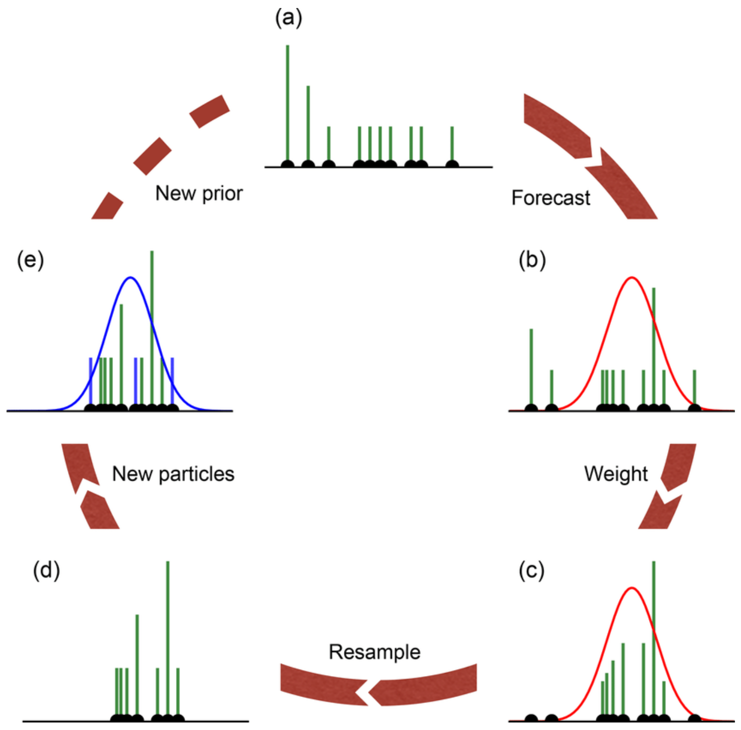
2.160 Identification, Estimation, and Learning

Part 2 Estimation

Lecture 12

Particle Filter

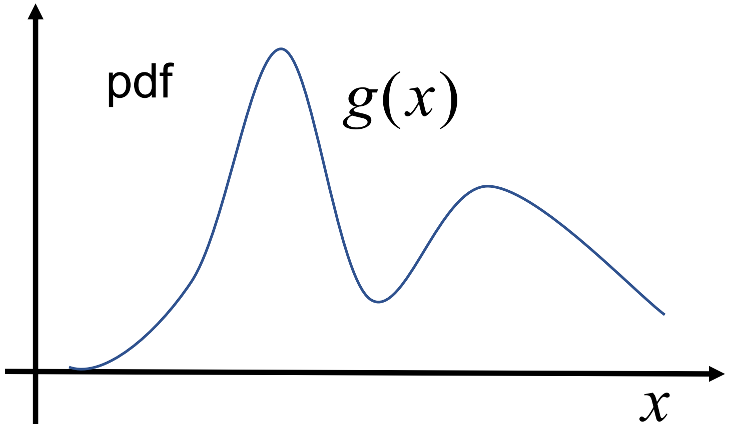


H. Harry Asada
Department of Mechanical Engineering
MIT

Recap

Bayes Filter

Belief

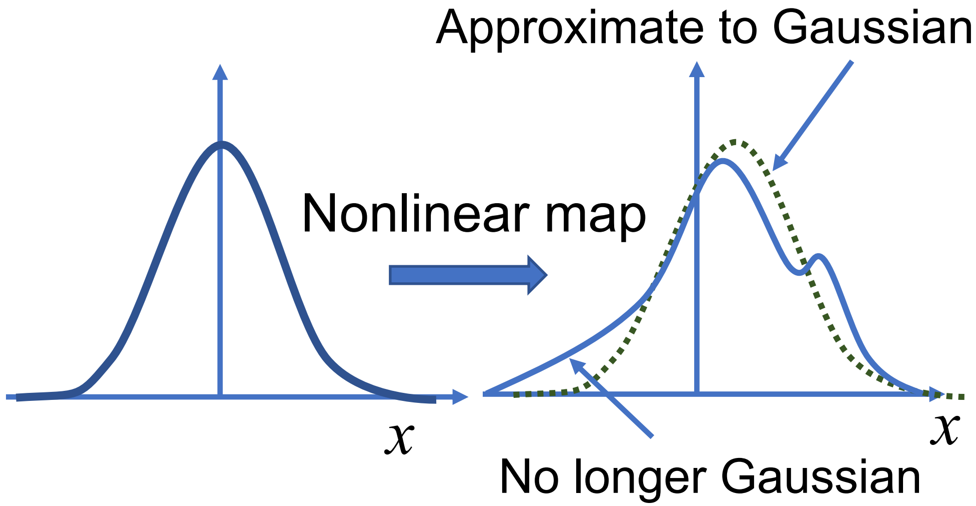


Non-Gaussian Distribution

A mean value does not represent the true distribution of the random variables.

Multi-Hypothesis Estimate

Limitation to UKF



Belief $g(x)$: the entire pdf distribution rather than a single value.

Bayes Filter predicts the pdf distribution of a random variable.

$$g_{t-1}(x_{t-1}) \quad \mapsto \quad g_{t|t-1}(x_t) \quad \mapsto \quad g_{t|t}(x_t)$$

Propagate　　Update

at time t-1　　a priori estimate.　a posteriori estimate

Recap

Bayes Filter

1. Initial Conditions: $g_0(x_0)$ set $t=1$;
2. Belief Propagation:

$$x_t = f(x_{t-1}, u_{t-1}) + w_{t-1}$$

Chapman-Kolmogorov Eq.

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} f_W(x_t - f(x_{t-1}, u_{t-1}))\, g_{t-1}(x_{t-1})\, dx_{t-1}$$

