(x) \mathrm{d}x$ where $p(x) = \mathcal{N}(0, 1)$.

Starting Simple: A Prior and One Measurement

Lets start by considering just a **single correction step**. That is, we have access to

- some prior information of our state $p(\mathbf{x}_0)$,
- one measurement $\mathbf{y}_0$ with measurement model,

$$\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0) + \mathbf{v}_0, \qquad \mathbf{v}_0 \sim p(\mathbf{v}_0) \tag{22}$$

  and hence we assume that we know $p(\mathbf{y}_0|\mathbf{x}_0)$.

The posterior distribution is

$$p(\mathbf{x}_0|\mathbf{y}_0). \tag{23}$$

Importance Sampling

- In this case, integrals to compute will be of the form $\int \mathbf{h}(\mathbf{x}_0)p(\mathbf{x}_0|\mathbf{y}_0)\mathrm{d}\mathbf{x}_0$.
- Unfortunately, even sampling from $p(\mathbf{x}_0|\mathbf{y}_0)$ is difficult, if not impossible.
- Hence, we will introduce an *importance distribution* $\pi(\mathbf{x}_0|\mathbf{y}_0)$ that we **can** sample from,

$$\int \mathbf{h}(\mathbf{x}_0)p(\mathbf{x}_0|\mathbf{y}_0)\mathrm{d}\mathbf{x}_0 = \int \left(\mathbf{h}(\mathbf{x}_0)\frac{p(\mathbf{x}_0|\mathbf{y}_0)}{\pi(\mathbf{x}_0|\mathbf{y}_0)}\right)\pi(\mathbf{x}_0|\mathbf{y}_0)\mathrm{d}\mathbf{x}_0. \tag{24}$$

- As such, after sampling $\mathbf{x}^{(i)} \sim \pi(\mathbf{x}_0|\mathbf{y}_0)$, the integral can be approximated with

$$E[\mathbf{h}(\mathbf{x}_0)] = \frac{1}{N}\sum_{i=1}^{N}\frac{p(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}{\pi(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}\mathbf{h}(\mathbf{x}_0^{(i)}) \tag{25}$$

$$\triangleq \sum_{i=1}^{N} w^{(i)}\mathbf{h}(\mathbf{x}_0^{(i)}), \qquad w^{(i)} = \frac{1}{N}\frac{p(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}{\pi(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}. \tag{26}$$

Importance Sampling

- …except we cannot even evaluate $p(\mathbf{x}_0^{(i)}|\mathbf{y}_0)$, in general.
- But, using Bayes' rule,

$$p(\mathbf{x}_0|\mathbf{y}_0) = \frac{p(\mathbf{y}_0|\mathbf{x}_0)p(\mathbf{x}_0)}{\int p(\mathbf{y}_0|\mathbf{x}_0)p(\mathbf{x}_0)\mathrm{d}\mathbf{x}_0}. \tag{27}$$

Importance Sampling

- …except we cannot even evaluate $p(\mathbf{x}_0^{(i)}|\mathbf{y}_0)$, in general.
- But, using Bayes' rule,

$$p(\mathbf{x}_0|\mathbf{y}_0) = \frac{p(\mathbf{y}_0|\mathbf{x}_0)p(\mathbf{x}_0)}{\int p(\mathbf{y}_0|\mathbf{x}_0)p(\mathbf{x}_0)\mathrm{d}\mathbf{x}_0}. \tag{27}$$

- Hence,

$$E[\mathbf{h}(\mathbf{x}_0)] = \int \mathbf{h}(\mathbf{x}_0)p(\mathbf{x}_0|\mathbf{y}_0)\mathrm{d}\mathbf{x}_0 = \frac{\int \mathbf{h}(\mathbf{x}_0)p(\mathbf{y}_0|\mathbf{x}_0)p(\mathbf{x}_0)\mathrm{d}\mathbf{x}_0}{\int p(\mathbf{y}_0|\mathbf{x}_0)p(\mathbf{x}_0)\mathrm{d}\mathbf{x}_0} \tag{28}$$

