# Monte Carlo Methods in Filtering

Vassili Korotkine

McGill University, Department of Mechanical Engineering

May 21, 2024

# Robot Localization

- Where am I?


[1] Image taken from https://www.hiig.de/en/robots-be-like-buddha-why-we-think-wall-e-and-bb8-are-cute-and-fortune-teller-robots-are-creepy/

# Robot Localization

- Where am I?
- Given: Noisy sensor measurements $\mathbf{y}_k$.


[1] Image taken from `https://www.hiig.de/en/robots-be-like-buddha-why-we-think-wall-e-and-bb8-are-cute-and-fortune-teller-robots-are-creepy/`

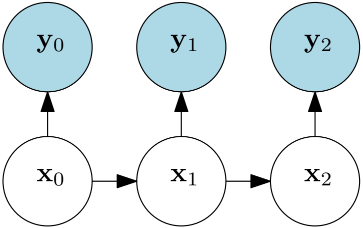
# Robot Localization

- Where am I?
- Given: Noisy sensor measurements $\mathbf{y}_k$.
- Want: Uncertain robot state at given timestep $\mathbf{x}_k$.


[1] Image taken from https://www.hiig.de/en/robots-be-like-buddha-why-we-think-wall-e-and-bb8-are-cute-and-fortune-teller-robots-are-creepy/

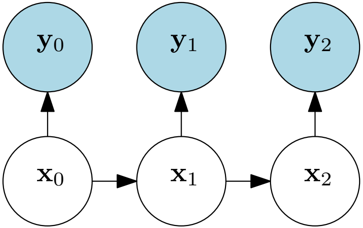
# Problem Structure

- Hidden Markov Model.



# Problem Structure

- Hidden Markov Model.

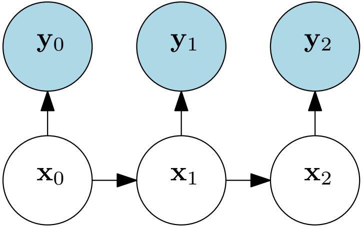



- Conditional measurement independence: Each measurement $\mathbf{y}_k$ is conditionally independent given the state it depends on,

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

# Problem Structure

- Hidden Markov Model.



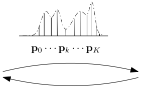
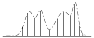
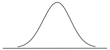

- Conditional measurement independence: Each measurement $\mathbf{y}_k$ is conditionally independent given the state it depends on,

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

- Markov assumption: Each state is conditionally independent of previous measurements given the previous hidden state.

$$p(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{y}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}). \tag{2}$$

# Non-Exhaustive Taxonomy of Estimation Methods

| | Non-Parametric | Parametric |
|---|---|---|
| Smoothed $p(\mathbf{x}_k\|\mathbf{y}_{1:\boxed{K}})$ | Particle Smoother  $\mathbf{p}_0 \cdots \mathbf{p}_k \cdots \mathbf{p}_K$ <br> $\mathbf{x}_0 \; \cdots \; \mathbf{x}_k \; \cdots \; \mathbf{x}_N$ <br> Normalizing Flow <br> iSAM <br> Nested Sampling on Factor Graphs | Batch  $\mathbf{x}_0 \; \cdots \; \mathbf{x}_k \; \cdots \; \mathbf{x}_N$ <br> Sliding Window  $\mathbf{x}_0 \; \mathbf{x}_1 \; \cdots \; \mathbf{x}_{k-2} \; \mathbf{x}_{k+2} \; \cdots \; \mathbf{x}_N$ <br> RTS Smoother  $\mathbf{x}_0 \; \cdots \; \mathbf{x}_k \; \cdots \; \mathbf{x}_N$ |
| Filtering $p(\mathbf{x}_k\|\mathbf{y}_{1:\boxed{k}})$ | Particle Filter  $\mathbf{p}_0 \cdots \mathbf{p}_k \cdots \mathbf{p}_K$ | Gaussian Filter  $\left\{\begin{array}{l}\text{Extended (linearization)} \\ \text{Sigma-point (numerical integration)}\end{array}\right.$ |

# Bayes Filter

- Seek *marginal* distribution of $\mathbf{x}_k$, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$.

# Bayes Filter

- Seek *marginal* distribution of $\mathbf{x}_k$, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$.
- Bayes rule for general random variables $\mathbf{x}$ and $\mathbf{y}$

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{\eta}p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{3}$$

# Bayes Filter

- Seek *marginal* distribution of $\mathbf{x}_k$, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$.
- Bayes rule for general random variables $\mathbf{x}$ and $\mathbf{y}$

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{\eta} p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}). \tag{3}$$

- For the filtering case

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}). \tag{4}$$

# Bayes Filter

- Seek *marginal* distribution of $\mathbf{x}_k$, $p(\mathbf{x}_k|\mathbf{y}_{1:k})$.
- Bayes rule for general random variables $\mathbf{x}$ and $\mathbf{y}$

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{\eta}p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \tag{3}$$

- For the filtering case

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta}p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1}). \tag{4}$$

- Reverse marginalization gives *Chapman-Kolmogorov* equation

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) &= \int p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\mathrm{d}\mathbf{x}_{k-1} & (5) \\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\mathrm{d}\mathbf{x}_{k-1}. & (6) \end{aligned}$$

# Bayes Filter

- The marginal $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is then

$$
\begin{aligned}
p(\mathbf{x}_k|\mathbf{y}_{1:k}) &= \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k}) &(7)\\
&= \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1}. &(8)
\end{aligned}
$$

# Bayes Filter

- The marginal $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is then

$$
\begin{aligned}
p(\mathbf{x}_k|\mathbf{y}_{1:k}) &= \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k}) && (7) \\
&= \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \underbrace{\int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1}.}_{\text{Prediction}} && (8)
\end{aligned}
$$

# Bayes Filter

- The marginal $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is then

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

$$= \frac{1}{\eta} \underbrace{p(\mathbf{y}_k|\mathbf{x}_k)}_{\text{Correction}} \underbrace{\int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1}.}_{\text{Prediction}} \tag{8}$$

# Bayes Filter

- The marginal $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is then

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

$$= \frac{1}{\eta}\underbrace{p(\mathbf{y}_k|\mathbf{x}_k)}_{\text{Correction}}\underbrace{\int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\mathrm{d}\mathbf{x}_{k-1}}_{\text{Prediction}}. \tag{8}$$

- The prediction integral is intractable in general, as is the normalization constant.

# Bayes Filter

- The marginal $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is then

$$
\begin{aligned}
p(\mathbf{x}_k|\mathbf{y}_{1:k}) &= \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k}) && (7) \\
&= \frac{1}{\eta} \underbrace{p(\mathbf{y}_k|\mathbf{x}_k)}_{\text{Correction}} \underbrace{\int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1}}_{\text{Prediction}}. && (8)
\end{aligned}
$$

- The prediction integral is intractable in general, as is the normalization constant.
- Choice of how to parametrize state belief.

# Bayes Filter

- The marginal $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is then

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

$$= \frac{1}{\eta} \underbrace{p(\mathbf{y}_k|\mathbf{x}_k)}_{\text{Correction}} \underbrace{\int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1}.}_{\text{Prediction}} \tag{8}$$

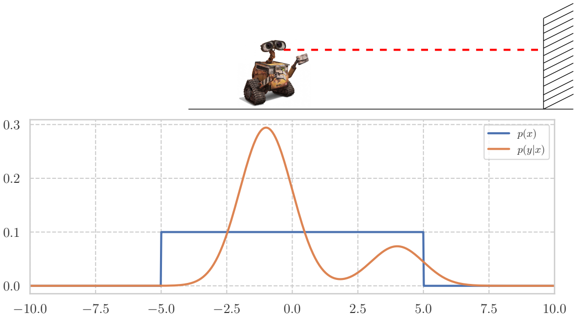
- The prediction integral is intractable in general, as is the normalization constant.
- Choice of how to parametrize state belief.
  - Parametric: Gaussian, multimodal.
  - Non-parametric: Particles.

