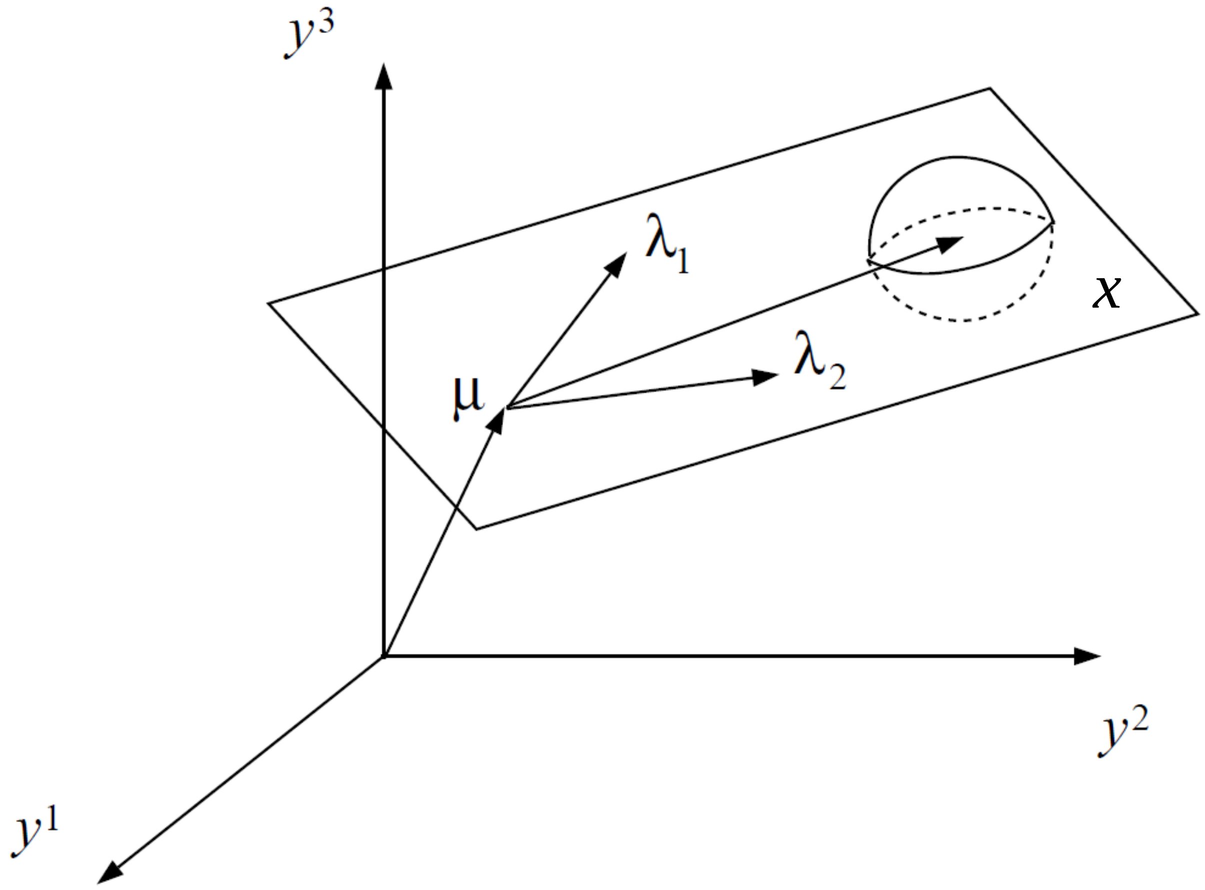
Computer Science

CSC 665-1: Advanced Topics in Probabilistic Graphical Models

State-Space Models & Dynamical Systems

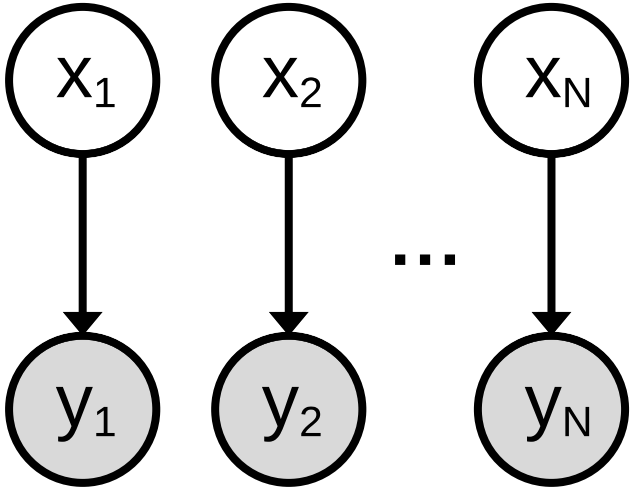
Instructor: Prof. Jason Pacheco

(Bayesian) Principal Component Analysis



Latent: $x \in \mathbb{R}^p$ Data: $y \in \mathbb{R}^q$

Typically p<q for dimension reduction



$$x \sim \mathcal{N}(0, I)$$

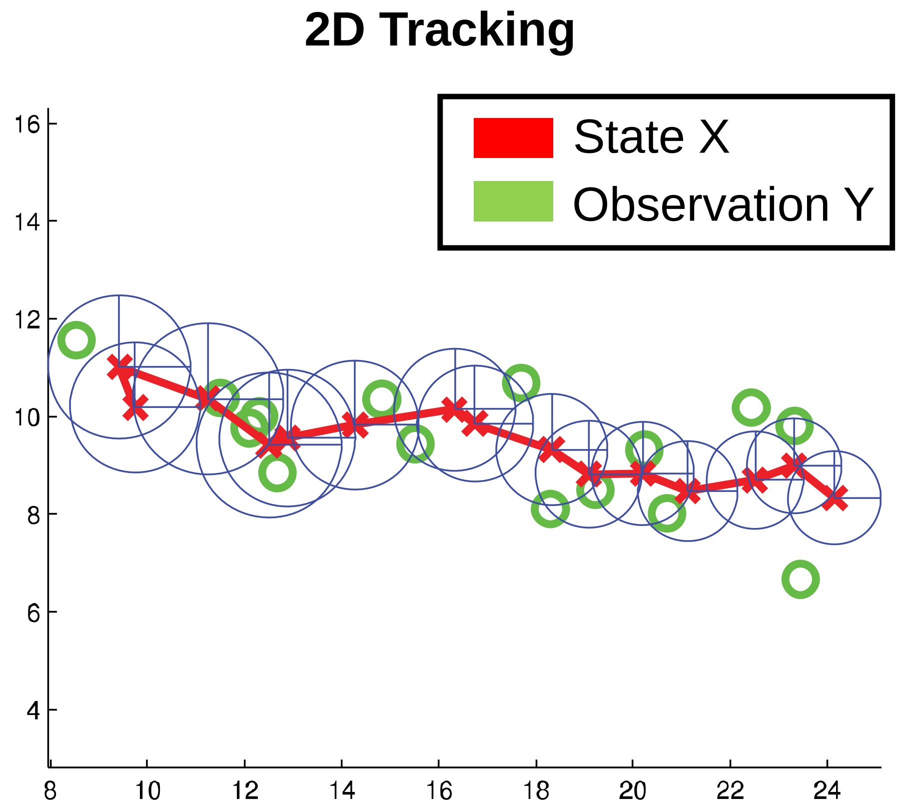
$$y \mid x \sim \mathcal{N}(\Lambda x + \mu, \sigma^2 I)$$