3. Assimilate y_t and update the a priori belief

Bayes Rule

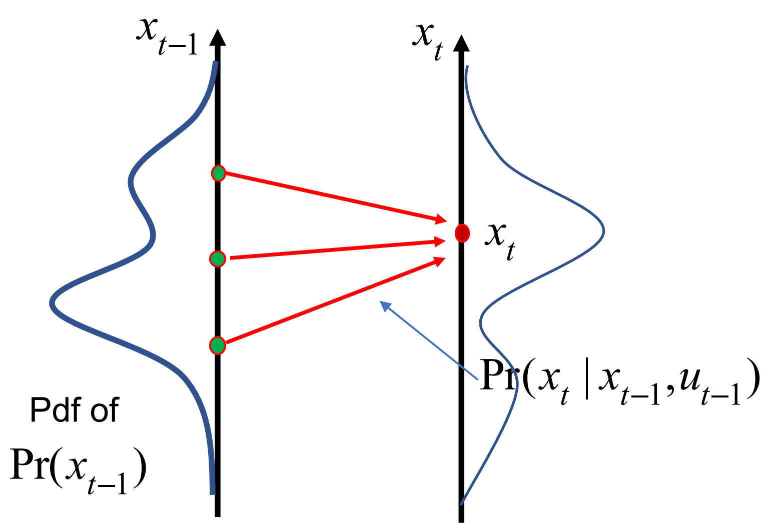
$$y_t = h(x_t, t) + v_t$$

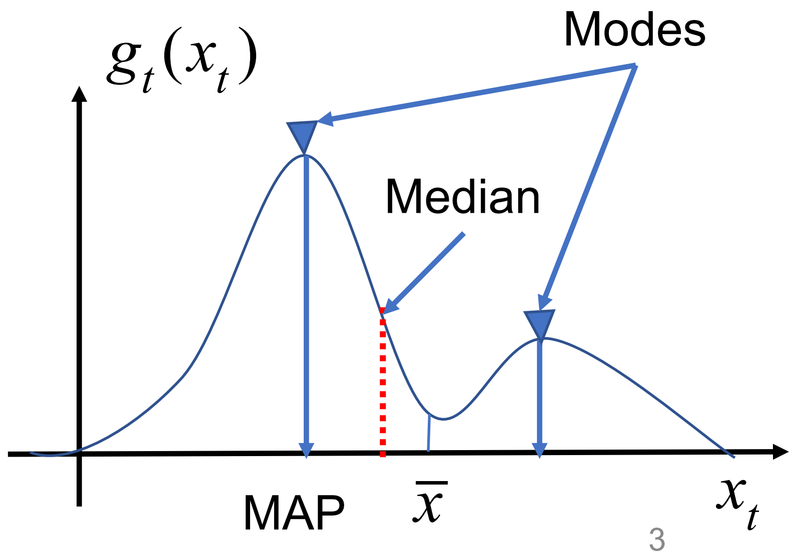
$$g_t(x_t) = \eta f_V(y_t - h(x_t, t)) g_{t|t-1}(x_t)$$

4. Return $g_t(x_t)$. Set $t = t+1$ and repeat.

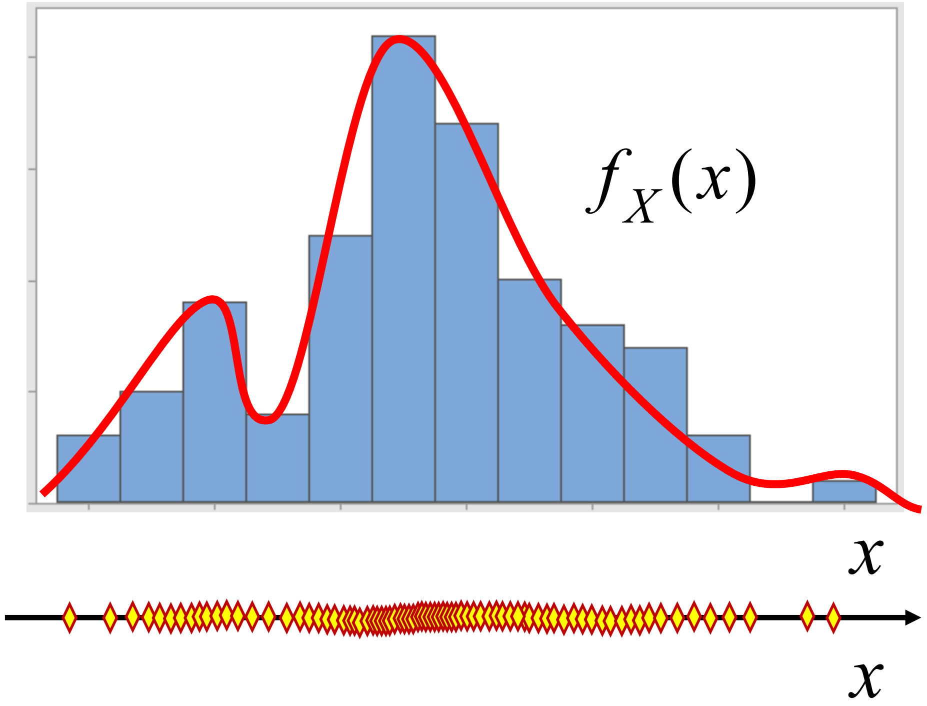
Computationally too expensive.
Not feasible for real-time applications.

State propagation





Particle: Non-Parametric Representation of Probability Density



- Parametric Distribution
  - Gaussian, Poisson, Gamma, Chi-square
- Non-Parametric: Arbitrary Distribution
  - Histogram
  - Particles
- Draw samples with probability density $f_X(x)$ and form a data set:

$$\tilde{X} = \{x^{(1)}, x^{(2)}, \cdots, x^{(M)}\}$$

- Particle $f_X(x)$ may be populated more densely at a region where $f_X(x)$ is large, reflecting the probability density of $f_X(x)$.

- The samples collectively represent the probability distribution.

Each drawn sample is called a "Particle"

Monte Carlo Approximation

- "Particles" facilitate the computation of expectation.
  Example, k-th order moment of random variable X.

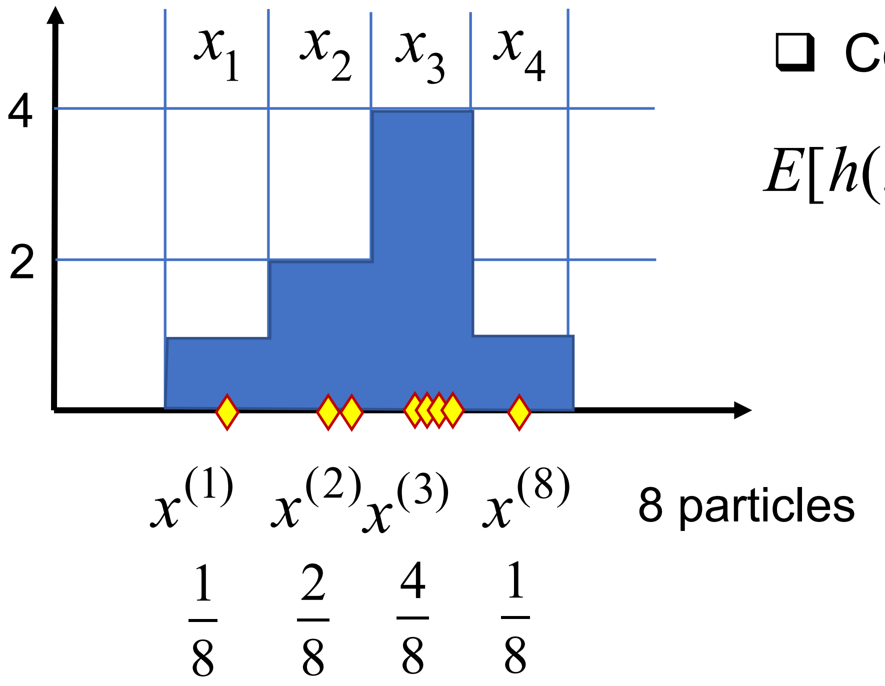
$$E[X^k] = \int_{-\infty}^{\infty} x^k f_X(x)\,dx \cong \frac{1}{M}\sum_{i=1}^{M}\left(x^{(i)}\right)^k$$

where $x^{(1)}, x^{(2)}, \cdots, x^{(M)}$ are particles drawn from $f_X(x)$.