# Non-Gaussian Belief

- Gaussian filter, nonlinear-least-squares optimization $\rightarrow$ *Gaussian belief*.
- Particle filtering $\rightarrow$ *non-Gaussian beliefs about the state estimate*.

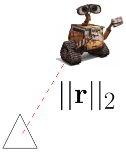
# Non-Gaussian Belief

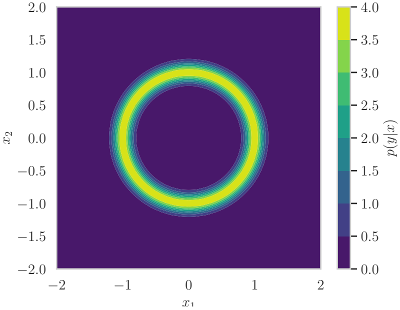
- Gaussian filter, nonlinear-least-squares optimization $\rightarrow$ *Gaussian belief*.
- Particle filtering $\rightarrow$ *non-Gaussian beliefs about the state estimate*.
  - Non-Gaussian sensor noise



# Non-Gaussian Belief

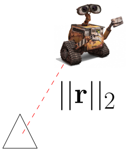
- Gaussian filter, nonlinear-least-squares optimization $\rightarrow$ *Gaussian belief*.
- Particle filtering $\rightarrow$ *non-Gaussian beliefs about the state estimate*.
  - Non-Gaussian sensor noise
  - Range-only localization

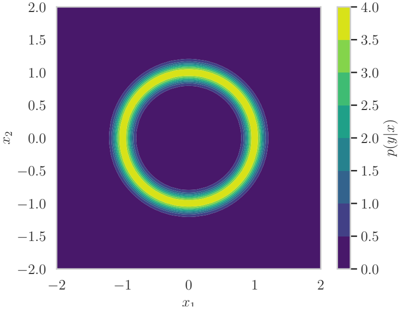


# Non-Gaussian Belief

- Gaussian filter, nonlinear-least-squares optimization $\rightarrow$ *Gaussian belief*.
- Particle filtering $\rightarrow$ *non-Gaussian beliefs about the state estimate*.
  - Non-Gaussian sensor noise
  - Range-only localization
  - Strongly nonlinear models, ambiguous data associations, loop closures



# Particle Representations of Probability Distribution Functions (PDFs)

- What can we do with a PDF?

# Particle Representations of Probability Distribution Functions (PDFs)

- What can we do with a PDF?

$$\mathrm{E}[\mathbf{f}(\mathbf{x})] = \int \mathbf{f}(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{9}$$

  Compute expectations!

# Particle Representations of Probability Distribution Functions (PDFs)

- What can we do with a PDF?

$$\mathrm{E}[\mathbf{f}(\mathbf{x})] = \int \mathbf{f}(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{9}$$

  Compute expectations!

- Examples:

$$\mathbf{f}(\mathbf{x}) = \mathbf{x} \rightarrow \mathrm{E}[\mathbf{x}]$$

  where $\mathcal{D}$ is a domain and $I(\mathbf{x} \in \mathcal{D})$ is the indicator function.

# Particle Representations of Probability Distribution Functions (PDFs)

- What can we do with a PDF?

$$\mathrm{E}[\mathbf{f}(\mathbf{x})] = \int \mathbf{f}(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{9}$$

  Compute expectations!

- Examples:

$$\mathbf{f}(\mathbf{x}) = \mathbf{x} \to \mathrm{E}[\mathbf{x}]$$
$$\mathbf{f}(\mathbf{x}) = (\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^{\mathsf{T}} \to \mathrm{Cov}[\mathbf{x}]$$

  where $\mathcal{D}$ is a domain and $I(\mathbf{x} \in \mathcal{D})$ is the indicator function.

# Particle Representations of Probability Distribution Functions (PDFs)

- What can we do with a PDF?

$$\mathrm{E}[\mathbf{f}(\mathbf{x})] = \int \mathbf{f}(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{9}$$

  Compute expectations!

- Examples:

$$\begin{aligned}
\mathbf{f}(\mathbf{x}) &= \mathbf{x} \rightarrow \mathrm{E}[\mathbf{x}] \\
\mathbf{f}(\mathbf{x}) &= (\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^{\mathsf{T}} \rightarrow \mathrm{Cov}[\mathbf{x}] \\
\mathbf{f}(\mathbf{x}) &= I(\mathbf{x} \in \mathcal{D}) \rightarrow P(\mathbf{x} \in \mathcal{D}),
\end{aligned}$$

  where $\mathcal{D}$ is a domain and $I(\mathbf{x} \in \mathcal{D})$ is the indicator function.

# Particle Representations of Probability Distribution Functions (PDFs)

- Numerical approximation

$$\int \mathbf{f}(\mathbf{x}) p(\mathbf{x}) \mathrm{d}\mathbf{x} \approx \sum_{i=1}^{N} w_i \mathbf{f}(\mathbf{x}_i) \tag{10}$$

(11)

# Particle Representations of Probability Distribution Functions (PDFs)

- Numerical approximation

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \sum_{i=1}^{N} w_i \mathbf{f}(\mathbf{x}_i) \tag{10}$$

$$= \int \mathbf{f}(\mathbf{x}) \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i)\mathrm{d}\mathbf{x}. \tag{11}$$

# Particle Representations of Probability Distribution Functions (PDFs)

- Numerical approximation

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \sum_{i=1}^{N} w_i \mathbf{f}(\mathbf{x}_i) \tag{10}$$

$$= \int \mathbf{f}(\mathbf{x}) \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i)\mathrm{d}\mathbf{x}. \tag{11}$$

- $\delta(\mathbf{x} - \mathbf{x}_i)$ is the Dirac delta, which has the *sifting property*

$$\int \mathbf{f}(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}_i)\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}_i). \tag{12}$$

# Particle Representations of Probability Distribution Functions (PDFs)

- Numerical approximation

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \sum_{i=1}^{N} w_i \mathbf{f}(\mathbf{x}_i) \tag{10}$$

$$= \int \mathbf{f}(\mathbf{x}) \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i)\mathrm{d}\mathbf{x}. \tag{11}$$

- $\delta(\mathbf{x} - \mathbf{x}_i)$ is the Dirac delta, which has the *sifting property*

$$\int \mathbf{f}(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}_i)\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}_i). \tag{12}$$

- PDF expressed as

$$\boxed{p(\mathbf{x}) \approx \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i), \quad \sum_{i=1}^{N} w_i = 1.} \tag{13}$$

# General Sampling Methods: Importance Sampling

- If able to directly sample $p(\mathbf{x})$,

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_i), \quad \mathbf{x}_i \sim p(\mathbf{x}). \tag{14}$$

# General Sampling Methods: Importance Sampling

- If able to directly sample $p(\mathbf{x})$,

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_i), \quad \mathbf{x}_i \sim p(\mathbf{x}). \tag{14}$$

- Typically impossible.