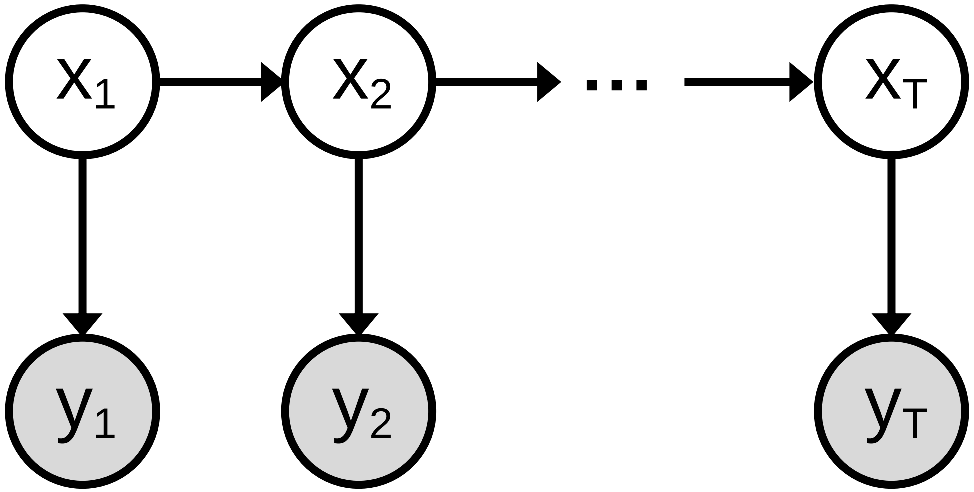
Data are exchangeable linear Gaussian projections of latent quantities

[Source: M. I. Jordan]

Gaussian Linear Dynamical System (LDS)

Temporal extension of probabilistic PCA…





$$x_0 \sim \mathcal{N}(0, I)$$

$$x_t \mid x_{t-1} \sim \mathcal{N}(Fx_{t-1}, \Sigma)$$

$$y_t \mid x_t \sim \mathcal{N}(Hx_t, R)$$

Data are time-dependent and non-exchangeable

Linear State-Space Model

- Consider the state vector:

$$x_t = \begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix} \quad \text{where} \quad x(t) \text{ : Position} \quad \dot{x}(t) \triangleq \frac{d}{dt}x(t) \text{ : Velocity}$$

- Differential equations for constant velocity dynamics:

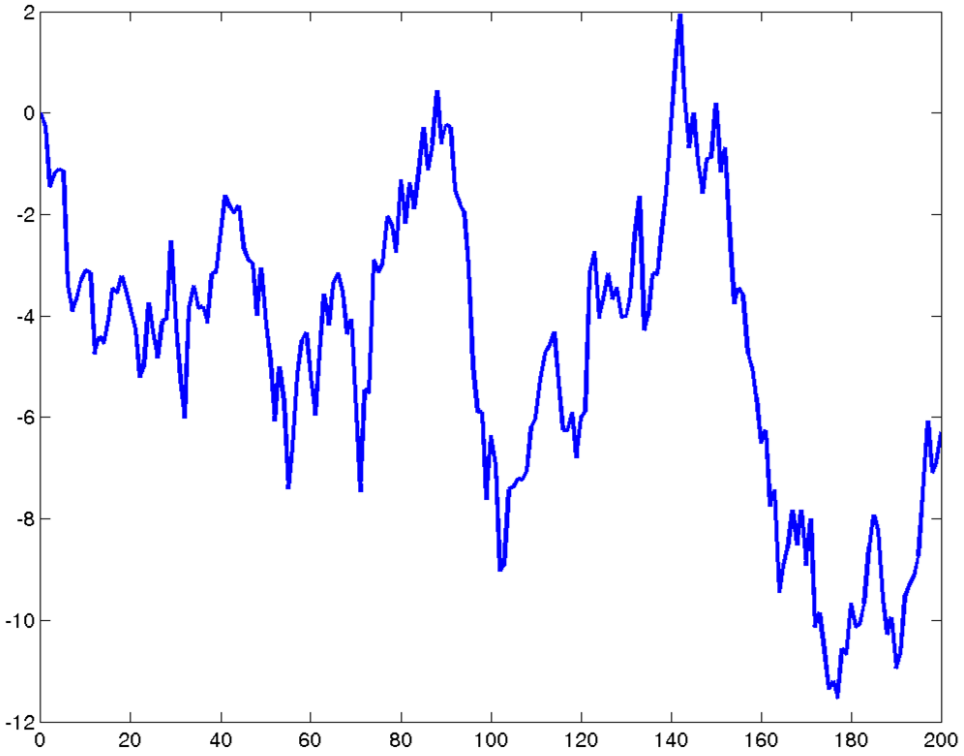
$$x(t) = x(t-1) + \dot{x}(t-1) \qquad \dot{x}(t) = \dot{x}(t-1)$$