- In general,

$$E\left[h(x)\right] \cong \frac{1}{M}\sum_{i=1}^{M} h(x^{(i)})$$

Example: Monte Carlo Approximation



Probability Mass Function

- The distribution of X is represented by histogram and particles.
- Compute the expectation of $h(x)$ from particles.

$$E[h(x)] = \sum_{j=1}^{4} h(x_j)\cdot p(x_j) \qquad \text{Definition of expectation}$$

$$= h(x_1)\frac{1}{8} + h(x_2)\frac{2}{8} + h(x_3)\frac{4}{8} + h(x_4)\frac{1}{8}$$

$$= \frac{1}{8}\left(h(x^{(1)}) + \underbrace{h(x^{(2)}) + h(x^{(3)})}_{2h(x_2)} + \underbrace{h(x^{(4)}) + \cdots}_{4h(x_3)} + h(x^{(8)}) \right)$$

$$= \frac{1}{8}\sum_{i=1}^{8} h(x^{(i)})$$

The probability is replaced by the multiple particles within the same interval.

How to Generate Particles: $\tilde{X}$

Objective: Generate M particles approximating pdf $f_X(x)$

The Cumulative Distribution Function (cdf) Method

1. Construct the cumulative distribution function

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi)\, d\xi$$

2. Draw a sample from a uniform distribution:

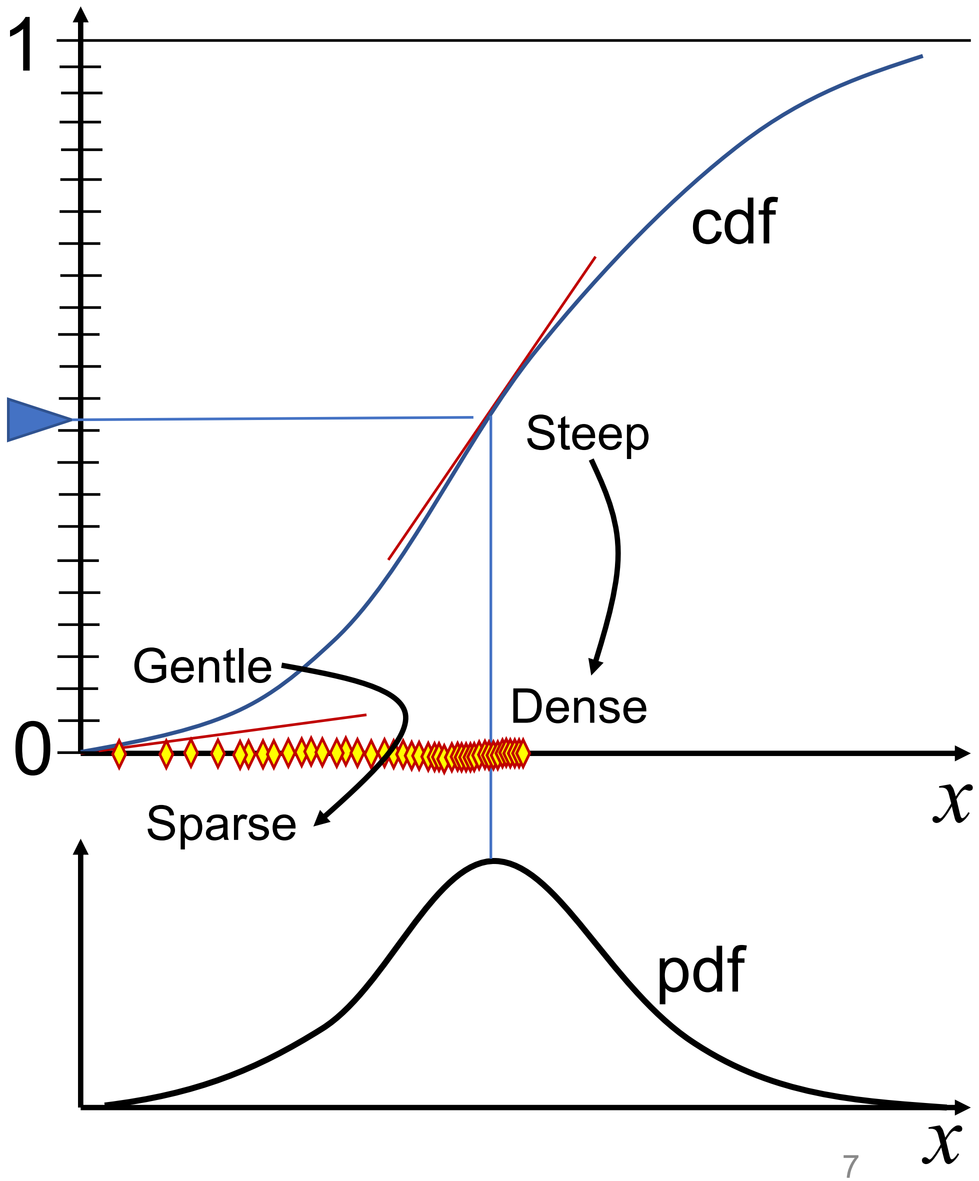
$$0 \le y^{(i)} \le 1$$

3. Convert the sample $y^{(i)}$ to $x^{(i)}$ by solving

$$F(x^{(i)}) = y^{(i)}$$

4. Repeat M times and form a data set

$$\tilde{X} = \left\{ \; x^{(1)} \quad x^{(2)} \quad \cdots \quad x^{(M)} \; \right\}$$



8.2 Implementing the Bayes Filter Using Particles

Step 1: Propagation

Instead of computing the Chapman-Kolmogorov equation, we generate M particles representing $g_{t|t-1}(x_t)$ from $g_{t-1}(x_{t-1})$

a posteriori belief

a priori belief

$$\tilde{X}_{t-1} = \left\{ \underbrace{x_{t-1}^{(1)} \quad x_{t-1}^{(2)} \quad \cdots \quad x_{t-1}^{(M)}} \right\} \Rightarrow \tilde{X}_{t|t-1} = \left\{ \underbrace{x_{t|t-1}^{(1)} \quad x_{t|t-1}^{(2)} \quad \cdots \quad x_{t|t-1}^{(M)}} \right\}$$