# General Sampling Methods: Importance Sampling

- If able to directly sample $p(\mathbf{x})$,

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_i), \quad \mathbf{x}_i \sim p(\mathbf{x}). \tag{14}$$

- Typically impossible.
- Typically only know nominator,

$$p(\mathbf{x}) = \frac{1}{\eta}\tilde{p}(\mathbf{x}). \tag{15}$$

# General Sampling Methods: Importance Sampling

- If able to directly sample $p(\mathbf{x})$,

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_i), \quad \mathbf{x}_i \sim p(\mathbf{x}). \tag{14}$$

- Typically impossible.
- Typically only know nominator,

$$p(\mathbf{x}) = \frac{1}{\eta}\tilde{p}(\mathbf{x}). \tag{15}$$

A solution: *proposal distribution* $q(\mathbf{x})$,

$$\mathrm{E}[\mathbf{f}(\mathbf{x})] = \frac{1}{\eta}\int \mathbf{f}(\mathbf{x})\tilde{p}(\mathbf{x})\mathrm{d}\mathbf{x} \tag{16}$$

(17)

(18)

# General Sampling Methods: Importance Sampling

- If able to directly sample $p(\mathbf{x})$,

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_i), \quad \mathbf{x}_i \sim p(\mathbf{x}). \tag{14}$$

- Typically impossible.
- Typically only know nominator,

$$p(\mathbf{x}) = \frac{1}{\eta}\tilde{p}(\mathbf{x}). \tag{15}$$

A solution: *proposal distribution* $q(\mathbf{x})$,

$$\mathrm{E}[\mathbf{f}(\mathbf{x})] = \frac{1}{\eta}\int \mathbf{f}(\mathbf{x})\tilde{p}(\mathbf{x})\mathrm{d}\mathbf{x} \tag{16}$$

$$= \frac{1}{\eta}\int \mathbf{f}(\mathbf{x})\frac{\tilde{p}(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})\mathrm{d}\mathbf{x} \tag{17}$$

$$\tag{18}$$

# General Sampling Methods: Importance Sampling

- If able to directly sample $p(\mathbf{x})$,

$$\int \mathbf{f}(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}(\mathbf{x}_i), \quad \mathbf{x}_i \sim p(\mathbf{x}). \tag{14}$$

- Typically impossible.
- Typically only know nominator,

$$p(\mathbf{x}) = \frac{1}{\eta}\tilde{p}(\mathbf{x}). \tag{15}$$

A solution: *proposal distribution* $q(\mathbf{x})$,

$$\begin{aligned}
\mathrm{E}[\mathbf{f}(\mathbf{x})] &= \frac{1}{\eta}\int \mathbf{f}(\mathbf{x})\tilde{p}(\mathbf{x})\mathrm{d}\mathbf{x} &(16)\\
&= \frac{1}{\eta}\int \mathbf{f}(\mathbf{x})\frac{\tilde{p}(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})\mathrm{d}\mathbf{x} &(17)\\
&\approx \frac{1}{\eta}\frac{1}{N}\sum \mathbf{f}(\mathbf{x}_i)\frac{\tilde{p}(\mathbf{x}_i)}{q(\mathbf{x}_i)}, \quad \mathbf{x}_i \sim q(\mathbf{x}). &(18)
\end{aligned}$$

# Importance Sampling Normalization Constant

- Evaluating $\eta$

$$p(\mathbf{x}) = \frac{1}{\eta}\tilde{p}(\mathbf{x}). \tag{19}$$

# Importance Sampling Normalization Constant

- Evaluating $\eta$

$$p(\mathbf{x}) = \frac{1}{\eta}\tilde{p}(\mathbf{x}). \tag{19}$$

- $\eta = \int \tilde{p}(\mathbf{x})\mathrm{d}\mathbf{x} \rightarrow$ Importance sampling approximation

$$\eta = \int \tilde{p}(\mathbf{x})\mathrm{d}\mathbf{x} = \int \frac{\tilde{p}(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x})\mathrm{d}\mathbf{x} = \frac{1}{N}\sum_{i=1}^{N} \frac{\tilde{p}(\mathbf{x}_i)}{q(\mathbf{x})}, \quad \mathbf{x}_i \sim q(\mathbf{x}). \tag{20}$$

# Importance Sampling Normalization Constant

- Evaluating $\eta$

$$p(\mathbf{x}) = \frac{1}{\eta}\tilde{p}(\mathbf{x}). \tag{19}$$

- $\eta = \int \tilde{p}(\mathbf{x})\mathrm{d}\mathbf{x} \to$ Importance sampling approximation

$$\eta = \int \tilde{p}(\mathbf{x})\mathrm{d}\mathbf{x} = \int \frac{\tilde{p}(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x})\mathrm{d}\mathbf{x} = \frac{1}{N}\sum_{i=1}^{N} \frac{\tilde{p}(\mathbf{x}_i)}{q(\mathbf{x})}, \quad \mathbf{x}_i \sim q(\mathbf{x}). \tag{20}$$

- Unnormalized weights $\tilde{w}_i = \frac{\tilde{p}(\mathbf{x}_i)}{q(\mathbf{x}_i)}$, samples $\mathbf{x}_i \sim q(\mathbf{x})$

$$\mathrm{E}[\mathbf{f}(\mathbf{x})] = \sum_{i=1}^{N} \frac{\tilde{w}_i}{\sum_{j=1}^{N} \tilde{w}_j} \mathbf{f}(\mathbf{x}_i). \tag{21}$$

# Approximating a Probability Density Function with Importance Sampling

- The importance sampling approximation to $p(\mathbf{x})$ given the unnormalized distribution $\tilde{p}(\mathbf{x})$, and a proposal distribution $q(\mathbf{x})$, is thus given by

$$\boxed{p(\mathbf{x}) \approx \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i), \quad \mathbf{x}_i \sim q(\mathbf{x}),} \tag{22}$$

# Approximating a Probability Density Function with Importance Sampling

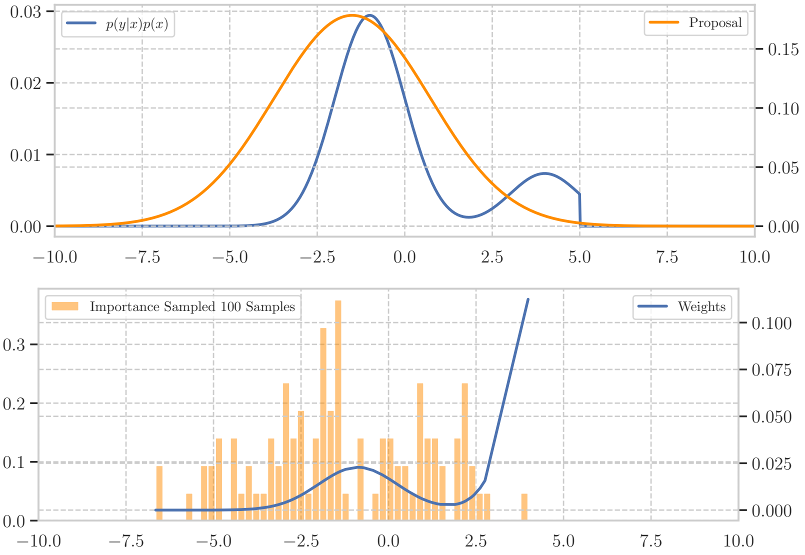
- The importance sampling approximation to $p(\mathbf{x})$ given the unnormalized distribution $\tilde{p}(\mathbf{x})$, and a proposal distribution $q(\mathbf{x})$, is thus given by

$$\boxed{p(\mathbf{x}) \approx \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i), \quad \mathbf{x}_i \sim q(\mathbf{x}),} \tag{22}$$