- Linear Gaussian state-space model

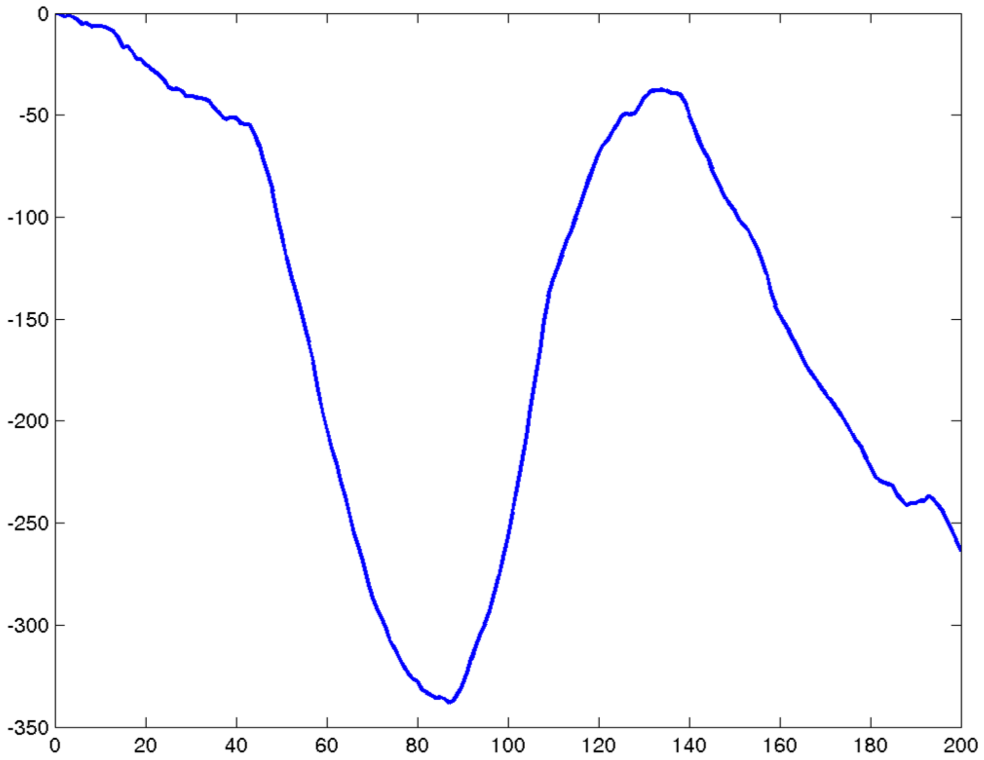
$$x_t = Fx_{t-1} + \epsilon \quad \text{where} \quad F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \epsilon \sim \mathcal{N}(0, \Sigma)$$

Simple Linear Gaussian Dynamics

Random Walk (Brownian Motion)



Constant Velocity (a.k.a. zero acceleration)

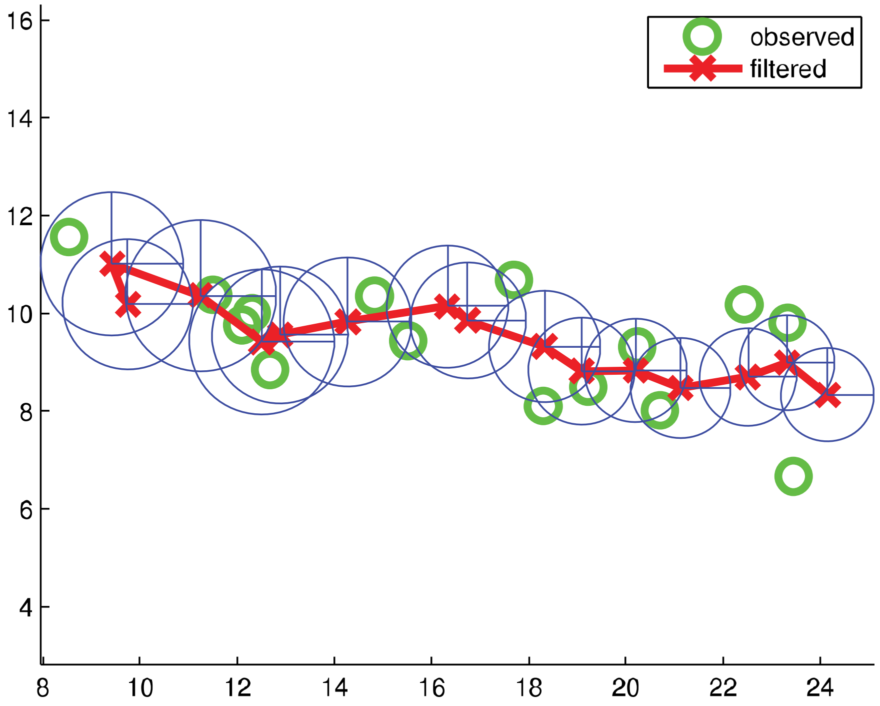


Acceleration can be included in higher-order models as well

Dynamical System Inference

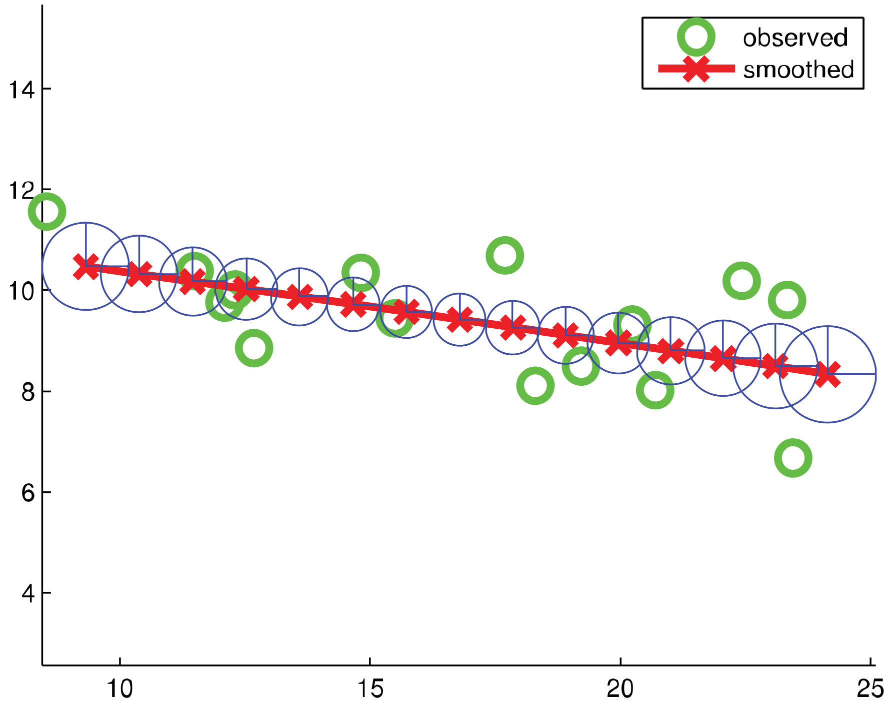
Define shorthand notation: $y_1^{t-1} \triangleq \{y_1, \ldots, y_{t-1}\}$