Recall

$$\int_{-\infty}^{\infty} \underline{\Pr(x_t \mid x_{t-1}, u_{t-1})} g_{t-1}(x_{t-1}) dx_{t-1} = g_{t|t-1}(x_t)$$

This resembles Monte Carlo Approximation:

But, this is not an algebraic function, but a pdf.

$$E\left[h(x)\right] = \int_{-\infty}^{\infty} h(x) f_X(x) dx \cong \frac{1}{M} \sum_{i=1}^{M} h(x^{(i)})$$

Suppose that many particles exist in

$$x_{t-1} \le x_{t-1}^{(i)} \le x_{t-1} + \Delta x_{t-1}$$

Draw $x_{t|t-1}^{(i)}$ from $p(x_t \mid \bar{x}_{t-1,j}, u_{t-1})$

As many times as the number of particles within

the interval of Δx_{t-1} .

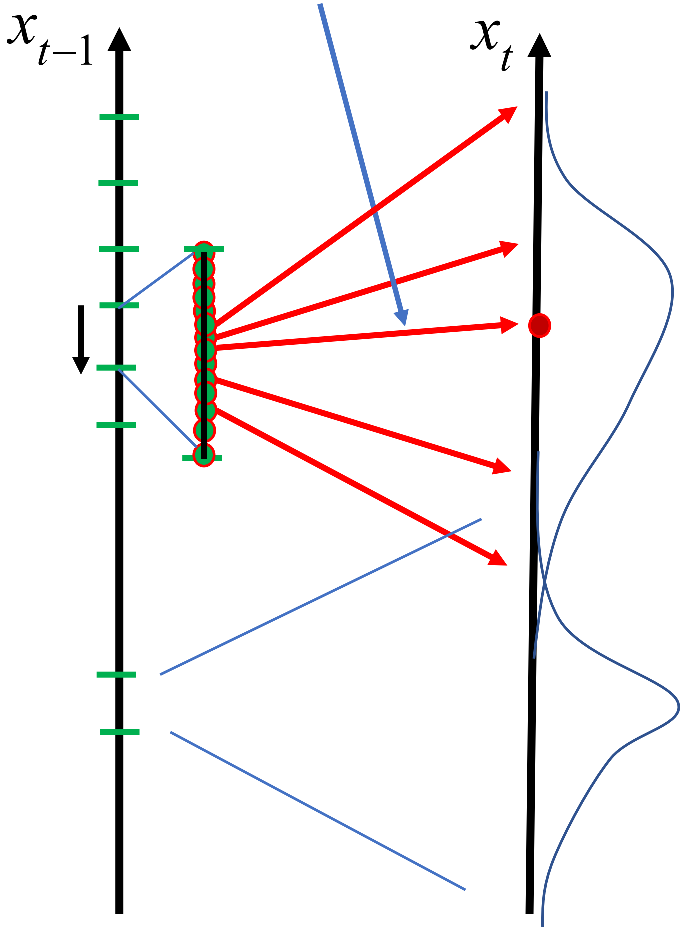
Then we can approximate $p(x_t \mid \bar{x}_{t-1,j}, u_{t-1})$.

Repeating this for all small intervals Δx_{t-1}
we can form the particle distribution.

$\Pr(x_t \mid x_{t-1}, u_{t-1})$ comes from the process noise pdf

$$x_{t|t-1}^{(i)} = \underbrace{f(x_{t-1}, u_{t-1})}_{\text{Deterministic state transition}} + w_{t-1}^{(i)}$$

Draw from pdf $f_W(w_{t-1})$



In reality, random variable x_{t-1} is a continuous variable. We draw $x_{t|t-1}^{(i)}$ for each particle $x_{t-1}^{(i)}$ from f_W .

Pseudo Code

For i = 1 to M

Draw $w_{t-1}^{(i)}$ from $f_W(w_{t-1})$

$$x_{t|t-1}^{(i)} = f(x_{t-1}^{(i)}, u_{t-1}) + w_{t-1}^{(i)}$$

End

Form

$$\tilde{X}_{t|t-1} = \left\{ \begin{matrix} x_{t|t-1}^{(1)} & x_{t|t-1}^{(2)} & \cdots & x_{t|t-1}^{(M)} \end{matrix} \right\}$$

This approximates $g_{t|t-1}(x_t)$.

Step 2: Update

Assimilate a new observation y_t and update the particles $\tilde{X}_{t|t-1}$.

$$\tilde{X}_{t|t-1} \xrightarrow{\quad y_t \quad} \tilde{X}_t$$

Bayes' Rule $$g_t(x_t) = \eta \underbrace{p(y_t \mid x_t)}_{} g_{t|t-1}(x_t)$$

$$f_V(v_t) : y_t = h(x_t) + v_t$$

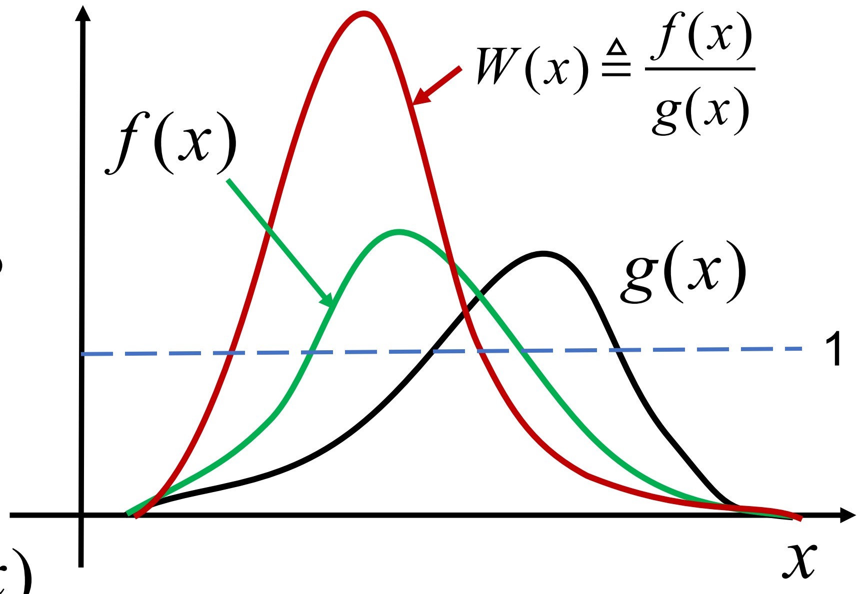
Monte Carlo Approximation cannot be used in its original formula.
The key technique for computing this is Importance Sampling.