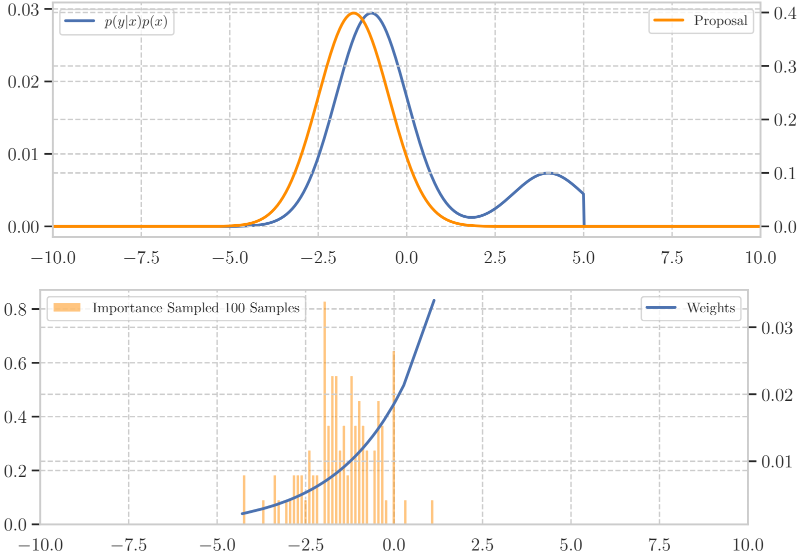
with the weights $w_i$ given by

$$\boxed{w_i = \frac{\tilde{p}(\mathbf{x}_i)/q(\mathbf{x}_i)}{\sum_{j=1}^{N} \tilde{p}(\mathbf{x}_j)/q(\mathbf{x}_j)}.} \tag{23}$$

# Importance Sampling Illustration

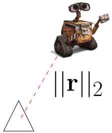


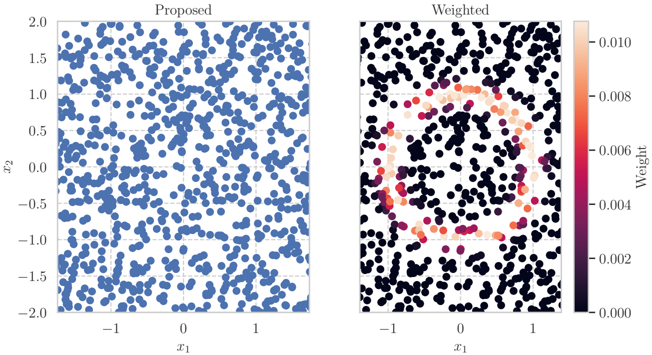
# Importance Sampling Illustration



# Importance Sampling Illustration

- Single measurement.
- Uniform proposal distribution.



# General Sampling Methods: Markov Chain Monte Carlo

- Given Markov chain with transition probability

$$T\left(\mathbf{x}^{(m+1)}, \mathbf{x}^{(m)}\right) = p\left(\mathbf{x}^{(m+1)} | \mathbf{x}^{(m)}\right) \tag{24}$$

# General Sampling Methods: Markov Chain Monte Carlo

- Given Markov chain with transition probability

$$T\left(\mathbf{x}^{(m+1)}, \mathbf{x}^{(m)}\right) = p\left(\mathbf{x}^{(m+1)} | \mathbf{x}^{(m)}\right) \tag{24}$$

- A distribution $p^*(\mathbf{x})$ *invariant* w.r.t. given Markov chain if each step leaves it unchanged

$$p(\mathbf{x}^{(m)}) = p^*(\mathbf{x}^{(m)}) \Rightarrow p(\mathbf{x}^{(m+1)}) = p^*(\mathbf{x}^{(m+1)}). \tag{25}$$

# General Sampling Methods: Markov Chain Monte Carlo

- Given Markov chain with transition probability

$$T\left(\mathbf{x}^{(m+1)}, \mathbf{x}^{(m)}\right) = p\left(\mathbf{x}^{(m+1)} | \mathbf{x}^{(m)}\right) \tag{24}$$

- A distribution $p^*(\mathbf{x})$ *invariant* w.r.t. given Markov chain if each step leaves it unchanged

$$p(\mathbf{x}^{(m)}) = p^*(\mathbf{x}^{(m)}) \Rightarrow p(\mathbf{x}^{(m+1)}) = p^*(\mathbf{x}^{(m+1)}). \tag{25}$$

  Formally,

$$p^*(\mathbf{x}) = \int T(\mathbf{x}, \mathbf{x}')p^*(\mathbf{x}')\mathrm{d}\mathbf{x}'. \tag{26}$$

# General Sampling Methods: Markov Chain Monte Carlo

- Given Markov chain with transition probability

$$T\left(\mathbf{x}^{(m+1)}, \mathbf{x}^{(m)}\right) = p\left(\mathbf{x}^{(m+1)} | \mathbf{x}^{(m)}\right) \tag{24}$$

- A distribution $p^*(\mathbf{x})$ *invariant* w.r.t. given Markov chain if each step leaves it unchanged

$$p(\mathbf{x}^{(m)}) = p^*(\mathbf{x}^{(m)}) \Rightarrow p(\mathbf{x}^{(m+1)}) = p^*(\mathbf{x}^{(m+1)}). \tag{25}$$

  Formally,

$$p^*(\mathbf{x}) = \int T(\mathbf{x}, \mathbf{x}')p^*(\mathbf{x}')\mathrm{d}\mathbf{x}'. \tag{26}$$

- Markov chain is called *ergodic* if $p(\mathbf{x}^{(m)})$ converges to $p^*(\mathbf{x})$.

# General Sampling Methods: Markov Chain Monte Carlo

- Given Markov chain with transition probability

$$T\left(\mathbf{x}^{(m+1)}, \mathbf{x}^{(m)}\right) = p\left(\mathbf{x}^{(m+1)} | \mathbf{x}^{(m)}\right) \tag{24}$$

- A distribution $p^*(\mathbf{x})$ *invariant* w.r.t. given Markov chain if each step leaves it unchanged

$$p(\mathbf{x}^{(m)}) = p^*(\mathbf{x}^{(m)}) \Rightarrow p(\mathbf{x}^{(m+1)}) = p^*(\mathbf{x}^{(m+1)}). \tag{25}$$

Formally,

$$p^*(\mathbf{x}) = \int T(\mathbf{x}, \mathbf{x}')p^*(\mathbf{x}')\mathrm{d}\mathbf{x}'. \tag{26}$$

- Markov chain is called *ergodic* if $p(\mathbf{x}^{(m)})$ converges to $p^*(\mathbf{x})$.
- Different MCMC algorithms design different Markov chains.

# Metropolis Algorithm

- Start with transition proposal distribution $q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)}))$

# Metropolis Algorithm

- Start with transition proposal distribution $q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)}))$
- Symmetric

$$q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)})) = q(\mathbf{x}^{(k)}|(\mathbf{x}^{(k+1)})) \tag{27}$$

Example: Gaussian around $\mathbf{x}^{(k)}$.

# Metropolis Algorithm

- Start with transition proposal distribution $q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)}))$
- Symmetric

$$q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)})) = q(\mathbf{x}^{(k)}|(\mathbf{x}^{(k+1)})) \tag{27}$$

  Example: Gaussian around $\mathbf{x}^{(k)}$.
- At each iteration, given current sample $\mathbf{x}^{(k)}$

# Metropolis Algorithm

- Start with transition proposal distribution $q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)}))$
- Symmetric

$$q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)})) = q(\mathbf{x}^{(k)}|(\mathbf{x}^{(k+1)})) \tag{27}$$

  Example: Gaussian around $\mathbf{x}^{(k)}$.