Filtering



Compute $p(x_t \mid y_1^t)$ at each time t

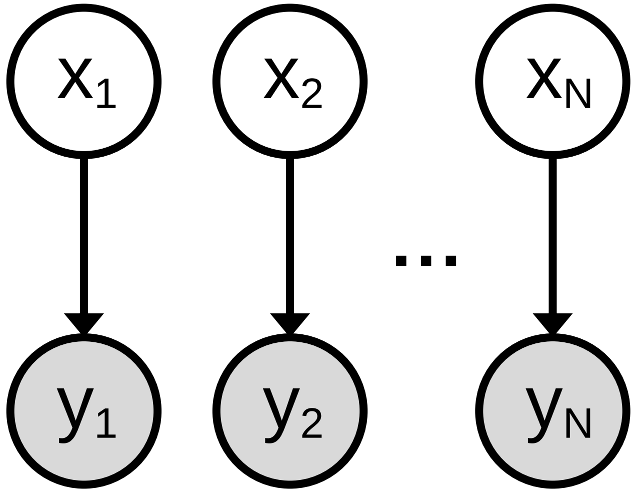
Smoothing



Compute full posterior marginal $p(x_t \mid y_1^T)$ at each time t

Linear Gaussian Inference

Generative Linear Regression



$$x \sim \mathcal{N}(\mu, \Sigma)$$

$$y \mid x \sim \mathcal{N}(Ax + b, R)$$

Marginal likelihood is Gaussian:

$$p(y) = \mathcal{N}(A\mu + b, R + A\Sigma A^T)$$

Posterior also Gaussian (surprise):

$$p(x \mid y) = \mathcal{N}(\mu_{x|y}, \Sigma_{x|y})$$

where,

$$\Sigma_{x|y}^{-1} = \Sigma^{-1} + A^T R^{-1} A$$

$$\mu_{x|y} = \Sigma_{x|y} \left[A^T R^{-1}(y - b) + \Sigma^{-1}\mu \right]$$

Building block for inference on linear Gaussian dynamical system

Gaussian LDS Filtering

- Suppose we have a Gaussian posterior at time t-1:

$$p(x_{t-1} \mid y_1^{t-1}) = \mathcal{N}(\mu_{t-1}, \Sigma_{t-1}) \qquad \text{where} \qquad y_1^{t-1} \triangleq \{y_1, \ldots, y_{t-1}\}$$

- Forward prediction at time t:

$$p(x_t \mid y_1^{t-1}) = \int p(x_t \mid x_{t-1}) p(x_{t-1} \mid y_1^{t-1}) \, dx_{t-1}$$

$$= \int \mathcal{N}(x_t \mid F x_{t-1}, \Sigma) \mathcal{N}(x_{t-1} \mid \mu_{t-1}, \Sigma_{t-1}) \, dx_{t-1}$$

Integrates to 1

$$= \mathcal{N}(x_t \mid F\mu_{t-1}, \Sigma + F\Sigma_{t-1}F^T) \int \mathcal{N}(x_{t-1} \mid \cdot, \cdot) \, dx_{t-1}$$

Same form as marginal likelihood on previous slide

Gaussian LDS Filtering

- Forward prediction at time t: $p(x_t \mid y_1^{t-1}) = \mathcal{N}(x_t \mid \mu_{t|t-1}, \Sigma_{t|t-1})$

where $\mu_{t|t-1} \triangleq F\mu_{t-1}$ and $\Sigma_{t|t-1} = \Sigma + F\Sigma_{t-1}F^T$

State Prediction **Predicted Covariance**

- Posterior at time t is also Gaussian:

$$p(x_t \mid y_1^t) \propto p(x_t \mid y_1^{t-1})p(y_t \mid x_t)$$

$$= \mathcal{N}(x_t \mid \mu_{t|t-1}, \Sigma_{t|t-1})\mathcal{N}(y_t \mid Hx_t, R) \propto \mathcal{N}(x_t \mid \mu_{t|t}, \Sigma_{t|t})$$

Gain Matrix: $K_t = \Sigma_{t|t-1}H^T(H\Sigma_{t|t-1}H^T + R)^{-1}$

Filter Covariance: $\Sigma_{t|t} = \Sigma_{t|t-1} - K_tH\Sigma_{t|t-1}$

Filter Mean: $\mu_{t|t} = \mu_{t|t-1} + K_t(y_t - H\mu_{t|t-1})$

Can be derived from Gaussian conditional formulas and Woodbury matrix identity

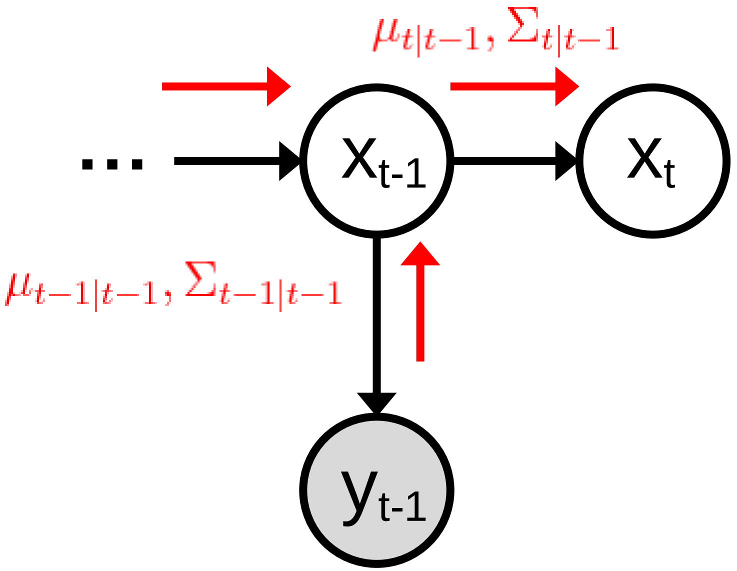
Kalman Filter

Prediction Step:

State Prediction: $\mu_{t|t-1} = F\mu_{t-1|t-1}$

Covariance Prediction:

$$\Sigma_{t|t-1} = \Sigma + F\Sigma_{t-1|t-1}F^T$$

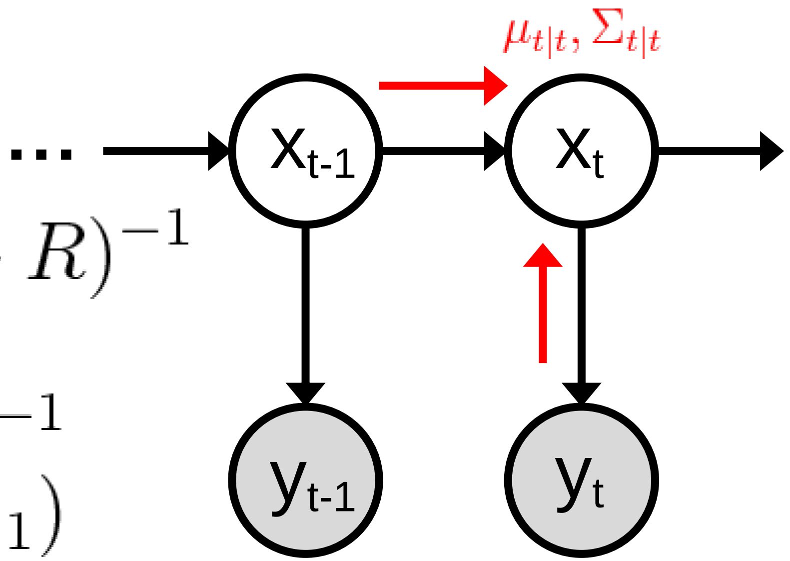


Measurement Update Step:

Gain Matrix: $K_t = \Sigma_{t|t-1}H^T(H\Sigma_{t|t-1}H^T + R)^{-1}$

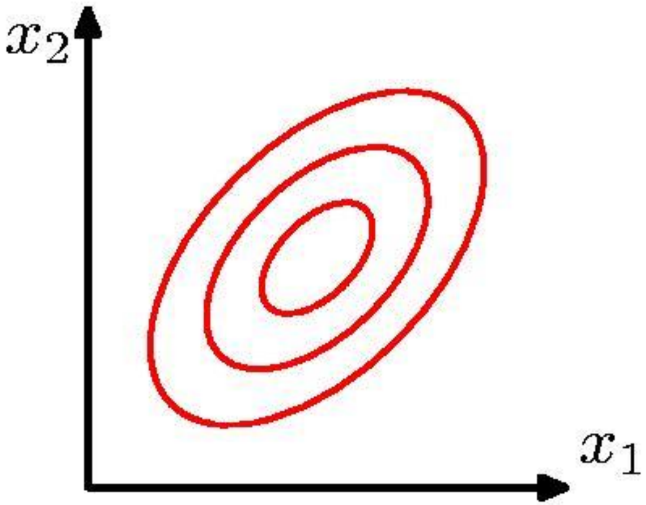
Filter Covariance: $\Sigma_{t|t} = \Sigma_{t|t-1} - K_tH\Sigma_{t|t-1}$

Filter Mean: $\mu_{t|t} = \mu_{t|t-1} + K_t(y_t - H\mu_{t|t-1})$



Gaussian Canonical Parameters

Mean Parameterization:



$$\mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{N/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

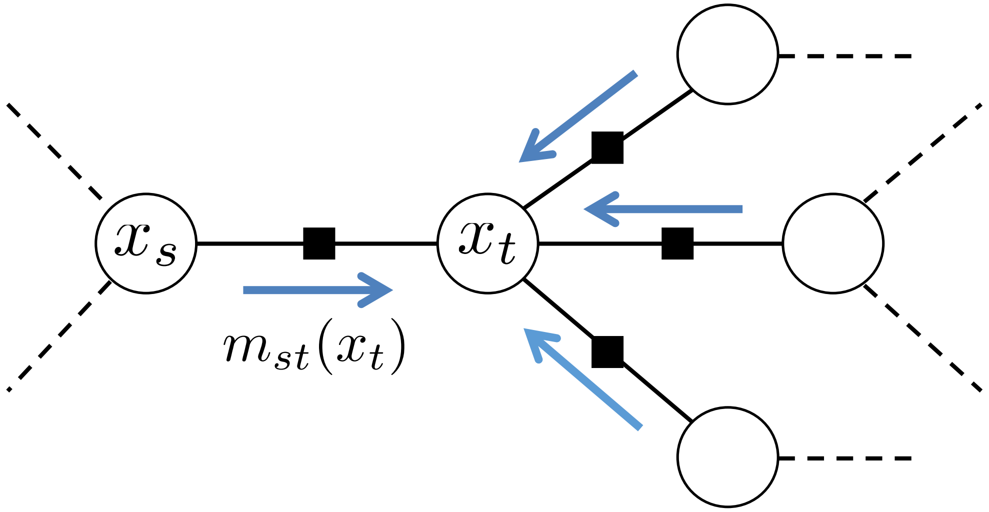
Canonical Parameterization (Information Parameters):

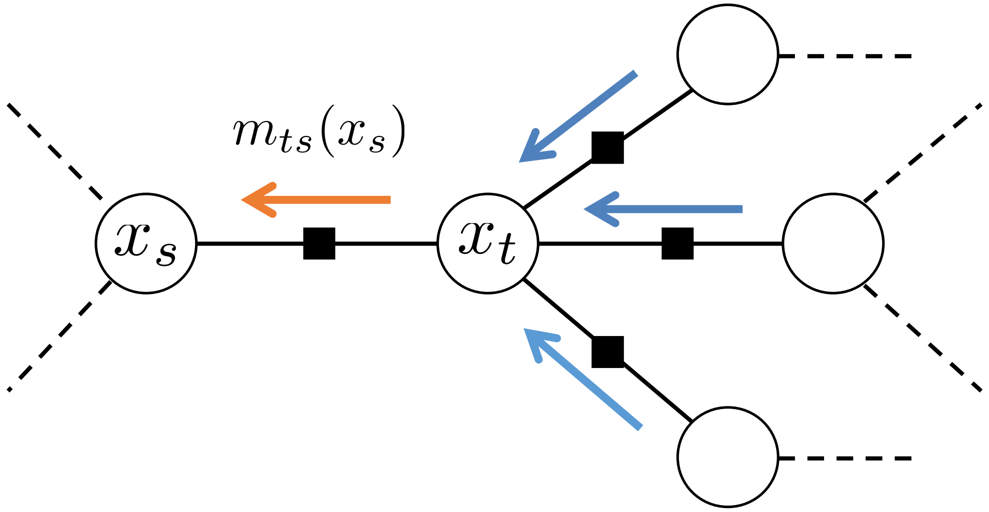
$$\mathcal{N}(x \mid \mu, \Sigma) \propto \exp\left\{-\frac{1}{2} x^T \Sigma^{-1} x + \mu^T \Sigma^{-1} x - \frac{1}{2} \mu^T \Sigma^{-1} \mu\right\}$$

$$\mathcal{N}^{-1}(x \mid \vartheta, \Lambda) \propto \exp\left\{-\frac{1}{2} x^T \Lambda x + \vartheta^T x\right\}$$ where $$\Lambda = \Sigma^{-1} \quad \vartheta = \Sigma^{-1}\mu$$

Recall general exponential family form: $$p(x \mid \theta) = \exp\{\theta^T \phi(x) - \Phi(\theta)\}$$