$$= \frac{\int \left(\frac{p(\mathbf{y}_0|\mathbf{x}_0)p(\mathbf{x}_0)}{\pi(\mathbf{x}_0|\mathbf{y}_0)}\mathbf{h}(\mathbf{x}_0)\right)\pi(\mathbf{x}_0|\mathbf{y}_0)\mathrm{d}\mathbf{x}_0}{\int \left(\frac{p(\mathbf{y}_0|\mathbf{x}_0)p(\mathbf{x}_0)}{\pi(\mathbf{x}_0|\mathbf{y}_0)})\right)\pi(\mathbf{x}_0|\mathbf{y}_0)\mathrm{d}\mathbf{x}_0} \tag{29}$$

$$\approx \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{p(\mathbf{y}_0|\mathbf{x}_0^{(i)})p(\mathbf{x}_0^{(i)})}{\pi(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}\mathbf{h}(\mathbf{x}_0^{(i)})}{\frac{1}{N}\sum_{i=1}^{N}\frac{p(\mathbf{y}_0|\mathbf{x}_0^{(i)})p(\mathbf{x}_0^{(i)})}{\pi(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}} \tag{30}$$

Importance Sampling

$$E[\mathbf{h}(\mathbf{x}_0)] \approx \sum_{i=1}^{N} \left(\frac{\frac{p(\mathbf{y}_0|\mathbf{x}_0^{(i)})p(\mathbf{x}_0^{(i)})}{\pi(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}}{\sum_{i=1}^{N} \frac{p(\mathbf{y}_0|\mathbf{x}_0^{(i)})p(\mathbf{x}_0^{(i)})}{\pi(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}} \right) \mathbf{h}(\mathbf{x}_0^{(i)}) \tag{31}$$

$$= \sum_{i=1}^{N} \underbrace{\left(\frac{w^{*(i)}}{\sum_{i=1}^{N} w^{*(i)}} \right)}_{w^{(i)}} \mathbf{h}(\mathbf{x}_0^{(i)}) \tag{32}$$

where the *un-normalized weights* are defined as

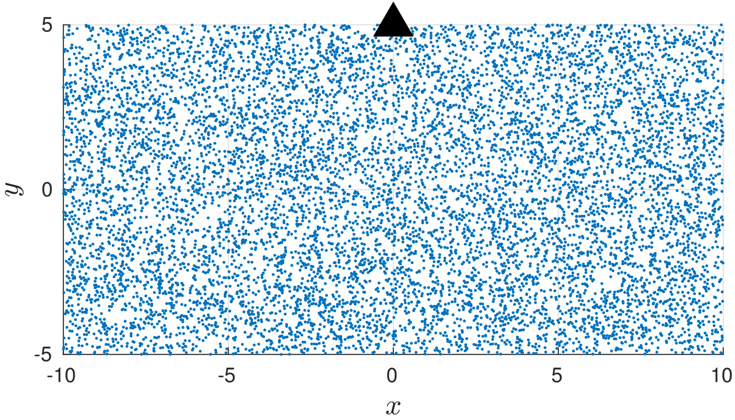
$$w^{*(i)} = \frac{p(\mathbf{y}_0|\mathbf{x}_0^{(i)})p(\mathbf{x}_0^{(i)})}{\pi(\mathbf{x}_0^{(i)}|\mathbf{y}_0)}. \tag{33}$$

- At last, (32) is something we can compute. The posterior can also be approximated as

$$p(\mathbf{x}_0|\mathbf{y}_0) \approx \sum_{i=1}^{N} w^{(i)} \delta(\mathbf{x}_0 - \mathbf{x}_0^{(i)}) \tag{34}$$

where $\delta(\cdot)$is the Dirac delta function.