- At each iteration, given current sample $\mathbf{x}^{(k)}$
  1. Generate candidate sample

$$\mathbf{x}_{\text{cand}}^{(k+1)} \sim q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)})). \tag{28}$$

# Metropolis Algorithm

- Start with transition proposal distribution $q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)}))$
- Symmetric

$$q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)})) = q(\mathbf{x}^{(k)}|(\mathbf{x}^{(k+1)})) \tag{27}$$

  Example: Gaussian around $\mathbf{x}^{(k)}$.
- At each iteration, given current sample $\mathbf{x}^{(k)}$
  1. Generate candidate sample

$$\mathbf{x}_{\text{cand}}^{(k+1)} \sim q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)})). \tag{28}$$

  2. Accept with probability

$$P(\text{accept}|\mathbf{x}_{\text{cand}}^{(k+1)}, \mathbf{x}^{(k)}) = \min\left(1, \frac{\tilde{p}(\mathbf{x}_{\text{cand}}^{k+1})}{\tilde{p}(\mathbf{x}^{k})}\right). \tag{29}$$

# Metropolis Algorithm

- Start with transition proposal distribution $q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)}))$
- Symmetric

$$q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)})) = q(\mathbf{x}^{(k)}|(\mathbf{x}^{(k+1)})) \tag{27}$$

  Example: Gaussian around $\mathbf{x}^{(k)}$.
- At each iteration, given current sample $\mathbf{x}^{(k)}$
  1. Generate candidate sample

$$\mathbf{x}_{\text{cand}}^{(k+1)} \sim q(\mathbf{x}^{(k+1)}|(\mathbf{x}^{(k)})). \tag{28}$$

  2. Accept with probability

$$P(\text{accept}|\mathbf{x}_{\text{cand}}^{(k+1)}, \mathbf{x}^{(k)}) = \min\left(1, \frac{\tilde{p}(\mathbf{x}_{\text{cand}}^{k+1})}{\tilde{p}(\mathbf{x}^{k})}\right). \tag{29}$$

- Animation!

# Sequential Monte Carlo

- Recall Bayes filter,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1}. \tag{30}$$

# Sequential Monte Carlo

- Recall Bayes filter,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \frac{1}{\eta}p(\mathbf{y}_k|\mathbf{x}_k)\int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\mathrm{d}\mathbf{x}_{k-1}. \tag{30}$$

- The distribution of the state at timestep $k-1$, $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$, is represented by set of particles,

$$p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) = \sum_{i=1}^{N} w_{i,k-1}\delta(\mathbf{x}-\mathbf{x}_{i,k-1}). \tag{31}$$

# Sequential Monte Carlo

- Recall Bayes filter,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \mathrm{d}\mathbf{x}_{k-1}. \tag{30}$$

- The distribution of the state at timestep $k-1$, $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$, is represented by set of particles,

$$p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) = \sum_{i=1}^{N} w_{i,k-1} \delta(\mathbf{x} - \mathbf{x}_{i,k-1}). \tag{31}$$

- Thus, (30) becomes

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) \sum_{i=1}^{N} w_{i,k-1} \delta(\mathbf{x} - \mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_{k-1}. \tag{32}$$

# Sequential Monte Carlo

- The sum may be taken outside of the integral such that

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) \delta(\mathbf{x} - \mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_{k-1} \tag{33}$$

$$= \boxed{\frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}).} \tag{34}$$

# Sequential Monte Carlo

- The sum may be taken outside of the integral such that

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta}p(\mathbf{y}_k|\mathbf{x}_k)\sum_{i=1}^{N} w_{i,k-1}\int p(\mathbf{x}_k|\mathbf{x}_{k-1})\delta(\mathbf{x}-\mathbf{x}_{i,k-1})\mathrm{d}\mathbf{x}_{k-1} \tag{33}$$

$$= \boxed{\frac{1}{\eta}p(\mathbf{y}_k|\mathbf{x}_k)\sum_{i=1}^{N} w_{i,k-1}p(\mathbf{x}_k|\mathbf{x}_{i,k-1}).} \tag{34}$$

- The marginal posterior (34) is only a function of $\mathbf{x}_k$.

# Sequential Monte Carlo

- The sum may be taken outside of the integral such that

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta}p(\mathbf{y}_k|\mathbf{x}_k)\sum_{i=1}^{N} w_{i,k-1}\int p(\mathbf{x}_k|\mathbf{x}_{k-1})\delta(\mathbf{x}-\mathbf{x}_{i,k-1})\mathrm{d}\mathbf{x}_{k-1} \tag{33}$$

$$= \boxed{\frac{1}{\eta}p(\mathbf{y}_k|\mathbf{x}_k)\sum_{i=1}^{N} w_{i,k-1}p(\mathbf{x}_k|\mathbf{x}_{i,k-1}).} \tag{34}$$

- The marginal posterior (34) is only a function of $\mathbf{x}_k$.

*Can use any Monte Carlo sampling method!*

# Integration Nuance

- Filtering distribution given by

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{35}$$

# Integration Nuance

- Filtering distribution given by

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{35}$$

- Expectation of an arbitrary $\mathbf{f}(\mathbf{x})$ is

$$\int \mathbf{f}(\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k}) \mathrm{d}\mathbf{x}_k = \frac{1}{\eta} \int \mathbf{f}(\mathbf{x}_k) p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_k \tag{36}$$

(37)

# Integration Nuance

- Filtering distribution given by

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{35}$$

- Expectation of an arbitrary $\mathbf{f}(\mathbf{x})$ is

$$\int \mathbf{f}(\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k}) \mathrm{d}\mathbf{x}_k = \frac{1}{\eta} \int \mathbf{f}(\mathbf{x}_k) p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_k \tag{36}$$

$$= \frac{1}{\eta} \sum_{i=1}^{N} w_{i,k-1} \int \mathbf{f}(\mathbf{x}_k) p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_k. \tag{37}$$

# Integration Nuance

- Filtering distribution given by

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{35}$$

- Expectation of an arbitrary $\mathbf{f}(\mathbf{x})$ is

$$\int \mathbf{f}(\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k}) \mathrm{d}\mathbf{x}_k = \frac{1}{\eta} \int \mathbf{f}(\mathbf{x}_k) p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_k \tag{36}$$

$$= \frac{1}{\eta} \sum_{i=1}^{N} w_{i,k-1} \int \mathbf{f}(\mathbf{x}_k) p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_k. \tag{37}$$

- By using (35) in an arbitrary Monte Carlo solver, we develop a set of particles that approximate the integral in (36).

# Integration Nuance

- Filering distribution given by

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{\eta} p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{35}$$

- Expectation of an arbitrary $\mathbf{f}(\mathbf{x})$ is

$$\int \mathbf{f}(\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k}) \mathrm{d}\mathbf{x}_k = \frac{1}{\eta} \int \mathbf{f}(\mathbf{x}_k) p(\mathbf{y}_k|\mathbf{x}_k) \sum_{i=1}^{N} w_{i,k-1} p(\mathbf{x}_k|\mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_k \tag{36}$$

$$= \frac{1}{\eta} \sum_{i=1}^{N} w_{i,k-1} \int \mathbf{f}(\mathbf{x}_k) p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{x}_{i,k-1}) \mathrm{d}\mathbf{x}_k. \tag{37}$$

- By using (35) in an arbitrary Monte Carlo solver, we develop a set of particles that approximate the integral in (36).
- By using the $N$ integrands in (37) in an arbitrary Monte Carlo solver, we develop a set of particles that approximate the sum of the integrals in (37).

# Sequential Monte Carlo with Importance Sampling

- Using importance sampling on (37) with proposal distribution $q(\mathbf{x}_k|\mathbf{x}_{i,k-1})$ gives

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_k|\mathbf{x}_{i,k-1})} \delta(\mathbf{x} - \mathbf{x}_{i,k}) \quad \mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{38}$$

# Sequential Monte Carlo with Importance Sampling

- Using importance sampling on (37) with proposal distribution $q(\mathbf{x}_k|\mathbf{x}_{i,k-1})$ gives

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_k|\mathbf{x}_{i,k-1})} \delta(\mathbf{x} - \mathbf{x}_{i,k}) \quad \mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{38}$$

- The sequential update is then

$$\boxed{\mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}),} \tag{39}$$

$$\boxed{w_{i,k} \leftarrow w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_{i,k})p(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})}.} \tag{40}$$

# Sequential Monte Carlo with Importance Sampling

- Using importance sampling on (37) with proposal distribution $q(\mathbf{x}_k|\mathbf{x}_{i,k-1})$ gives

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_k|\mathbf{x}_{i,k-1})} \delta(\mathbf{x} - \mathbf{x}_{i,k}) \quad \mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{38}$$

- The sequential update is then

$$\boxed{\mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}),} \tag{39}$$

$$\boxed{w_{i,k} \leftarrow w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_{i,k})p(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})}.} \tag{40}$$

- This is called *Sequential Importance Sampling*.