Gaussian Belief Propagation





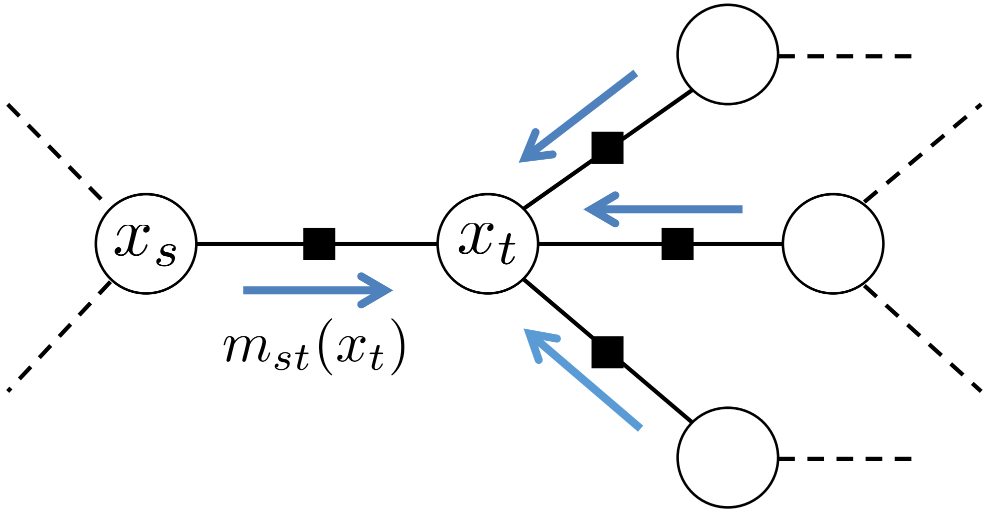
$$p_t(x_t) \propto \prod_{s \in \Gamma(t)} m_{st}(x_t) \qquad m_{ts}(x_s) = \int_{\mathcal{X}_t} \psi_{st}(x_s, x_t) \prod_{u \in \Gamma(t) \setminus s} m_{ut}(x_t)\, dx_t$$

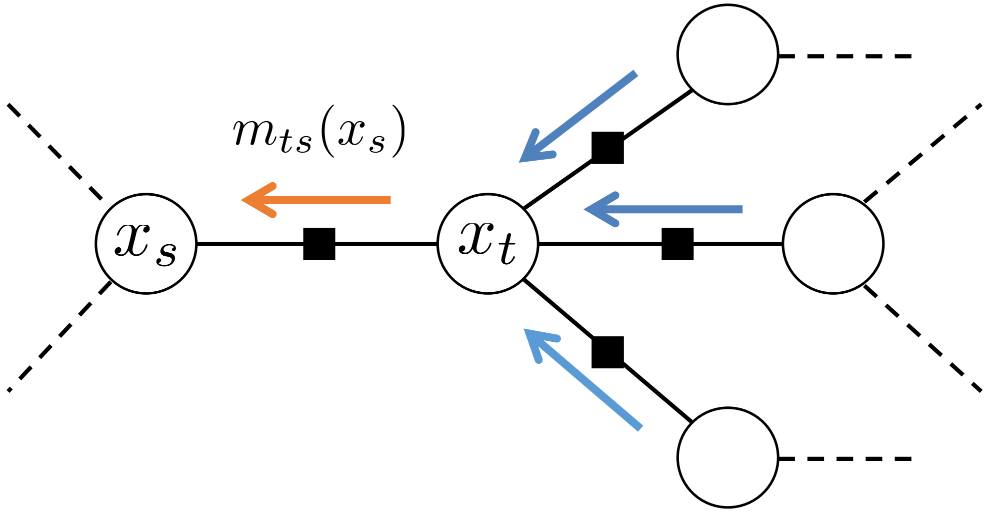
$$p_t(x_t) = \mathcal{N}^{-1}(x_t \mid \vartheta_t, \Lambda_t) \qquad m_{ts}(x_s) \propto \mathcal{N}^{-1}(x_s \mid \vartheta_{ts}, \Lambda_{ts})$$

Computing *marginal mean and covariance* from messages:

$$\Lambda_t = \sum_{s \in \Gamma(t)} \Lambda_{st} \qquad \vartheta_t = \sum_{s \in \Gamma(t)} \vartheta_{st}$$

Gaussian Belief Propagation





- Compute message mean & covar as algebraic function of incoming message mean & covar (generalizes Kalman)
- For tree of *N* nodes of dimension *d*, cost is $O(Nd^3)$

Computing *marginal mean and covariance* from messages:

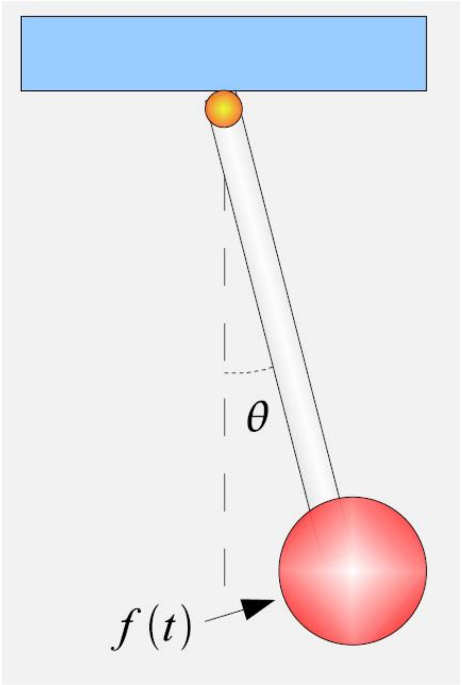
$$\Lambda_t = \sum_{s \in \Gamma(t)} \Lambda_{st} \qquad \vartheta_t = \sum_{s \in \Gamma(t)} \vartheta_{st}$$

Nonlinear Dynamical System

State dynamics and/or measurement may be nonlinear:

$$x_t = f(x_{t-1}) + \epsilon \sim \mathcal{N}(0, \Sigma)$$

$$y_t = h(x_t) + \omega \sim \mathcal{N}(0, R)$$



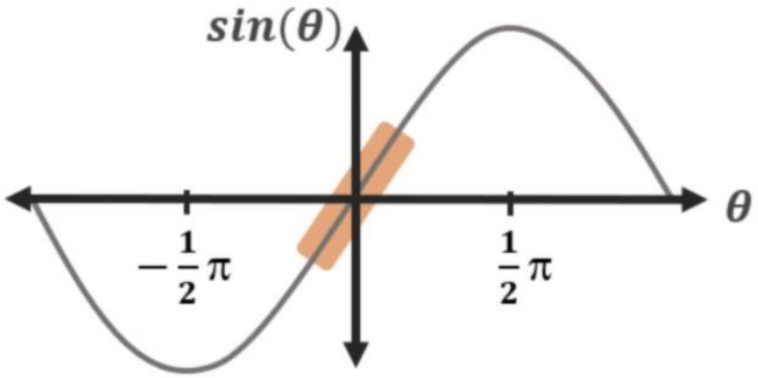
Filter equations lack a closed-form:

Prediction:

$$p(x_t \mid y_1^{t-1}) = \int \mathcal{N}(x_t \mid f(x_{t-1}), \Sigma) p(x_{t-1} \mid y_1^{t-1}) \, dx_t$$

Measurement Update:

$$p(x_t \mid y_1^t) \propto \mathcal{N}(y_t \mid h(x_t), R) p(x_t \mid y_1^{t-1})$$



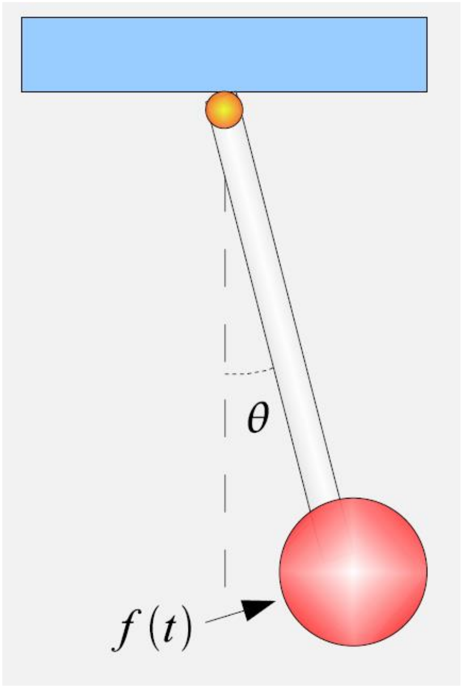
Extended Kalman Filter

Linearize f(.) and h(.) about a point m:

$$f(x) \approx f(m) + \mathbf{J}_f(m)(x - m)$$

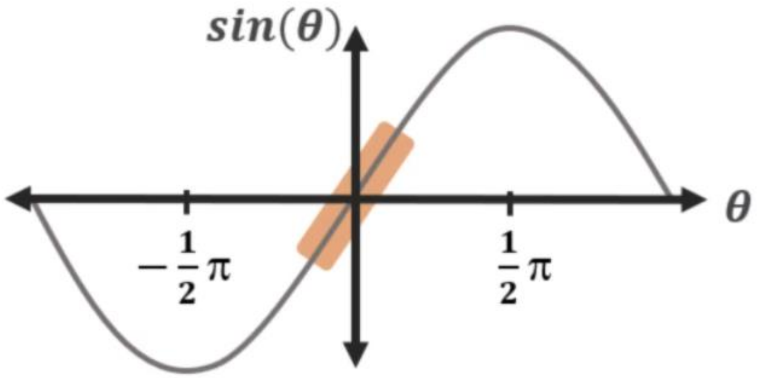
$$h(x) \approx h(m) + \mathbf{J}_h(m)(x - m)$$

where $\mathbf{J}_f = \left(\frac{\partial f_i}{\partial x_j}\right)$ is Jacobian matrix of partials



Assume *approximately* Gaussian marginals:

$$p(x_{t-1} \mid y_1^{t-1}) \approx \mathcal{N}(\mu_{t-1|t-1}, \Sigma_{t-1|t-1})$$



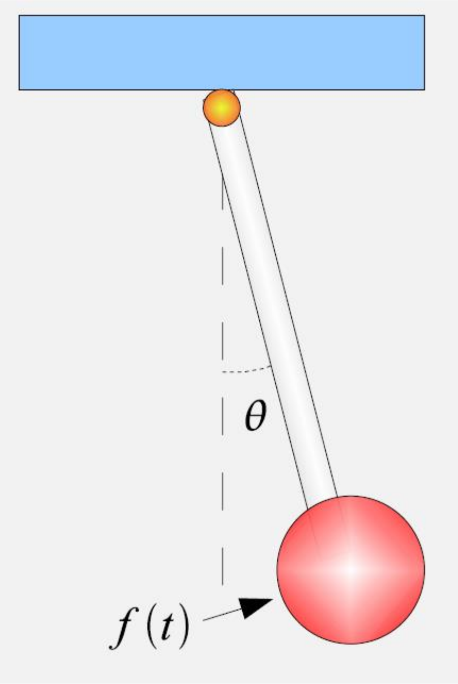
Filter equations:

- Remain (approximately) Gaussian
- Nearly identical to standard Kalman

Nonlinear Dynamical System

Pendulum with mass m=1,pole length L=1:

$$x_t = \begin{pmatrix} \theta_t \\ \dot{\theta}_t \end{pmatrix} = \underbrace{\begin{pmatrix} \theta_{t-1} + \dot{\theta}_{t-1} \\ \dot{\theta}_{t-1} - g\sin(\theta_{t-1}) \end{pmatrix}}_{f(x_{t-1})} + \epsilon$$

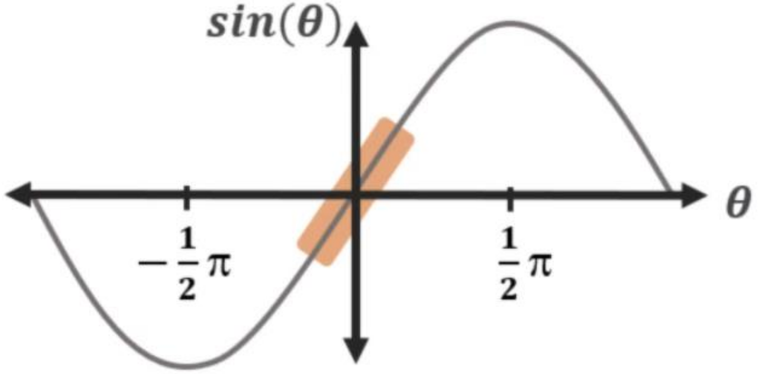


Noisy observation of X-position:

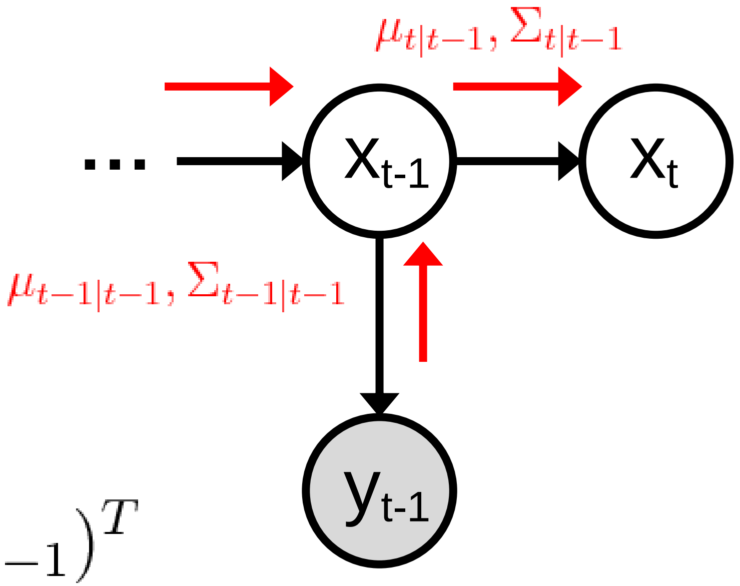
$$y_t = \underbrace{\sin(\theta_t)}_{h(x_t)} + \omega$$