Importance Sampling

Consider two pdf's, $f(x)$ and $g(x)$, where samples can be drawn from $g(x)$, but not from $f(x)$.

Assume $g(x) > 0$ for all x where $f(x) > 0$.

How can we generate a particle set to represent $f(x)$?



Importance Sampling

Represent $f(x)$ as a combination of Importance Density $g(x)$ and weight $W(x)$

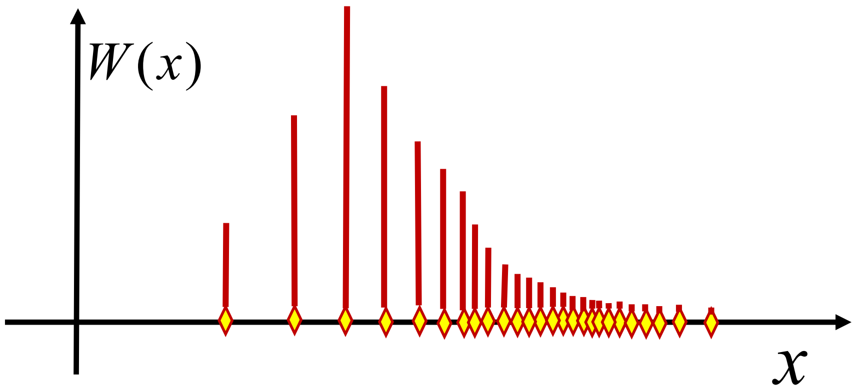
$$f(x) = \frac{f(x)}{g(x)} \cdot g(x)$$

$$W(x) \triangleq \frac{f(x)}{g(x)}$$

Importance Weight

Importance Density

Particles



Applying the Importance Sampling Method to Step 2 Update

$$g_t(x_t) = \eta \underline{p(y_t \mid x_t)} \underline{g_{t|t-1}(x_t)}$$

Given y_t , this term can be computed with the measurement noise model.
Treat this term as an importance weight: $W(x)$

This has been represented by particles

$$\tilde{X}_{t|t-1} = \left\{ \begin{array}{cccc} x_{t|t-1}^{(1)} & x_{t|t-1}^{(2)} & \cdots & x_{t|t-1}^{(M)} \end{array} \right\}$$

Samples have been drawn → $g(x)$

How can we generate particles representing $g_t(x_t)$ from $\tilde{X}_{t|t-1}$ and $W(x)$?

Use the Cumulative Density Function (cdf)

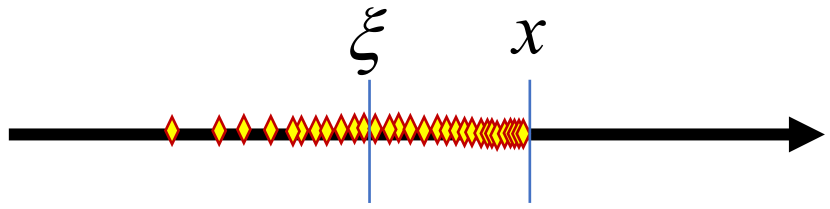
Consider cdf of $f(x)$

$$F(x) = \int_{-\infty}^{x} f(\xi)\,d\xi = \int_{-\infty}^{x} \frac{f(\xi)}{g(\xi)} g(\xi)\,d\xi = \int_{-\infty}^{x} w(\xi) g(\xi)\,d\xi$$

Collect particles up to x.

Define a membership function:

$$I(\xi, x) = \begin{cases} 1 : \xi \le x \\ 0 : \xi > x \end{cases}$$



Using this function yields $$F(x) \cong \frac{1}{W_0} \sum_{i=1}^{M} W(x^{(i)}) I(x^{(i)}, x)$$

where $$W_0 = \sum_{i=1}^{M} W(x^{(i)})$$

The Cumulative Density Function of $g_t(x_t)$

$$G_t(x_t) = \int_{-\infty}^{x_t} g_t(\xi)\, d\xi$$

$$= \frac{1}{W_0} \sum_{i=1}^{M} W(x_{t|t-1}^{(i)}; y_t) I(x_{t|t-1}^{(i)}, x_t)$$