# Sequential Monte Carlo with Importance Sampling

- Using importance sampling on (37) with proposal distribution $q(\mathbf{x}_k|\mathbf{x}_{i,k-1})$ gives

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_k|\mathbf{x}_{i,k-1})} \delta(\mathbf{x} - \mathbf{x}_{i,k}) \quad \mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{38}$$

- The sequential update is then

$$\boxed{\mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}),} \tag{39}$$

$$\boxed{w_{i,k} \leftarrow w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_{i,k})p(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})}.} \tag{40}$$

- This is called *Sequential Importance Sampling*.
- Setting $q(\mathbf{x}_k|\mathbf{x}_{i,k-1}) = p(\mathbf{x}_k|\mathbf{x}_{i,k-1})$ is the *bootstrap particle filter*.

# Sequential Monte Carlo with Importance Sampling

- Using importance sampling on (37) with proposal distribution $q(\mathbf{x}_k|\mathbf{x}_{i,k-1})$ gives

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \approx \sum_{i=1}^{N} w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_k|\mathbf{x}_{i,k-1})} \delta(\mathbf{x} - \mathbf{x}_{i,k}) \quad \mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}). \tag{38}$$

- The sequential update is then

$$\boxed{\mathbf{x}_{i,k} \sim q(\mathbf{x}_k|\mathbf{x}_{i,k-1}),} \tag{39}$$

$$\boxed{w_{i,k} \leftarrow w_{i,k-1} \frac{p(\mathbf{y}_k|\mathbf{x}_{i,k})p(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})}{q(\mathbf{x}_{i,k}|\mathbf{x}_{i,k-1})}.} \tag{40}$$

- This is called *Sequential Importance Sampling*.
- Setting $q(\mathbf{x}_k|\mathbf{x}_{i,k-1}) = p(\mathbf{x}_k|\mathbf{x}_{i,k-1})$ is the *bootstrap particle filter*.
- Running an MCMC method on each integral of (38), after resampling, corresponds to the *resample-move* algorithm. Only the last sample in the chain is kept.

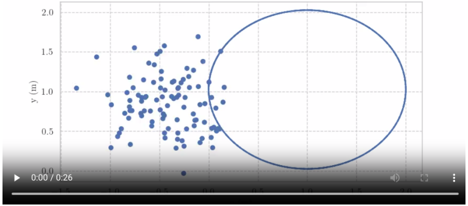
# Demo - A Simple Example

- Vector state $\mathbf{x}_k \in \mathbb{R}^2$.
- Single integrator process model,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{u}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k). \tag{41}$$

- Range measurement to anchor in the center of the scene,

$$\mathbf{y}_k = \|\mathbf{x}_k\|_2^2 + w_k, \quad w_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{42}$$



# Resampling

- The problem of all but a few weights going to zero is called the *sample degeneracy problem*.

# Resampling

- The problem of all but a few weights going to zero is called the *sample degeneracy problem*.
- Addressed by *resampling* in more probable regions. Given

$$p(\mathbf{x}) \approx \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i), \tag{43}$$

draw a new set $\mathbf{x}_j$ from the discrete distribution

$$P(\mathbf{x}_j) = w_j, \quad \mathbf{x}_j \in \{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots \mathbf{x}_N\}. \tag{44}$$

# Resampling

- The problem of all but a few weights going to zero is called the *sample degeneracy problem*.
- Addressed by *resampling* in more probable regions. Given

$$p(\mathbf{x}) \approx \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i), \tag{43}$$

draw a new set $\mathbf{x}_j$ from the discrete distribution

$$P(\mathbf{x}_j) = w_j, \quad \mathbf{x}_j \in \{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots \mathbf{x}_N\}. \tag{44}$$

- Concentrates particles in more likely regions.

# Resampling

- The problem of all but a few weights going to zero is called the *sample degeneracy problem*.
- Addressed by *resampling* in more probable regions. Given

$$p(\mathbf{x}) \approx \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i), \tag{43}$$

draw a new set $\mathbf{x}_j$ from the discrete distribution

$$P(\mathbf{x}_j) = w_j, \quad \mathbf{x}_j \in \{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots \mathbf{x}_N\}. \tag{44}$$

- Concentrates particles in more likely regions.
- Can cause *sample impoverishment* where particles lose diversity.

# Resampling

- The problem of all but a few weights going to zero is called the *sample degeneracy problem*.
- Addressed by *resampling* in more probable regions. Given

$$p(\mathbf{x}) \approx \sum_{i=1}^{N} w_i \delta(\mathbf{x} - \mathbf{x}_i), \tag{43}$$

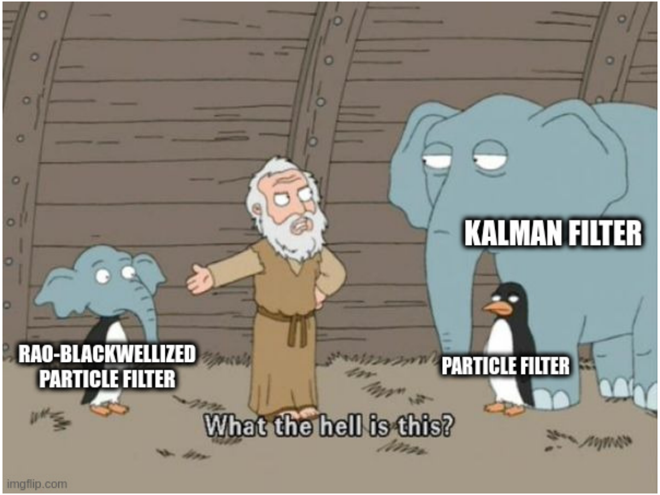
  draw a new set $\mathbf{x}_j$ from the discrete distribution

$$P(\mathbf{x}_j) = w_j, \quad \mathbf{x}_j \in \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots \mathbf{x}_N\}. \tag{44}$$

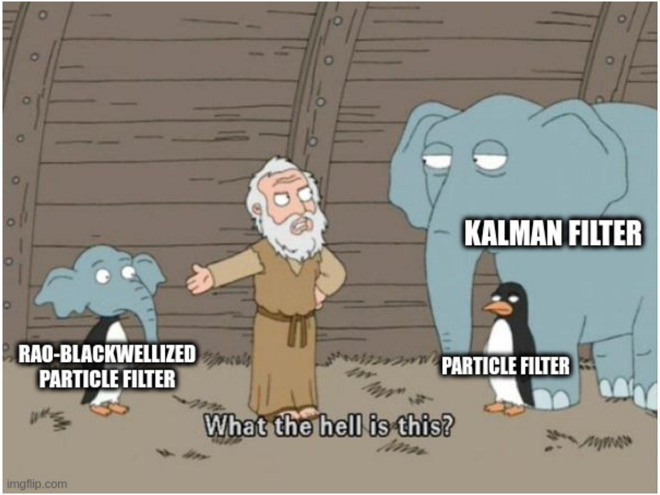
- Concentrates particles in more likely regions.
- Can cause *sample impoverishment* where particles lose diversity.
- Adaptive resampling - resample only when needed. For exaple, use effective number of particles as a threshold,

$$n_{\text{eff}} \approx \frac{1}{\sum_{i=1}^{N} {w_k^{(i)}}^2}. \tag{45}$$

# Rao-Blackwellization

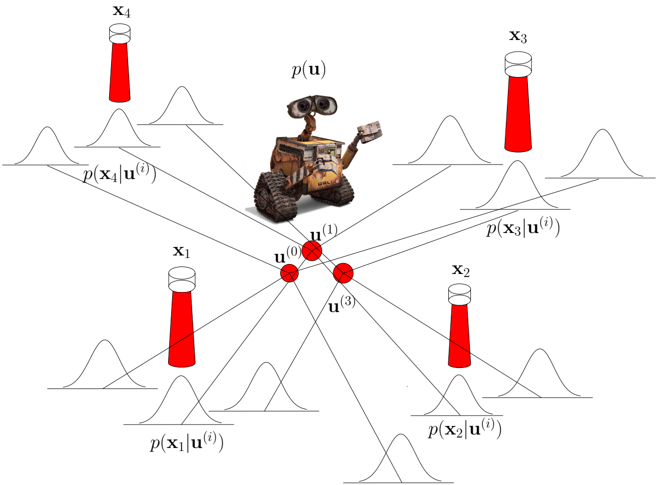


# Rao-Blackwellization




Partition state into non-Gaussian part $\mathbf{u}$ and *conditionally Gaussian* part $\mathbf{x}$.

