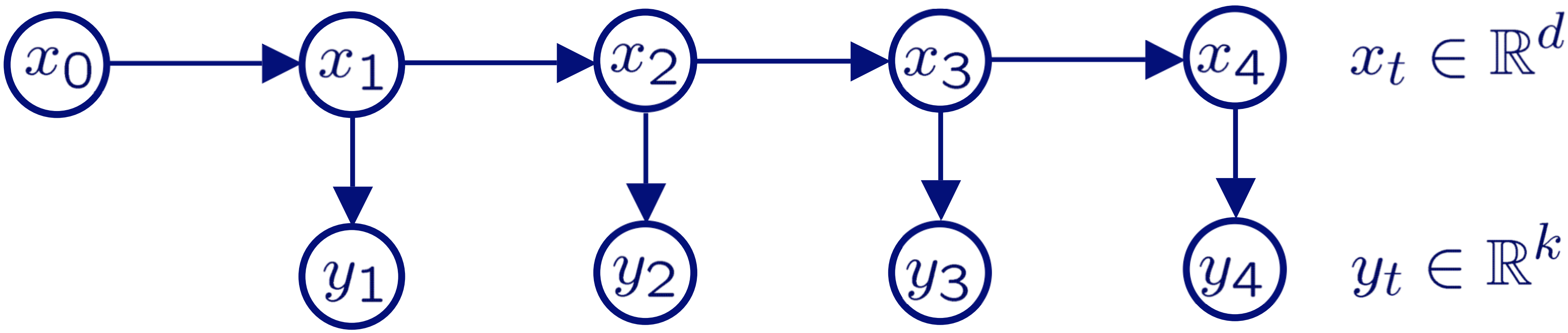
Learning and Inference in Probabilistic Graphical Models

Particle Filters and Sequential Monte Carlo

April 14, 2010

Linear State Space Models

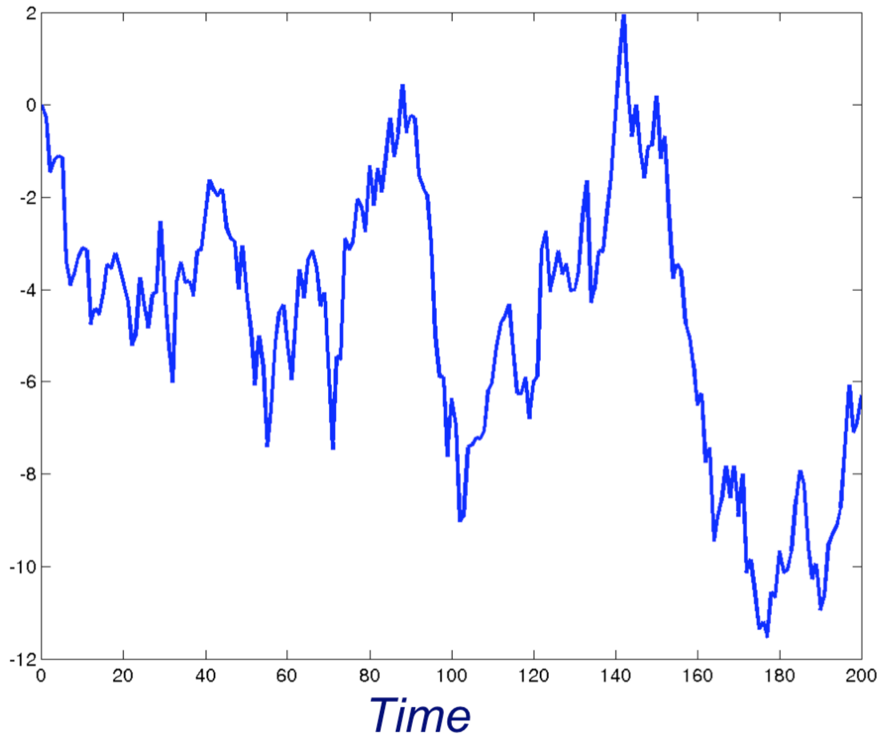


$$
\begin{aligned}
x_{t+1} &= Ax_t + w_t \qquad & w_t &\sim \mathcal{N}(0, Q) \\
y_t &= Cx_t + v_t \qquad & v_t &\sim \mathcal{N}(0, R)
\end{aligned}
$$

- States & observations jointly Gaussian:
  - All marginals & conditionals Gaussian
  - Linear transformations remain Gaussian