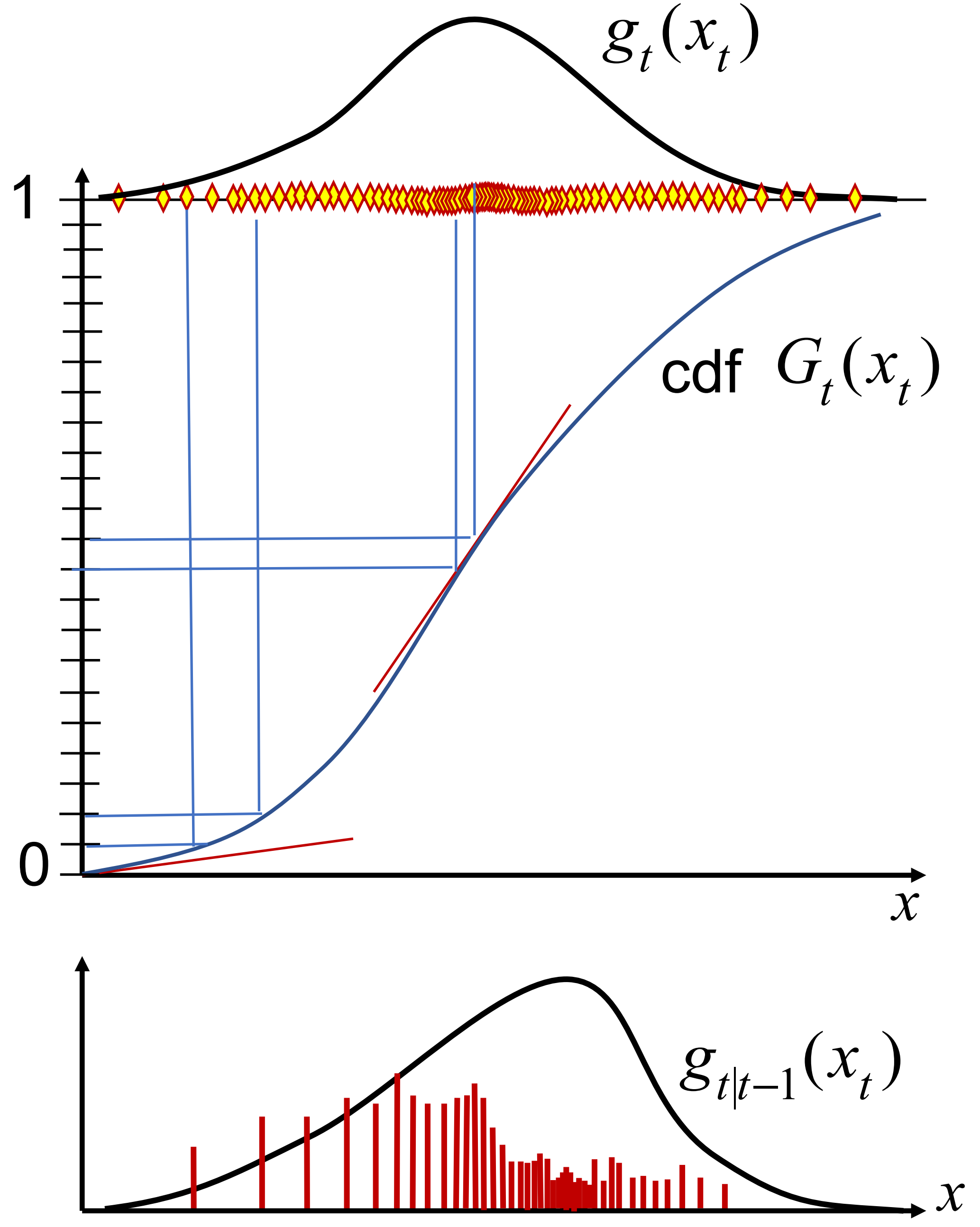
For additive measurement noise,

$$W(x_{t|t-1}^{(i)}; y_t) = f_V(y_t - h(x_{t|t-1}^{(i)}))$$

Recall $y_t = h(x_t) + v_t$

From $G_t(x_t)$ we can draw new M particles to form:

$$\tilde{X}_t = \left\{ \begin{array}{cccc} x_t^{(1)} & x_t^{(2)} & \cdots & x_t^{(M)} \end{array} \right\}$$



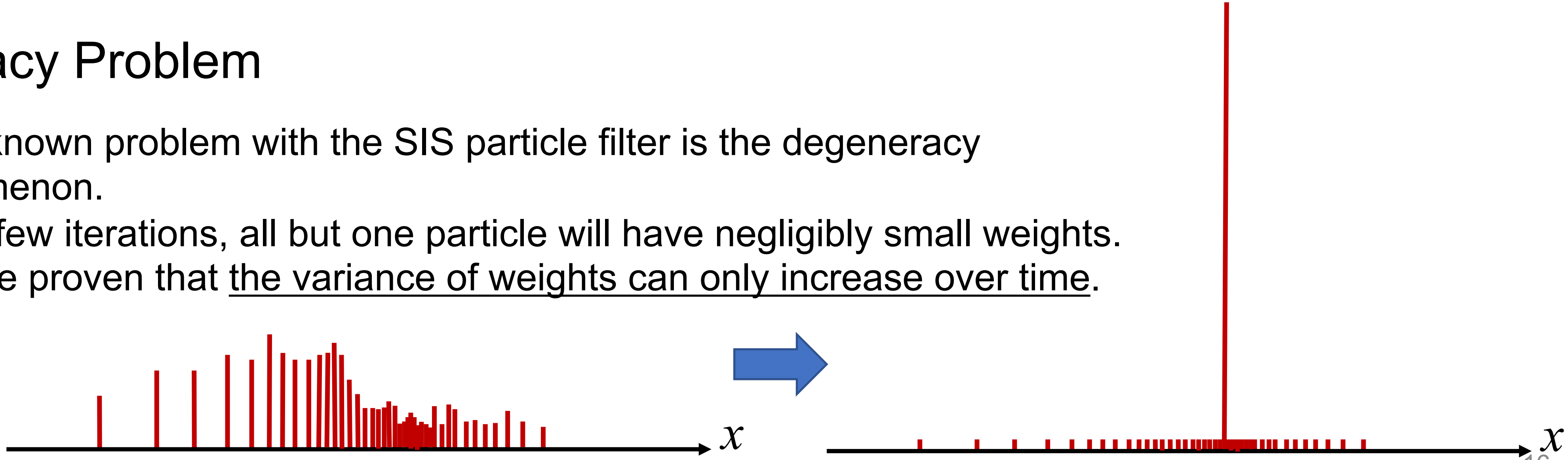
Sequential Importance Sampling (SIS)

Given $g_{t|t-1}(x_t)$ and y_t, SIS draws $x_t^{(i)}; i = 1 \cdots M$ from $g_{t|t-1}(x_t)$ and computes weights $W_t^{(i)}; i = 1 \cdots M$, sequentially.
The weights can be updated recursively as:

$$W_t^{(i)} \propto W_{t-1}^{(i)} \frac{p(y_t \mid x_t^{(i)}) p(x_t^{(i)} \mid x_{t-1}^{(i)})}{g_{t-1}(x_t^{(i)} \mid x_{t-1}^{(i)})}; \quad i = 1 \cdots M,$$