Importance Sampling Example

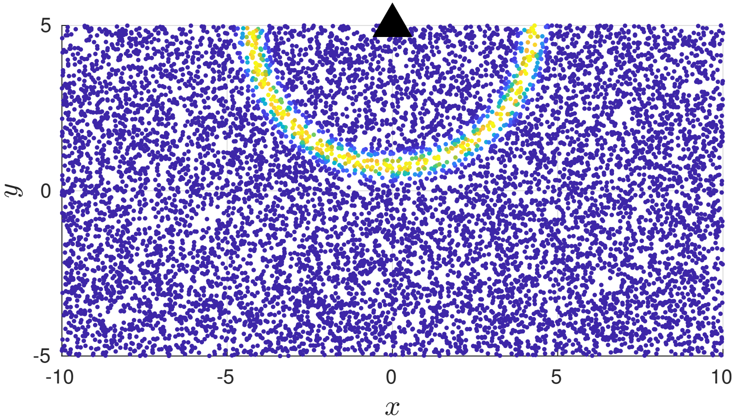


- We will use the earlier example, choosing

$$\pi(\mathbf{x}_0|\mathbf{y}_0) = p(\mathbf{x}_0) = \mathrm{Unif}([-10\ \ -5]^\mathsf{T}, [10, 5]^\mathsf{T}). \tag{35}$$

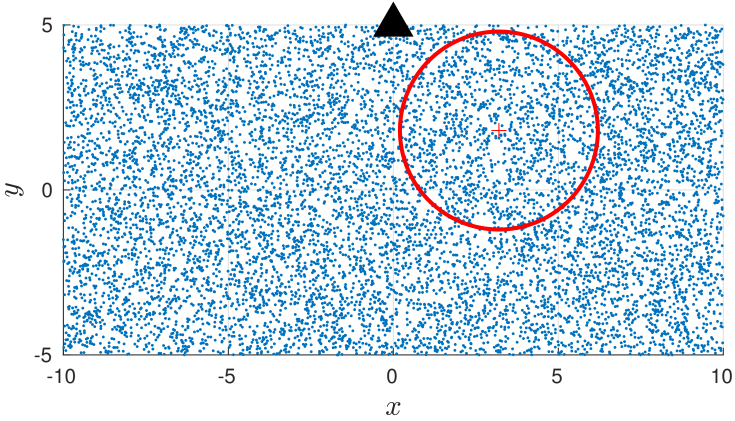
and we receive a distance measurement y_0 to the top landmark.

Importance Sampling Example



- The samples' color have been scaled according to their weight $w^{(i)}$.

Importance Sampling Example 2

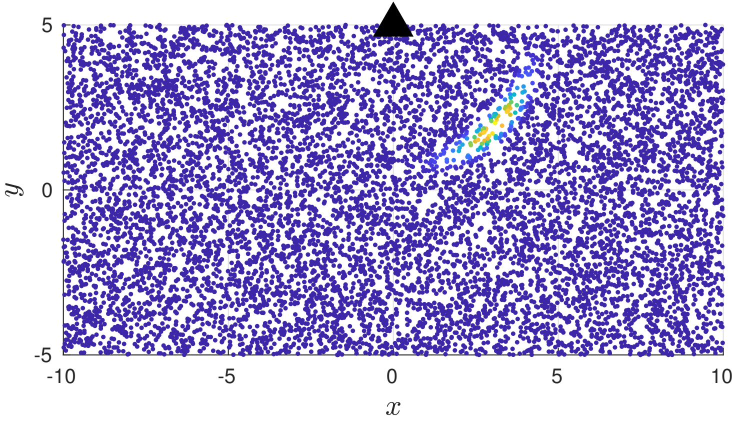


- As another example, we choose

$$\pi(\mathbf{x}_0|\mathbf{y}_0) = \mathrm{Unif}([-10\ \ -5]^\mathsf{T}, [10, 5]^\mathsf{T}). \tag{36}$$

and we have a prior $p(\mathbf{x}_0) = \mathcal{N}([3.2\ 1.8]^\mathsf{T}, \mathbf{1})$.

Importance Sampling Example 2



- The samples' color have been scaled according to their weight $w^{(i)}$.
- Now, what about the case with multiple measurements?