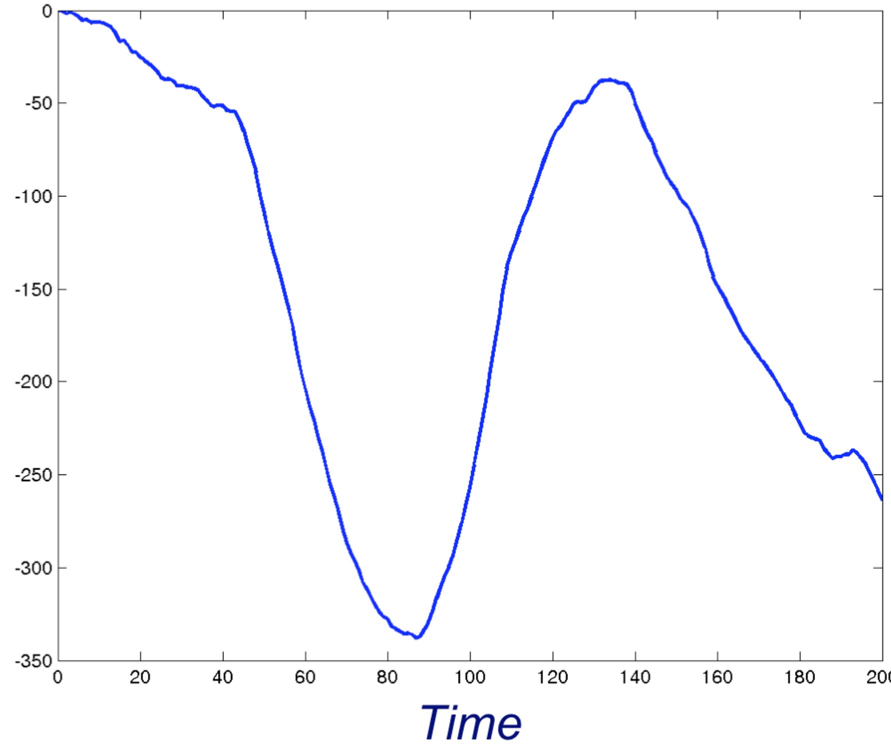
Simple Linear Dynamics

Brownian Motion



$$x_{t+1} = x_t + w_t$$

Constant Velocity



$$\begin{bmatrix} x_{t+1} \\ \delta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \delta_t \end{bmatrix} + w_t$$

Kalman Filter

$$x_{t+1} = Ax_t + w_t \qquad w_t \sim \mathcal{N}(0, Q)$$
$$y_t = Cx_t + v_t \qquad v_t \sim \mathcal{N}(0, R)$$

- Represent Gaussians by *mean* & *covariance*:

$$p(x_t \mid y_1, \ldots, y_{t-1}) = \mathcal{N}(x; \tilde{\mu}_t, \tilde{\Lambda}_t)$$
$$p(x_t \mid y_1, \ldots, y_t) = \mathcal{N}(x; \mu_t, \Lambda_t)$$

Prediction:

$$\boxed{\begin{aligned} \tilde{\mu}_t &= A\mu_{t-1} \\ \tilde{\Lambda}_t &= A\Lambda_{t-1}A^T + Q \end{aligned}}$$

Kalman Gain:

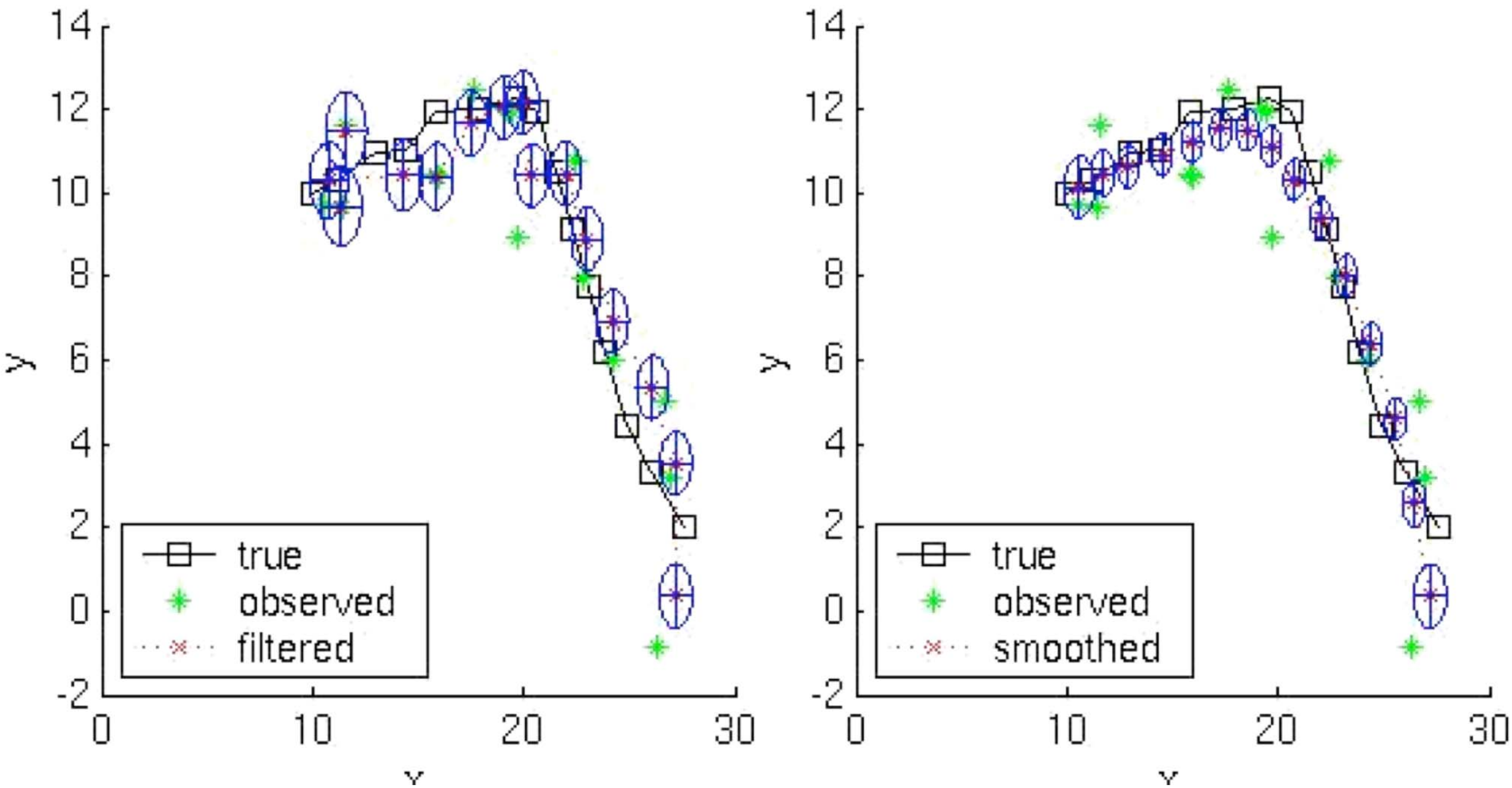
$$K_t = \tilde{\Lambda}_t C^T (C\tilde{\Lambda}_t C^T + R)^{-1}$$