# Rao-Blackwellization




Partition state into non-Gaussian part $\mathbf{u}$ and *conditionally Gaussian* part $\mathbf{x}$.

# Rao-Blackwellization

- Models of form

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1})), \mathbf{Q}_{k-1}(\mathbf{u}_{k-1})), \tag{46}$$

$$p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{g}(\mathbf{x}_k, \mathbf{u}_k), \mathbf{R}_k(\mathbf{u}_k)), \tag{47}$$

$$p(\mathbf{u}_k | \mathbf{u}_{k-1}) \sim \text{Any distribution.} \tag{48}$$

where $\mathbf{u}_k$ is *non-Gaussian part of the state.*

# Rao-Blackwellization

- Models of form

$$
\begin{aligned}
p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{u}_{k-1}) &= \mathcal{N}(\mathbf{x}_k;\mathbf{f}(\mathbf{x}_{k-1},\mathbf{u}_{k-1})),\mathbf{Q}_{k-1}(\mathbf{u}_{k-1})), && (46)\\
p(\mathbf{y}_k|\mathbf{x}_k,\mathbf{u}_k) &= \mathcal{N}(\mathbf{y}_k;\mathbf{g}(\mathbf{x}_k,\mathbf{u}_k),\mathbf{R}_k(\mathbf{u}_k)), && (47)\\
p(\mathbf{u}_k|\mathbf{u}_{k-1}) &\sim \text{Any distribution.} && (48)
\end{aligned}
$$

  where $\mathbf{u}_k$ is *non-Gaussian part of the state.*

- State belief

$$
p(\mathbf{x}_k,\mathbf{u}_k|\mathbf{y}_{1:k}) = \sum_{i=1}^{N} w_k^{(i)} \delta\left(\mathbf{u}_k - \mathbf{u}_k^{(i)}\right) \mathcal{N}\left(\mathbf{x}_k;\hat{\mathbf{x}}_k^{(i)},\hat{\mathbf{P}}_k^{(i)}\right). \tag{49}
$$

# Rao-Blackwellization

- Models of form

$$p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{u}_{k-1}) = \mathcal{N}(\mathbf{x}_k;\mathbf{f}(\mathbf{x}_{k-1},\mathbf{u}_{k-1})),\mathbf{Q}_{k-1}(\mathbf{u}_{k-1})), \tag{46}$$

$$p(\mathbf{y}_k|\mathbf{x}_k,\mathbf{u}_k) = \mathcal{N}(\mathbf{y}_k;\mathbf{g}(\mathbf{x}_k,\mathbf{u}_k),\mathbf{R}_k(\mathbf{u}_k)), \tag{47}$$

$$p(\mathbf{u}_k|\mathbf{u}_{k-1}) \sim \text{Any distribution.} \tag{48}$$

  where $\mathbf{u}_k$ is *non-Gaussian part of the state.*

- State belief

$$p(\mathbf{x}_k,\mathbf{u}_k|\mathbf{y}_{1:k}) = \sum_{i=1}^{N} w_k^{(i)}\delta\left(\mathbf{u}_k-\mathbf{u}_k^{(i)}\right)\mathcal{N}\left(\mathbf{x}_k;\hat{\mathbf{x}}_k^{(i)},\hat{\mathbf{P}}_k^{(i)}\right). \tag{49}$$

- State part $\mathbf{x}$ is conditionally Gaussian given $\mathbf{u}$,

$$p(\mathbf{x}_k|\mathbf{u}_k,\mathbf{y}_{1:k}) = \sum_{j=1}^{N} I(\mathbf{u}_j=\mathbf{u}_k)\mathcal{N}(\mathbf{x}_k|\mathbf{x}_k^{(j)},\mathbf{P}_k^{(j)}). \tag{50}$$

  where $I(*)$ is the indicator function that is equal to one if the input condition is fulfilled, and zero if not.

# Rao-Blackwellization

- The Bayes filter takes the form,

$$
\begin{aligned}
p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) &= \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{u}_k | \mathbf{y}_{k-1:k}) && (51) \\
&= \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{y}_k | \mathbf{u}_k, \mathbf{y}_{k-1}) p(\mathbf{u}_k | \mathbf{y}_{k-1}) && (52)
\end{aligned}
$$

# Rao-Blackwellization

- The Bayes filter takes the form,

$$p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) = \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{u}_k | \mathbf{y}_{k-1:k}) \tag{51}$$

$$= \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{y}_k | \mathbf{u}_k, \mathbf{y}_{k-1}) p(\mathbf{u}_k | \mathbf{y}_{k-1}) \tag{52}$$

- The predicted distribution of $\mathbf{u}$ is given by the Chapman-Kolmogorov equation,

$$p(\mathbf{u}_k | \mathbf{y}_{k-1}) = \int p(\mathbf{u}_k | \mathbf{u}_{k-1}) p(\mathbf{u}_{k-1} | \mathbf{y}_{k-1}) \mathrm{d}\mathbf{u}_{k-1} \tag{53}$$

$$= \int p(\mathbf{u}_k | \mathbf{u}_{k-1}) \sum_{i=1}^{N} w_{k-1}^{i} \delta(\mathbf{u}_{k-1} - \mathbf{u}_{k-1}^{(i)}) \mathrm{d}\mathbf{u}_{k-1} \tag{54}$$

$$= \sum_{i=1}^{N} w_{k-1}^{i} p(\mathbf{u}_k | \mathbf{u}_{k-1}^{(i)}). \tag{55}$$

# Rao-Blackwellization - Prediction step for non-Gaussian state

- Importance sampling. Define $q(\mathbf{u}_k|\mathbf{u}^i_{k-1})$ and sample a $\mathbf{u}^{(i)}_k$ with corresponding predicted weight

$$\check{w}^{(i)}_k = w^{(i-1)}_k \frac{p(\mathbf{u}^{(i)}_k|\mathbf{u}^{(i)}_{k-1})}{q(\mathbf{u}^{(i-1)}_k|\mathbf{u}^{(i)}_{k-1})}. \tag{56}$$

# Rao-Blackwellization - Prediction step for non-Gaussian state

- Importance sampling. Define $q(\mathbf{u}_k|\mathbf{u}_{k-1}^i)$ and sample a $\mathbf{u}_k^{(i)}$ with corresponding predicted weight

$$\check{w}_k^{(i)} = w_k^{(i-1)} \frac{p(\mathbf{u}_k^{(i)}|\mathbf{u}_{k-1}^{(i)})}{q(\mathbf{u}_k^{(i-1)}|\mathbf{u}_{k-1}^{(i)})}. \tag{56}$$

- The predicted distribution on $\mathbf{u}_k$ is thus

$$p(\mathbf{u}_k|\mathbf{y}_{k-1}) = \sum_{i=1}^{N} \check{w}_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}). \tag{57}$$