Sequential Importance Sampling

- Let's generalize the previous importance sampling procedure to the posterior given many measurements $p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}) = p(\mathbf{x}|\mathbf{y}, \mathbf{u})$.
- Using the Markov and conditional independence assumptions, as well as Bayes' rule,

$$p(\mathbf{x}|\mathbf{y}, \mathbf{u}) = \eta p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) p(\mathbf{x}_{0:k-1}|\mathbf{y}_{0:k-1}, \mathbf{u}). \tag{37}$$

- Repeating the same importance sampling derivation as before will eventually give the following un-normalized weights

$$w_k^{*(i)} = \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(i)}) p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_{k-1}) p(\mathbf{x}_{0:k-1}^{(i)}|\mathbf{y}_{0:k-1}, \mathbf{u})}{\pi(\mathbf{x}^{(i)}|\mathbf{y}, \mathbf{u})} \tag{38}$$

Sequential Importance Sampling

- The importance distribution can be written as,

$$\pi(\mathbf{x}|\mathbf{y},\mathbf{u}) = \pi(\mathbf{x}_k|\mathbf{x}_{0:k-1},\mathbf{y},\mathbf{u})\pi(\mathbf{x}_{0:k-1}|\mathbf{y}_{0:k-1},\mathbf{u}). \tag{39}$$

- Thus the un-normalized weights can be written as

$$w_k^{*(i)} = \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{u}_{k-1})}{\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{0:k-1}^{(i)},\mathbf{y},\mathbf{u})} \underbrace{\frac{p(\mathbf{x}_{0:k-1}^{(i)}|\mathbf{y}_{0:k-1},\mathbf{u})}{\pi(\mathbf{x}_{0:k-1}^{(i)}|\mathbf{y}_{0:k-1},\mathbf{u})}}_{\triangleq w_{k-1}^{(i)}}, \tag{40}$$

$$w_k^{*(i)} = w_{k-1}^{(i)} \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{u}_{k-1})}{\pi(\mathbf{x}_k^{(i)}|\mathbf{x}_{0:k-1}^{(i)},\mathbf{y},\mathbf{u})}, \tag{41}$$

  after which, they should be normalized to sum to 1.

- We have a recursive expression, where the weights are “updated”.

"Bootstrapping" the Importance Distribution

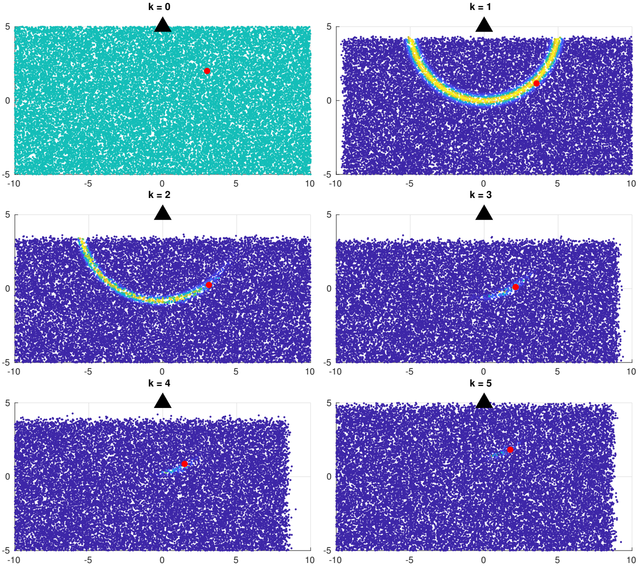
- If we choose $\pi(\mathbf{x}_k|\mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}, \mathbf{u}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_{k-1})$ as the importance distribution, the weights becomes

$$w_k^{*(i)} = w_{k-1}^{(i)} \frac{p(\mathbf{y}_k|\mathbf{x}_k^{(i)})p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_{k-1})}{p(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_{k-1})}, \tag{42}$$

$$= w_{k-1}^{(i)} p(\mathbf{y}_k|\mathbf{x}_k^{(i)}). \tag{43}$$

- This forms the basis of the most popular particle filter, the *bootstrap particle filter*.

Sequential Importance Sampling Example



Red dot is the true position.

$$p(\mathbf{x}_0) = \text{Unif}\left(\begin{bmatrix} -10 \\ -5 \end{bmatrix}, \begin{bmatrix} 10 \\ 5 \end{bmatrix}\right)$$

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \Delta t \mathbf{u}_{k-1}$$

$$\mathbf{u}(t) = [\cos(t) \;\; -\sin(t)]^{\mathsf{T}}$$

$$y_k = \left\| \mathbf{x}_k - \mathbf{r}_a^{\ell w} \right\|$$

Resampling

- As is, the filter will have a *degeneracy problem* [2, 3].
- That is, almost all of the weights will go to 0, except one, which will go to 1.