Update:

$$\boxed{\begin{aligned} \mu_t &= \tilde{\mu}_t + K_t(y_t - C\tilde{\mu}_t) \\ \Lambda_t &= \tilde{\Lambda}_t - K_t C\tilde{\Lambda}_t \end{aligned}}$$

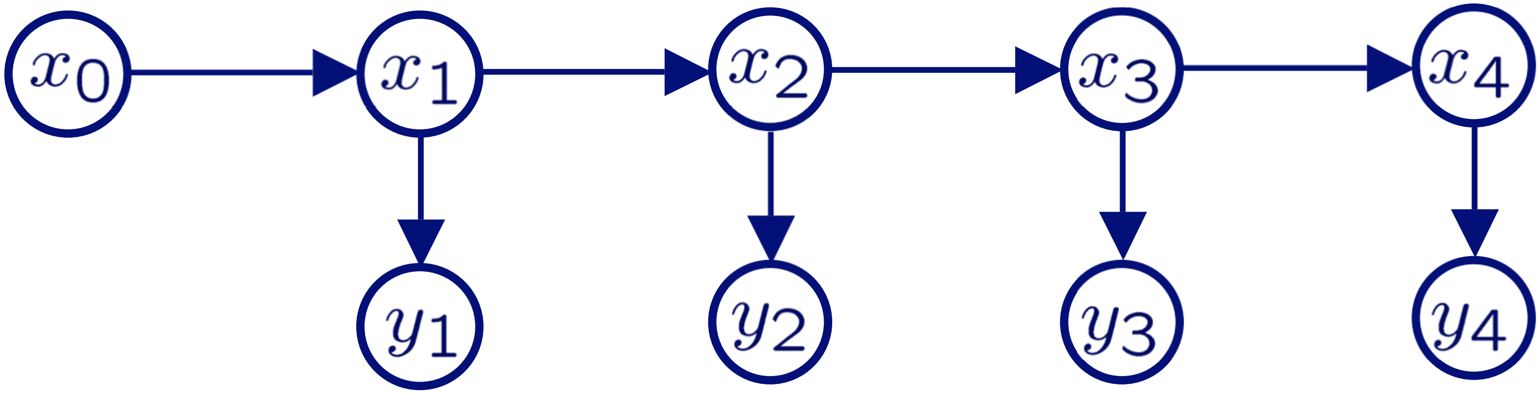
Constant Velocity Tracking

Kalman Filter *Kalman Smoother*



(K. Murphy, 1998)