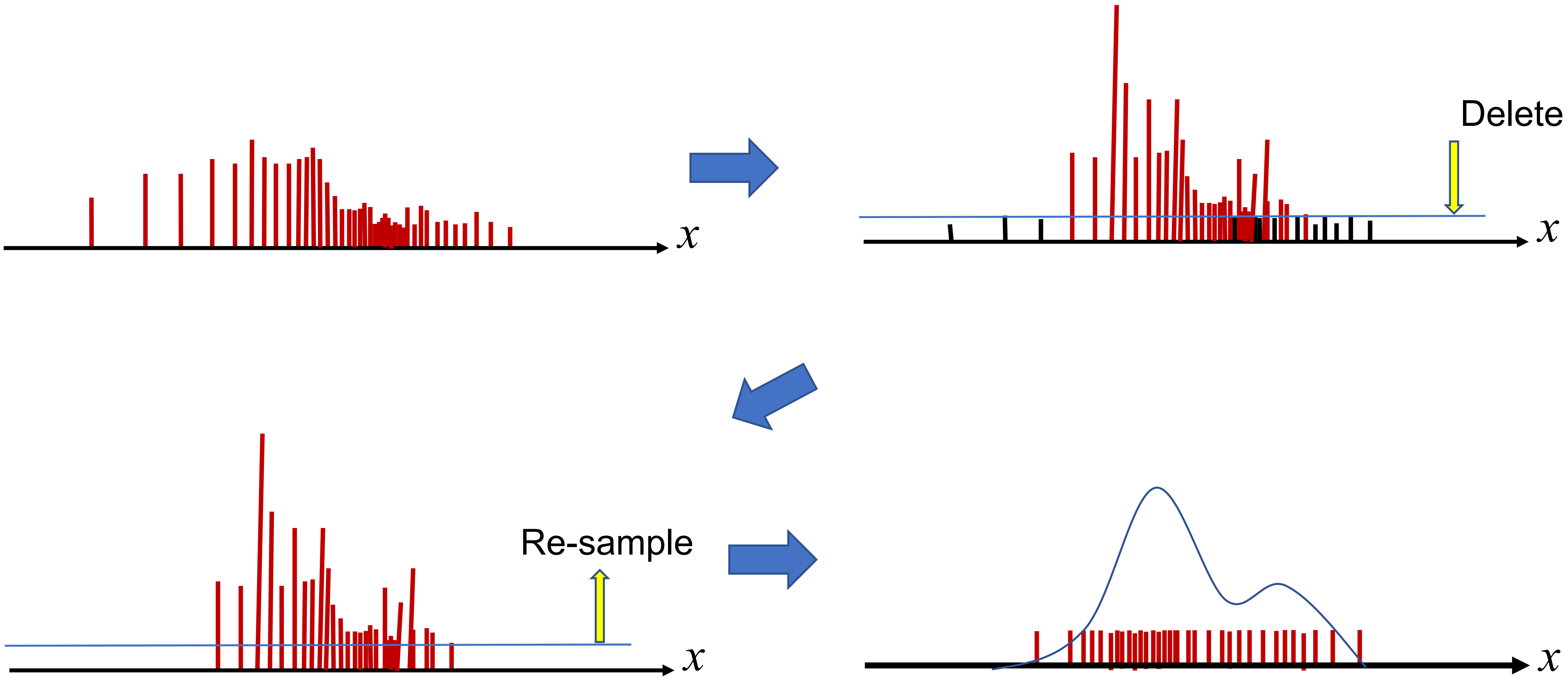
Degeneracy Problem

A well-known problem with the SIS particle filter is the degeneracy phenomenon.
After a few iterations, all but one particle will have negligibly small weights.
It can be proven that the variance of weights can only increase over time.



Re-sampling

Before the variance of weights becomes very large, delete particles with small weights using a threshold.