Resampling

- As is, the filter will have a *degeneracy problem* [2, 3].
- That is, almost all of the weights will go to 0, except one, which will go to 1.
- This problem can be solved by resampling.
  - Make copies of samples with high weights.
  - Discard samples with low weights.
- Interpret the weights as the probability of making a copy of the sample.

Resampling

- As is, the filter will have a *degeneracy problem* [2, 3].
- That is, almost all of the weights will go to 0, except one, which will go to 1.
- This problem can be solved by resampling.
  - Make copies of samples with high weights.
  - Discard samples with low weights.
- Interpret the weights as the probability of making a copy of the sample.
- There are many resampling strategies [3]:
  - multinomial resampling;
  - residual resampling;
  - stratified resampling;
  - **systematic resampling**.

Systematic Resampling

From N normalized weights $w^{(i)}$, systematic resampling proceeds as follows.

1. Create bins with boundaries β_m according to $\beta_m = \sum_{i=1}^{m} w^{(i)}$.
2. Select a random number $\Delta \sim \mathrm{Unif}(0, 1/N)$.
3. Draw N new samples using the look-up values

$$\ell_j = \Delta + j(1/N), \qquad j = 0, \ldots, N-1 \tag{44}$$

   and choosing the sample whose bin contains ℓ_j.
4. Reset all the weights to $w^{(i)} = 1/N$.

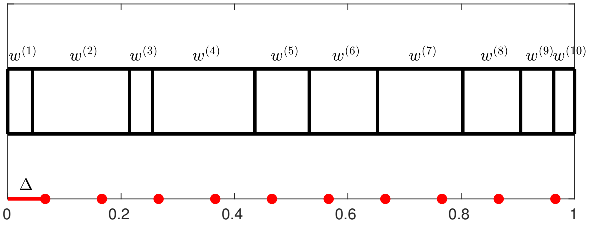


Figure 2: Systematic resampling schematic for $N = 10$. Red dots are ℓ_j values.

Systematic Resampling

This completes the plain-vanilla *bootstrap particle filter*.

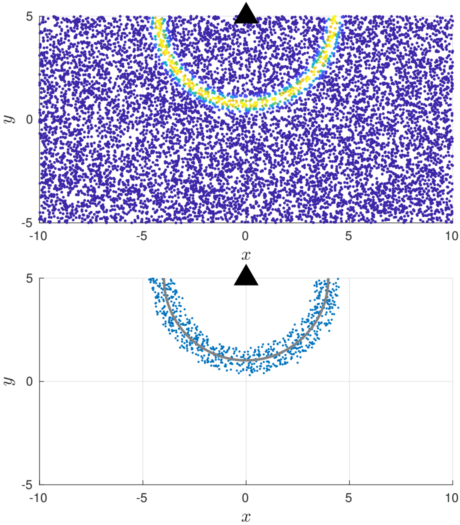


Figure 3: (top) Before resampling. (bottom) After resampling.

Sample Impoverishment

- Resampling can occasionally result in *sample impoverishment*.
- We end up with a large amount of copies of just a few samples.
- This often happens when process noise is low.
- Suggestions to fix this problem can be found in [3].