Nonlinear State Space Models



$x_t \in \mathbb{R}^d$

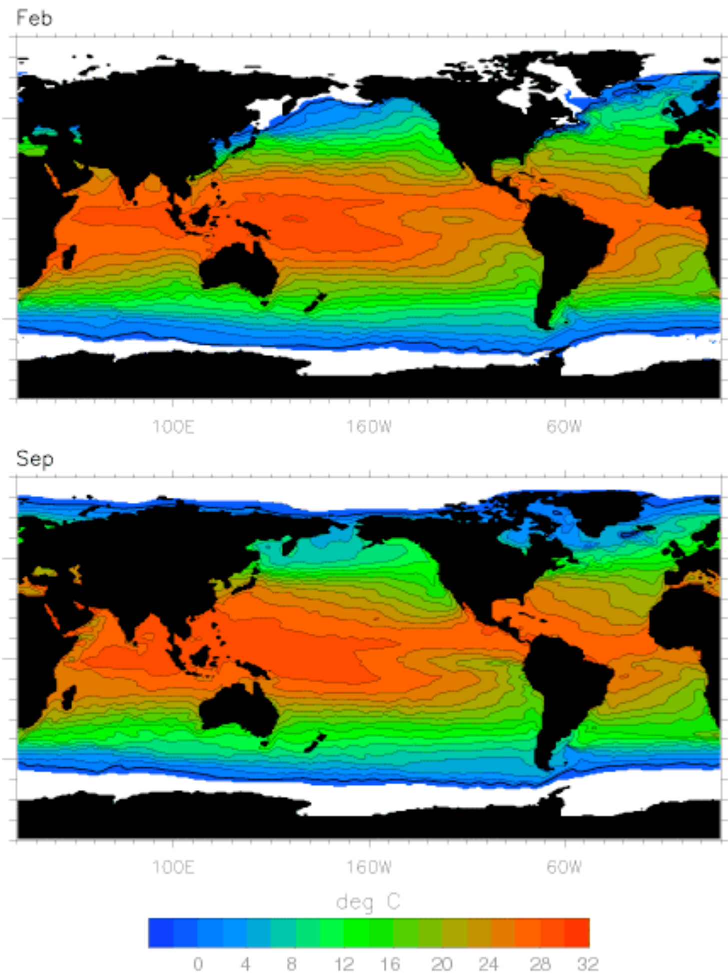
$y_t \in \mathbb{R}^k$

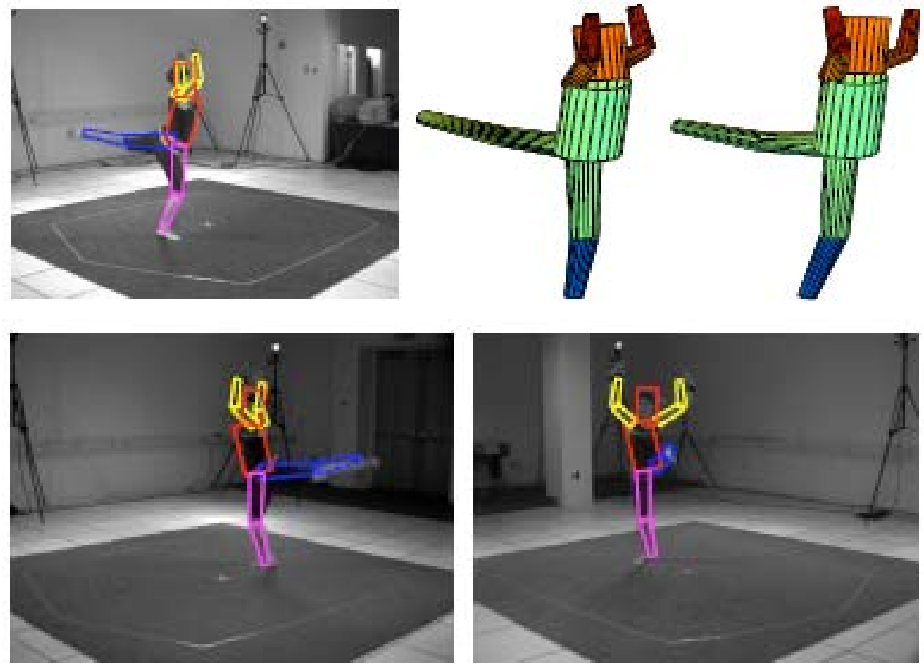
$$x_{t+1} = f(x_t, w_t) \qquad w_t \sim \mathcal{F}$$

$$y_t = g(x_t, v_t) \qquad v_t \sim \mathcal{G}$$

- State dynamics and measurements given by potentially complex *nonlinear functions*
- Noise sampled from *non-Gaussian* distributions