# Rao-Blackwellization - Correction step for non-Gaussian state

- The correction consists of updating the weights using the marginal likelihood $p(\mathbf{y}_k|\mathbf{u}_k, \mathbf{y}_{k-1})$, which is obtained by

$$p(\mathbf{y}_k|\mathbf{u}_k, \mathbf{y}_{k-1}) = \int p(\mathbf{y}_k, \mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1})\mathrm{d}\mathbf{x}_k \tag{58}$$

$$= \int p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{u}_k)p(\mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1})\mathrm{d}\mathbf{x}_k, \tag{59}$$

  where $p(\mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1})$ is given by

$$p(\mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1}) = \sum_{i=1}^{N} I(\mathbf{u}_i = \mathbf{u}_k)\mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}). \tag{60}$$

# Rao-Blackwellization - Correction step for non-Gaussian state

- The correction consists of updating the weights using the marginal likelihood $p(\mathbf{y}_k|\mathbf{u}_k, \mathbf{y}_{k-1})$, which is obtained by

$$p(\mathbf{y}_k|\mathbf{u}_k, \mathbf{y}_{k-1}) = \int p(\mathbf{y}_k, \mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1}) \mathrm{d}\mathbf{x}_k \tag{58}$$

$$= \int p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1}) \mathrm{d}\mathbf{x}_k, \tag{59}$$

where $p(\mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1})$ is given by

$$p(\mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1}) = \sum_{i=1}^{N} I(\mathbf{u}_i = \mathbf{u}_k) \mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}). \tag{60}$$

- The likelihood (59) becomes

$$p(\mathbf{y}_k|\mathbf{u}_k, \mathbf{y}_{k-1}) = \sum_{i=1}^{N} w_{k-1}^{(i)} I(\mathbf{u}_i = \mathbf{u}_k) \int p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{u}_k) \mathcal{N}\left(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}\right) \mathrm{d}\mathbf{x}_k, \tag{61}$$

where each integral obtained as the marginal measurement mean and covariance of the Gaussian filter update.

# Rao-Blackwellization - Updated Belief on non-Gaussian State

The likelihood (61) is thus combined with (57) to give

$$p(\mathbf{u}_k|\mathbf{y}_{k-1:k}) = \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}), \tag{62}$$

with

$$\mathbf{u}_k^{(i)} \sim q(\mathbf{u}_k^{(i-1)}|\mathbf{u}_{k-1}^{(i)}) \tag{63}$$

$$w_k^{(i)} = \underbrace{w_k^{(i-1)} \frac{p(\mathbf{u}_k^{(i)}|\mathbf{u}_{k-1}^{(i)})}{q(\mathbf{u}_k^{(i-1)}|\mathbf{u}_{k-1}^{(i)})}}_{\check{w}_k^{(i)}} \int p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{u}_k) \mathcal{N}\left(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}\right) \mathrm{d}\mathbf{x}_k. \tag{64}$$

# Rao-Blackwellization - Updating Conditionally Gaussian State

- The Bayes filter (51) becomes

$$p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) = \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{u}_k | \mathbf{y}_{k-1:k}) \tag{65}$$

(66)

(67)

# Rao-Blackwellization - Updating Conditionally Gaussian State

- The Bayes filter (51) becomes

$$p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) = \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{u}_k | \mathbf{y}_{k-1:k}) \tag{65}$$

$$= \frac{1}{\eta} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) \tag{66}$$

(67)

# Rao-Blackwellization - Updating Conditionally Gaussian State

- The Bayes filter (51) becomes

$$
\begin{aligned}
p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) &= \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{u}_k | \mathbf{y}_{k-1:k}) && (65) \\
&= \frac{1}{\eta} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) && (66) \\
&= \frac{1}{\eta} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) \sum_{i=1}^{N} I(\mathbf{u}_i = \mathbf{u}_k) \mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}). && (67)
\end{aligned}
$$

# Rao-Blackwellization - Updating Conditionally Gaussian State

- The Bayes filter (51) becomes

$$p(\mathbf{x}_k, \mathbf{u}_k|\mathbf{y}_{k-1:k}) = \frac{1}{\eta}p(\mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1:k})p(\mathbf{u}_k|\mathbf{y}_{k-1:k}) \tag{65}$$

$$= \frac{1}{\eta}p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{u}_k)p(\mathbf{x}_k|\mathbf{u}_k, \mathbf{y}_{k-1}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) \tag{66}$$

$$= \frac{1}{\eta}p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{u}_k) \sum_{i=1}^{N} I(\mathbf{u}_i = \mathbf{u}_k)\mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}). \tag{67}$$

- Which simplifies to

$$p(\mathbf{x}_k, \mathbf{u}_k|\mathbf{y}_{k-1:k}) = \frac{1}{\eta} \sum_{i=1}^{N} p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{u}_k)\mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}) w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) \tag{68}$$

$$\tag{69}$$

# Rao-Blackwellization - Updating Conditionally Gaussian State

- The Bayes filter (51) becomes

$$
\begin{aligned}
p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) &= \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{u}_k | \mathbf{y}_{k-1:k}) && (65) \\
&= \frac{1}{\eta} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) && (66) \\
&= \frac{1}{\eta} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) \sum_{i=1}^{N} I(\mathbf{u}_i = \mathbf{u}_k) \mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}). && (67)
\end{aligned}
$$

- Which simplifies to

$$
\begin{aligned}
p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) &= \frac{1}{\eta} \sum_{i=1}^{N} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) \mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}) w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) && (68) \\
&= \frac{1}{\eta} \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) \underbrace{p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) \mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)})}_{\text{Predict/Correct for each particle's } \mathbf{x}}, && (69)
\end{aligned}
$$

# Rao-Blackwellization - Updating Conditionally Gaussian State

- The Bayes filter (51) becomes

$$p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) = \frac{1}{\eta} p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1:k}) p(\mathbf{u}_k | \mathbf{y}_{k-1:k}) \tag{65}$$

$$= \frac{1}{\eta} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) p(\mathbf{x}_k | \mathbf{u}_k, \mathbf{y}_{k-1}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) \tag{66}$$

$$= \frac{1}{\eta} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) \sum_{i=1}^{N} I(\mathbf{u}_i = \mathbf{u}_k) \mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}) \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}). \tag{67}$$

- Which simplifies to

$$p(\mathbf{x}_k, \mathbf{u}_k | \mathbf{y}_{k-1:k}) = \frac{1}{\eta} \sum_{i=1}^{N} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) \mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)}) w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) \tag{68}$$

$$= \frac{1}{\eta} \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{u}_k - \mathbf{u}_k^{(i)}) \underbrace{p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{u}_k) \mathcal{N}(\mathbf{x}_k; \check{\mathbf{x}}_k^{(i)}, \check{\mathbf{P}}_k^{(i)})}_{\text{Predict/Correct for each particle's } \mathbf{x}}, \tag{69}$$

where the non-Gaussian state update for $w_k^{(i)}, \mathbf{u}_k^{(i)}$ is given by (64).

# References I


Cappe, Olivier, Simon J. Godsill, and Eric Moulines (2007). “An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo”. In: *Proceedings of the IEEE* 95.5, pp. 899–924.

Doucet, Arnaud, Nando De Freitas, Neil James Gordon, et al. (2001). *Sequential Monte Carlo methods in practice*. Vol. 1. 2. Springer.

Doucet, Arnaud, Adam M Johansen, et al. (2009). “A tutorial on particle filtering and smoothing: Fifteen years later”. In: *Handbook of nonlinear filtering* 12.656-704, p. 3.

Särkkä, Simo and Lennart Svensson (2023). *Bayesian filtering and smoothing*. Vol. 17. Cambridge university press.

Septier, François and Gareth W. Peters (2016). “Langevin and Hamiltonian Based Sequential MCMC for Efficient Bayesian Filtering in High-Dimensional Spaces”. In: *IEEE Journal of Selected Topics in Signal Processing* 10.2, pp. 312–327.