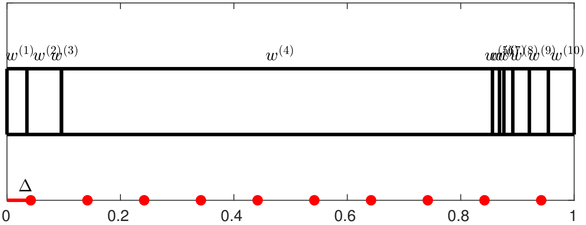
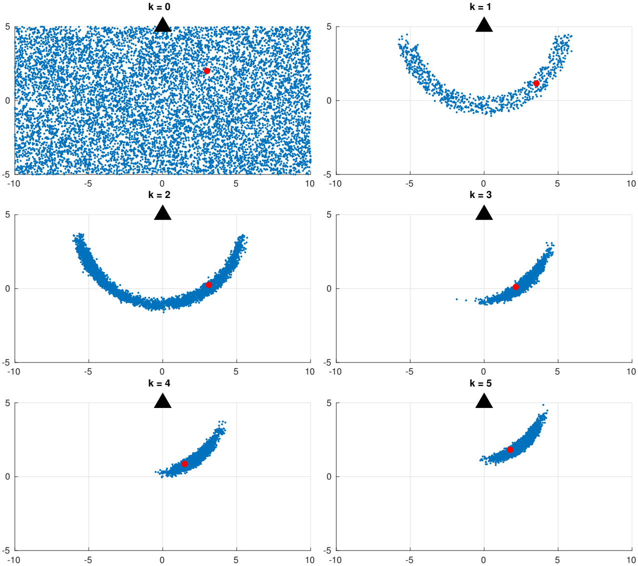


Figure 4: Excessive copies of a single sample, results in loss of diversity.

Sequential Importance Resampling (Particle Filter)



Red dot is the true position.

$$p(\mathbf{x}_0) = \text{Unif}\left(\begin{bmatrix} -10 \\ -5 \end{bmatrix}, \begin{bmatrix} 10 \\ 5 \end{bmatrix}\right)$$

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \Delta t \mathbf{u}_{k-1}$$

$$\mathbf{u}(t) = [\cos(t) \ \ -\sin(t)]^\mathsf{T}$$

$$y_k = \left\|\mathbf{x}_k - \mathbf{r}_a^{\ell w}\right\|$$

Summary

Bootstrap Particle Filter (resampling at every step)

Assuming that we have $\mathbf{x}_{k-1}^{(i)},\ i = 1, \ldots, N$ samples from the previous time step, which together represent $p(\mathbf{x}_{k-1}|\mathbf{y}_{0:k-1}, \mathbf{u}_{0:k-2})$, the PF proceeds as follows.

Predict:

1. Draw N noise samples $\mathbf{w}_k^{(i)}$ from $p(\mathbf{w}_k)$.
2. Compute the "predicted particles" with

$$\mathbf{x}_k^{(i)} = \mathbf{f}(\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_{k-1}, \mathbf{w}_k^{(i)}), \qquad i = 1, \ldots, N, \tag{45}$$

   which now approximate $p(\mathbf{x}_k|\mathbf{y}_{0:k-1}, \mathbf{u}_{0:k-1})$.

Correct:

1. Compute the un-normalized weights as

$$w_k^{*(i)} = p(\mathbf{y}_k|\mathbf{x}_k^{(i)}) \tag{46}$$

   and normalize them to sum to 1.
2. Do resampling.

Advantages and Disadvantages

Advantages:

- Easy to implement.
- Does not require analytical expressions for $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$, nor their derivatives.
- Works with any noise distribution, not just Gaussian.
- Can represent non-Gaussian posteriors.

Disadvantages:

- It has its own issues, such as sample impoverishment.
- Computationally demanding. For comparison:
  - EKF requires 1 function evaluation of $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$;
  - UKF requires $2L + 1$ (usually 30-50) function evaluations $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ where $L = \dim(\mathbf{x}_k) + \dim(\mathbf{w}_k)$;
  - PF requires N (anywhere from 500-50000+) $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ evaluations.

References

These slides are based on [2–4]

[1] T. D. Barfoot, J. R. Forbes, and D. Yoon, "Exactly Sparse Gaussian Variational Inference with Application to Derivative-Free Batch Nonlinear State Estimation (preprint),", vol. 1, no. 1, pp. 1–31, 2019. arXiv: 1911.08333. [Online]. Available: http://arxiv.org/abs/1911.08333.

[2] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2010, pp. 1–232.

[3] J. Elfring, E. Torta, and R. V. D. Molengraft, "Particle Filters : A Hands-On Tutorial," *Sensors*, vol. 21, no. 438, pp. 1–28, 2021.

[4] T. Barfoot, *State Estimation for Robotics*. Toronto, ON: Cambridge University Press, 2019.