Examples of Nonlinear Models

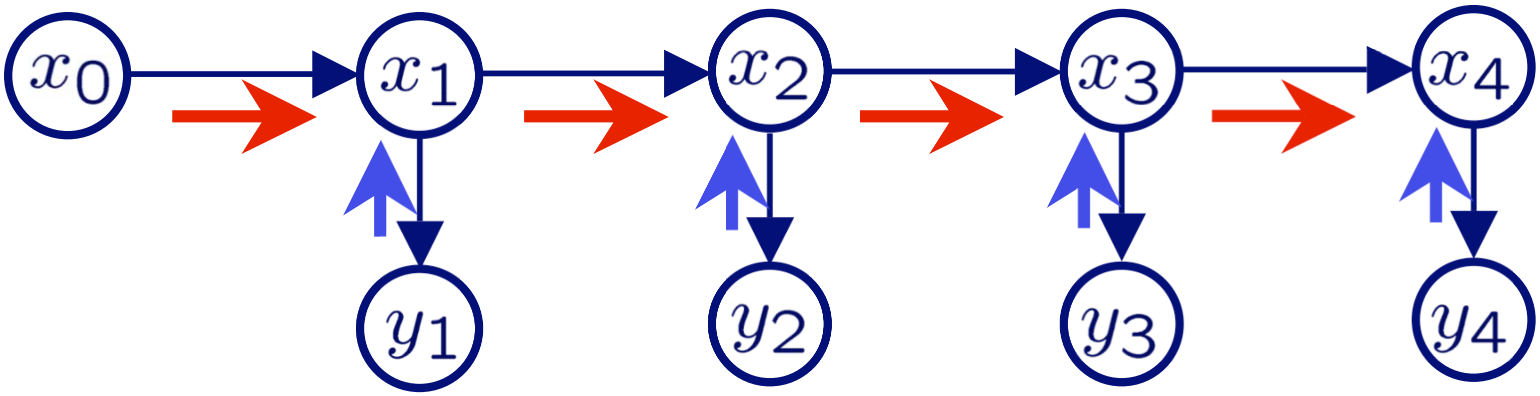




Observed image is a complex function of the 3D pose, other nearby objects & clutter, lighting conditions, camera calibration, etc.

Dynamics implicitly determined by geophysical simulations

Nonlinear Filtering



$$p(x_t \mid y_1, \ldots, y_{t-1}) = \tilde{q}_t(x_t)$$

$$p(x_t \mid y_1, \ldots, y_t) = q_t(x_t)$$

Prediction:

$$\tilde{q}_t(x_t) = \int p(x_t \mid x_{t-1}) q_{t-1}(x_{t-1}) \, dx_{t-1}$$

Update:

$$q_t(x_t) = \frac{1}{Z_t} \tilde{q}_t(x_t) p(y_t \mid x_t)$$

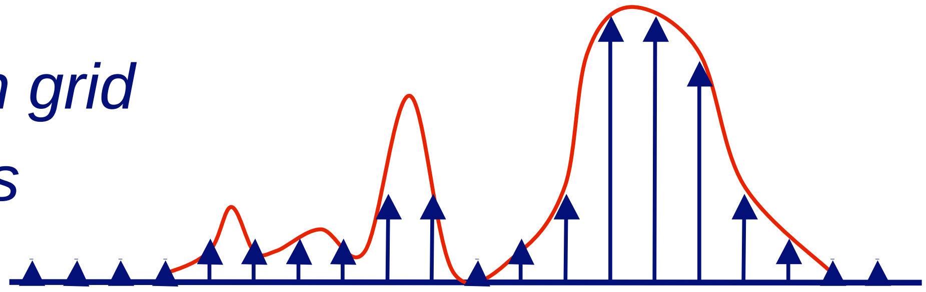
Approximate Nonlinear Filters

$$q_t(x_t) \propto p(y_t \mid x_t) \cdot \int p(x_t \mid x_{t-1}) q_{t-1}(x_{t-1}) \, dx_{t-1}$$

- No direct *representation* of continuous functions, or closed form for the prediction *integral*
- Big literature on approximate filtering:
  - *Histogram filters*
  - *Extended & unscented Kalman filters*
  - *Particle filters*
  - *…*

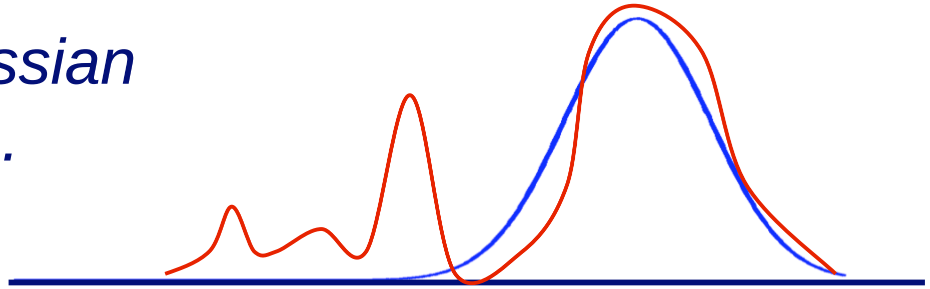
Nonlinear Filtering Taxonomy

Histogram Filter:



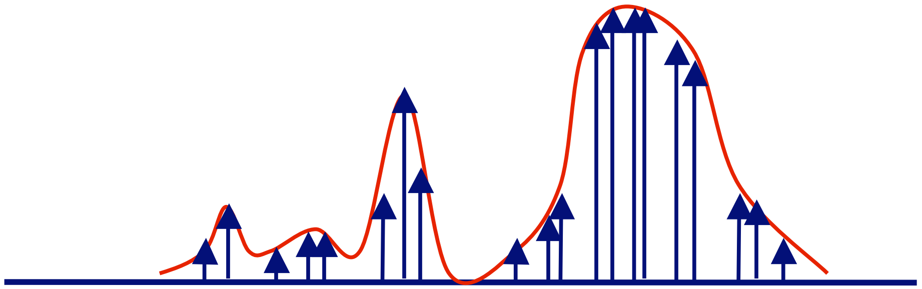
- Evaluate on fixed discretization grid
- Only feasible in low dimensions
- Expensive or inaccurate