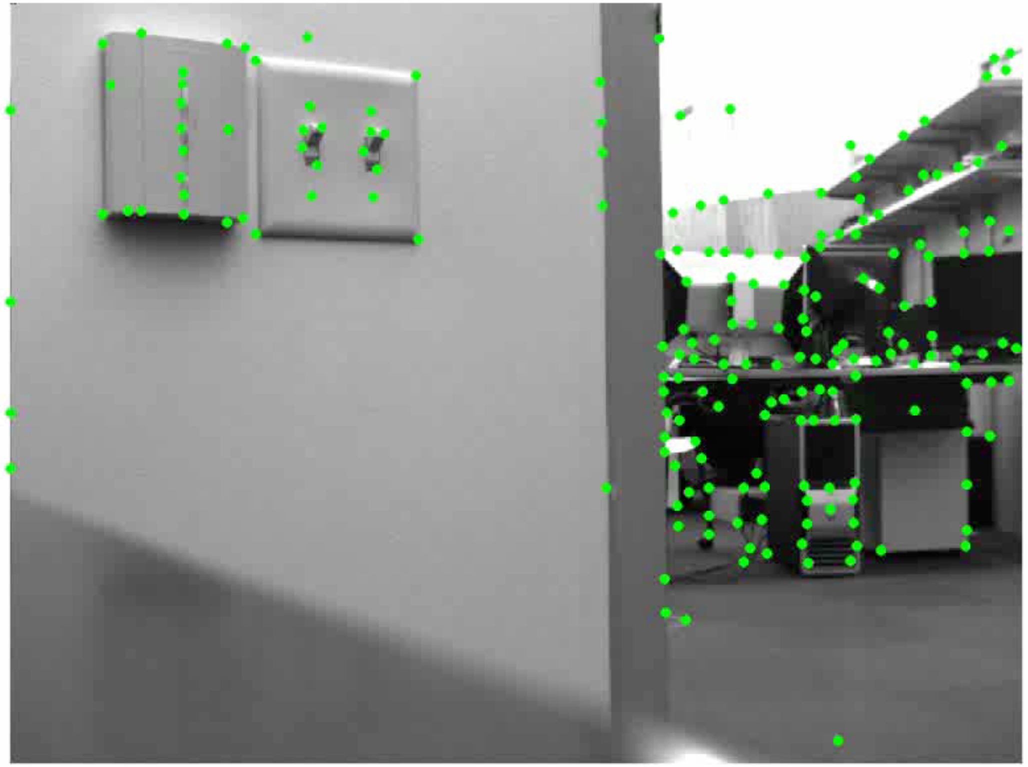


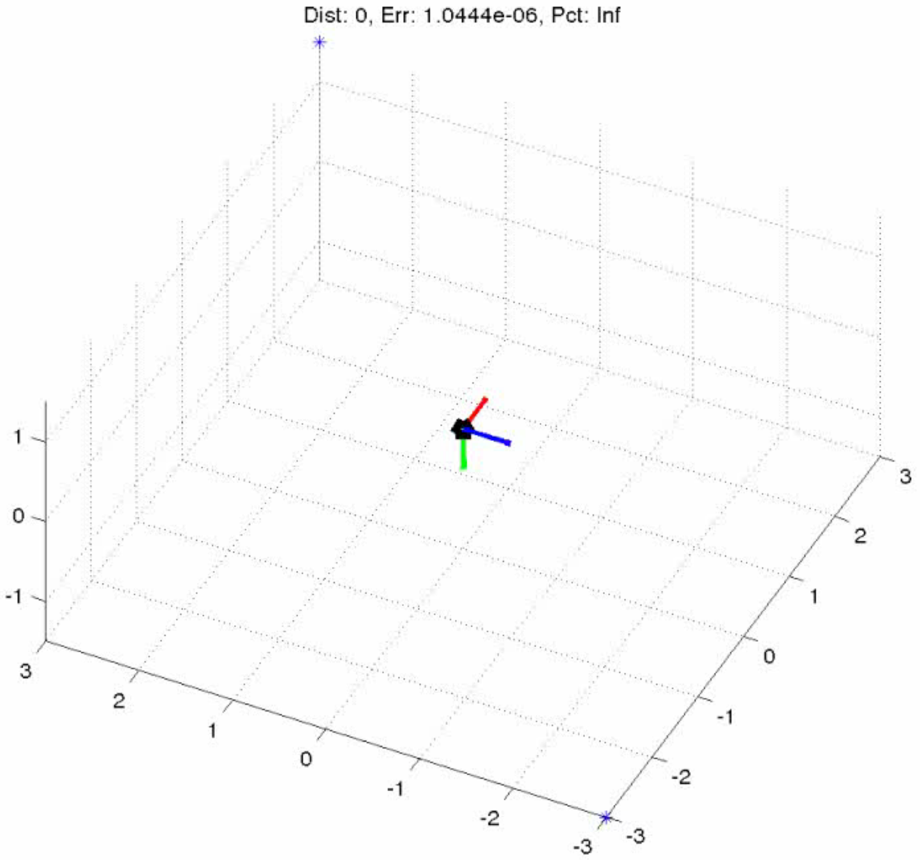
M particles

Uniform weights

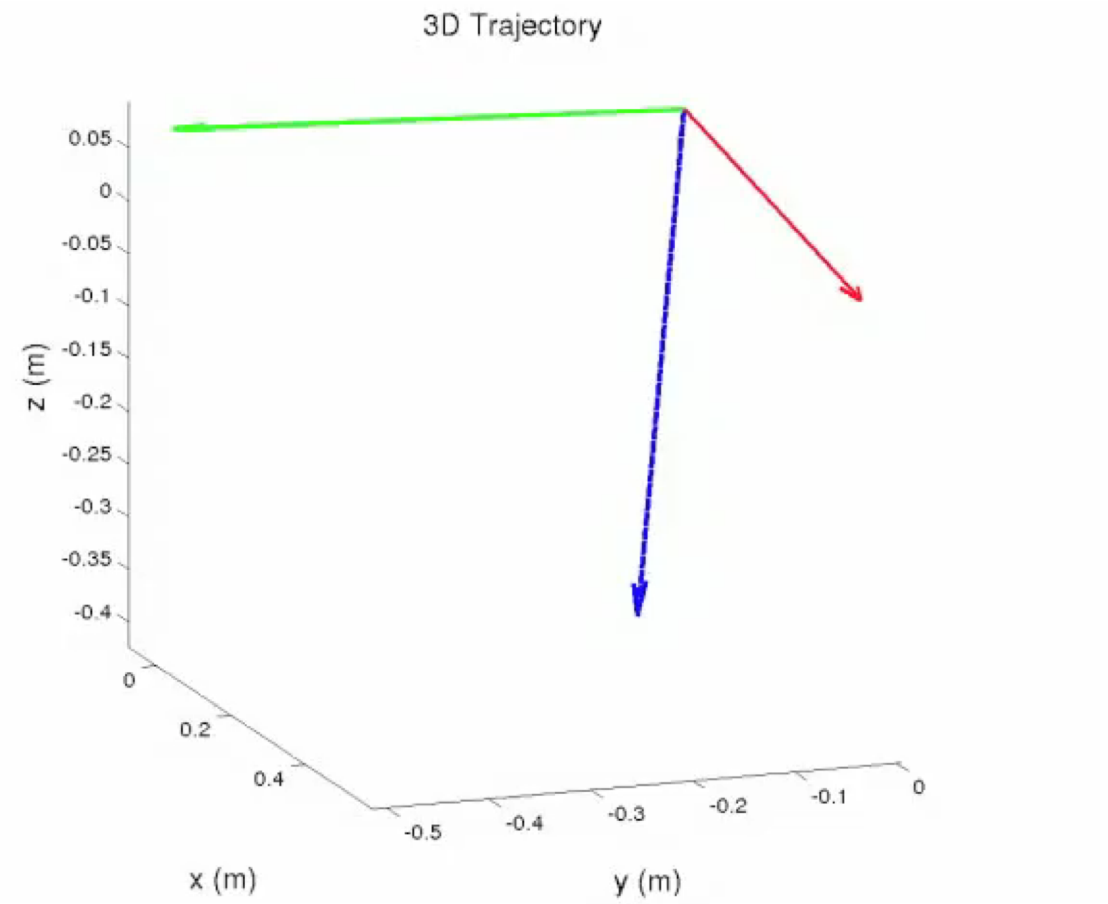
Cameras and IMU

Estimate the 3D position and orientation using Particle Filter





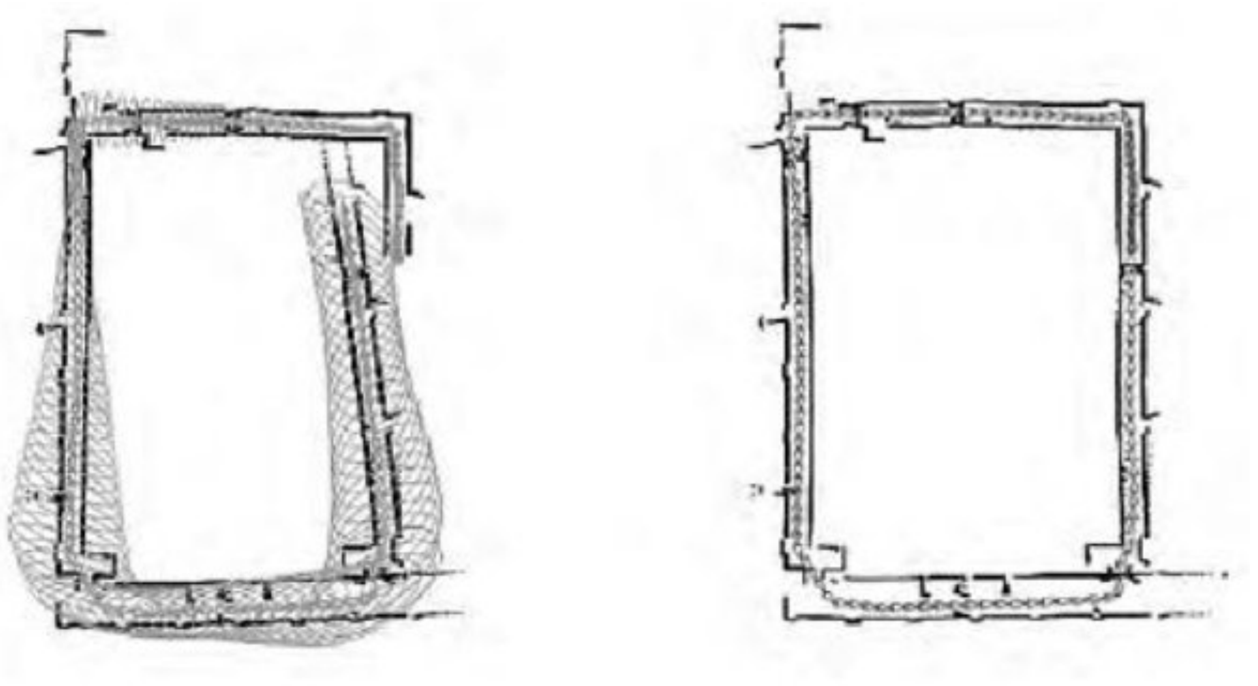
Credit: Professor John Leonard



Credit: Professor John Leonard

- Rao-Blackwellized Particle Filter
- Color-tile Particle Filter
- Fast SLAM

- Loop-Closure



(a) (b)