Extended/Unscented Kalman Filter:



- Approximate posterior as Gaussian via linearization, quadrature, …
- Inaccurate for multimodal posterior distributions

Particle Filter:

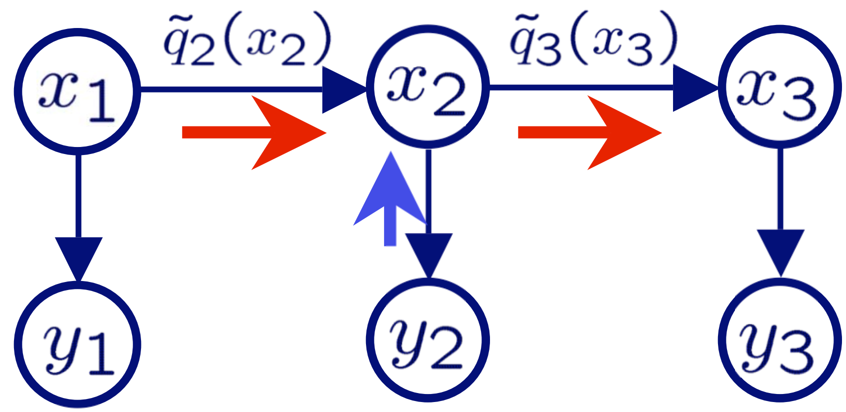


- Dynamically evaluate states with highest probability
- Monte Carlo approximation

Particle Filters

Condensation, Sequential Monte Carlo, Survival of the Fittest,…

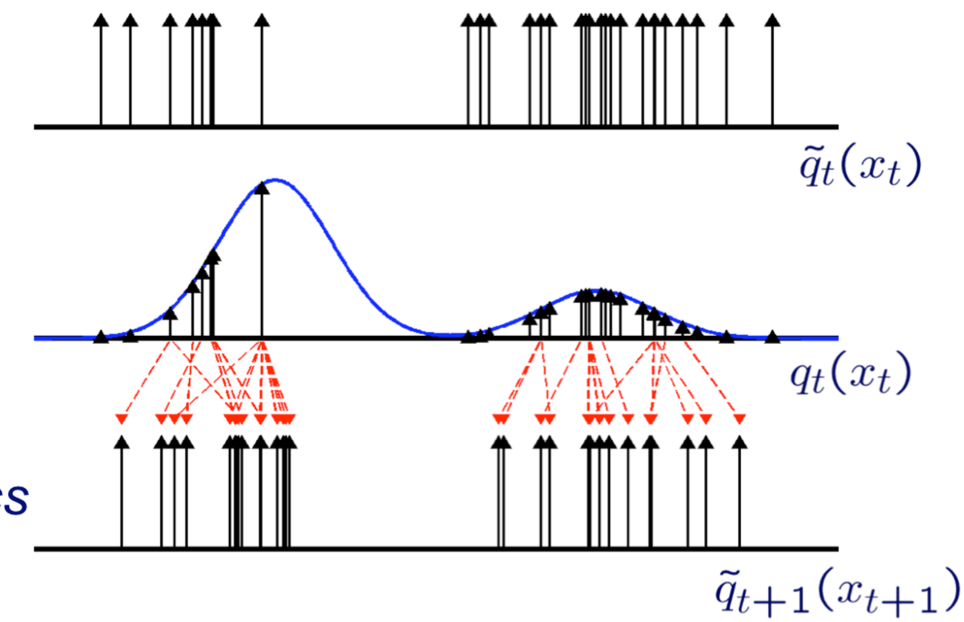
- Represent state estimates using a set of samples
- Propagate over time using *importance sampling*



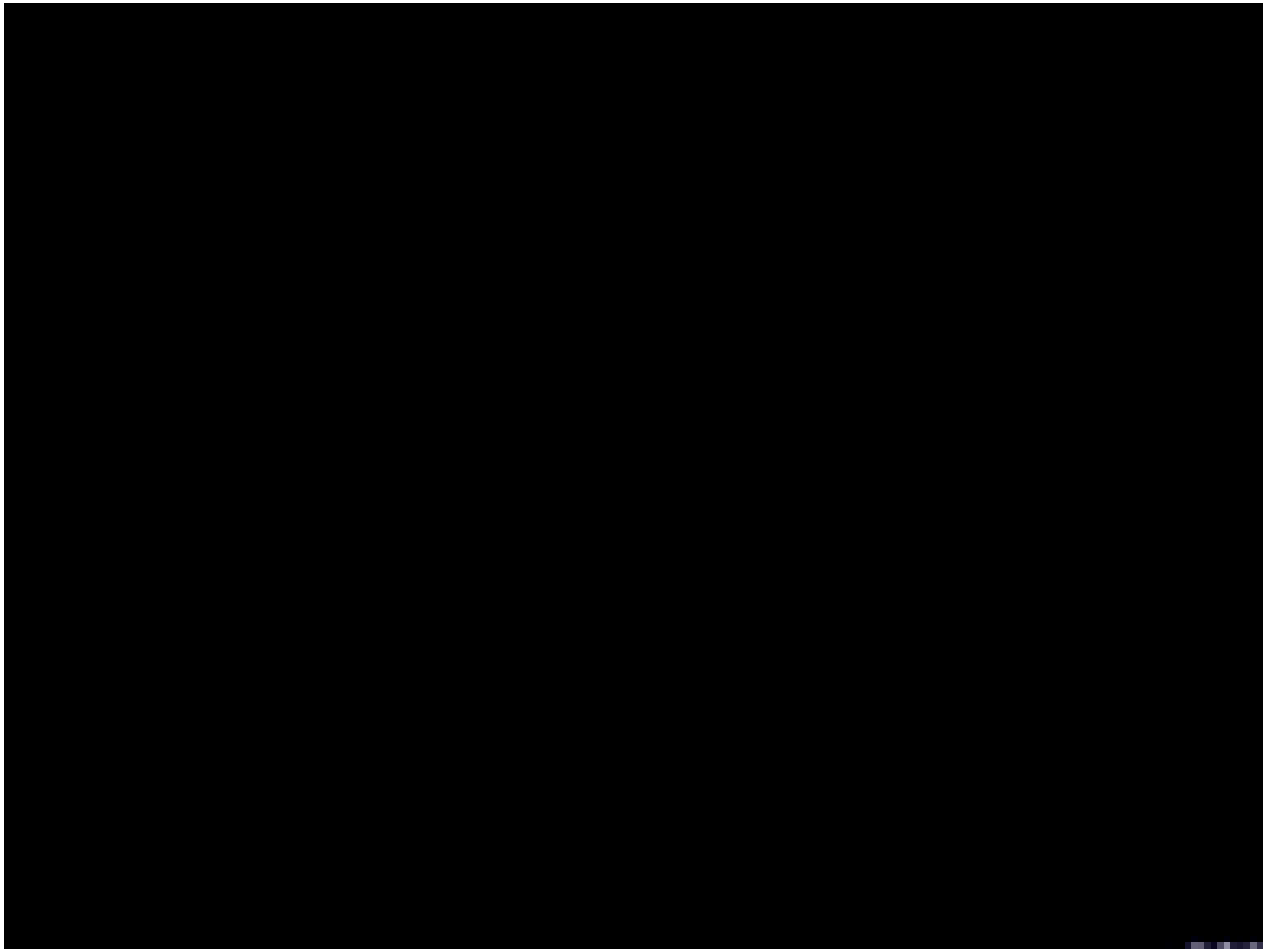
Sample-based density estimate

Weight by observation likelihood

Resample & propagate by dynamics



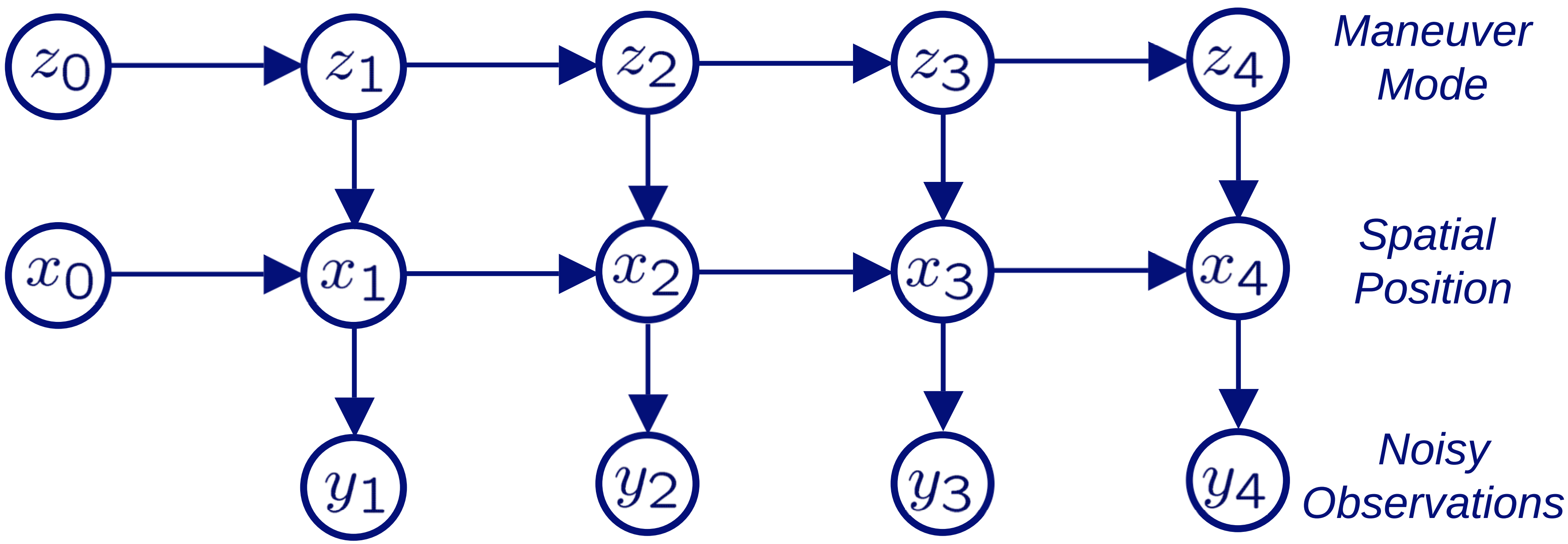
Particle Filtering Movie



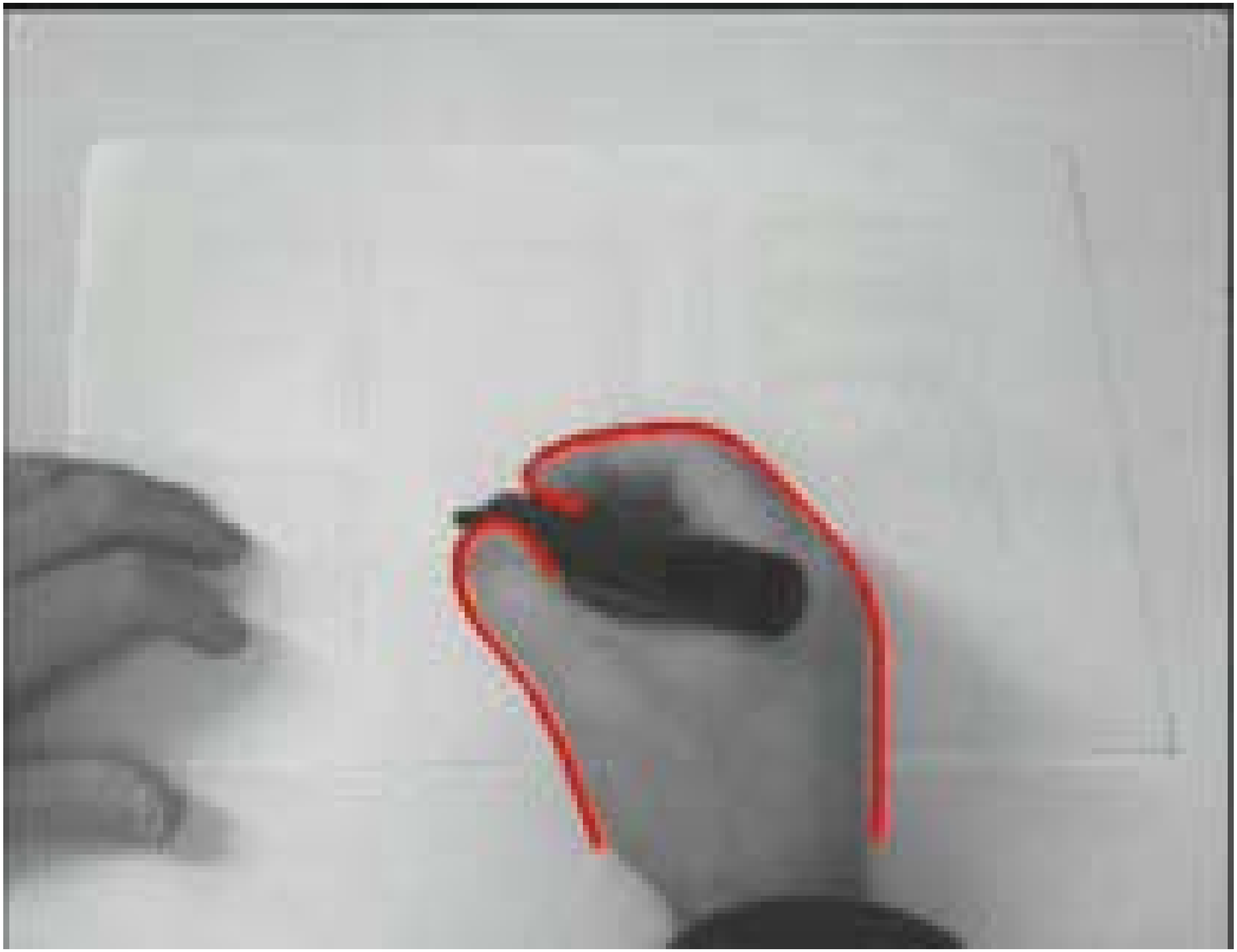
(M. Isard, 1996)

Dynamic Bayesian Networks

Specify and exploit internal structure in the hidden states underlying a time series



DBN Hand Tracking Video



Isard et. al., 1998

Particle Filtering Caveats

- Particle filters are easy to implement, and effective in many applications, BUT
  - *It can be difficult to know how many samples to use, or to tell when the approximation is poor*
  - *Sometimes suffer catastrophic failures, where NO particles have significant posterior probability*
  - *This is particularly true with “peaky” observations in high-dimensional spaces:*