Jacobian terms for linearization:

$$\mathbf{J}_f(x) = \begin{pmatrix} 1 & 1 \\ -g\cos(x^1) & 1 \end{pmatrix} \qquad \mathbf{J}_h(x) = (\cos(x^1), 0)$$



EKF Update Equations

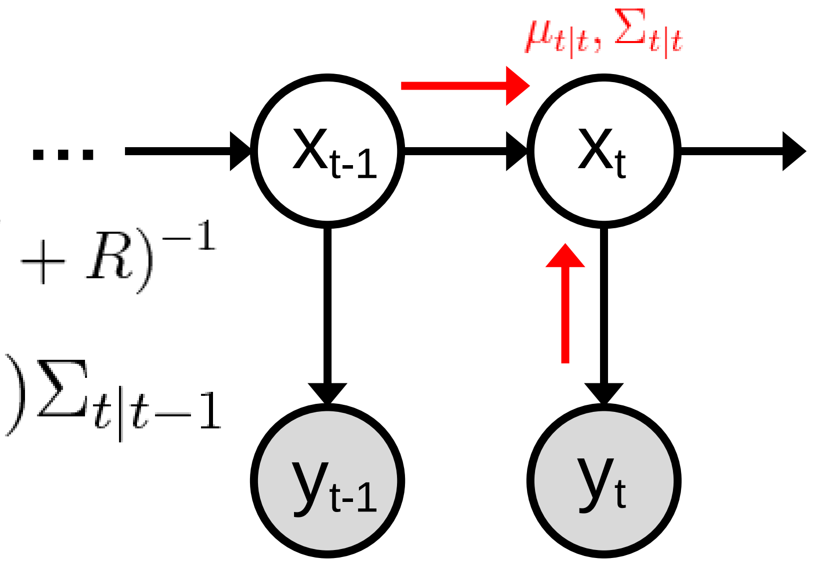


Prediction Step:

State Prediction: $\mu_{t|t-1} = f(\mu_{t-1|t-1})$

Covariance Prediction:

$$\Sigma_{t|t-1} = \Sigma + \mathbf{J}_f(\mu_{t-1|t-1})\Sigma_{t-1|t-1}\mathbf{J}_f(\mu_{t-1|t-1})^T$$



Measurement Update Step:

Gain: $K_t = \Sigma_{t|t-1}\mathbf{J}_h(\mu_{t|t-1})^T(\mathbf{J}_h(\mu_{t|t-1})\Sigma_{t|t-1}\mathbf{J}_h(\mu_{t|t-1})^T + R)^{-1}$

Filter Covariance: $\Sigma_{t|t} = \Sigma_{t|t-1} - K_t\mathbf{J}_h(\mu_{t|t-1})\Sigma_{t|t-1}$

Filter Mean: $\mu_{t|t} = \mu_{t|t-1} + K_t(y_t - h(\mu_{t|t-1}))$

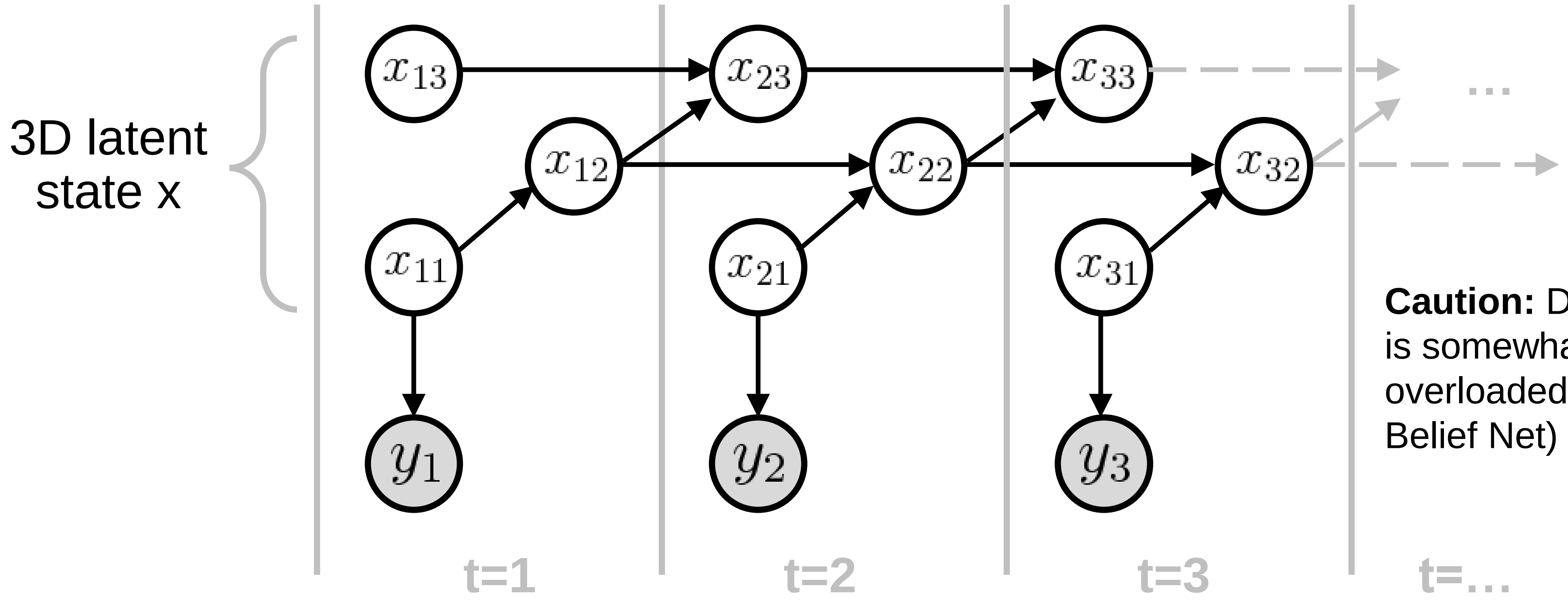
Extended Kalman Filter

- **PROS:**
  - Easy to implement – updates analogous to standard Kalman
  - Computationally efficient
  - Known theoretical stability results

- **CONS:**
  - Linearity assumption poor for highly nonlinear models
  - Requires differentiability
  - Jacobian matrices can be hard to calculate & implement

Unscented Kalman filter (UKF) typically more accurate in practice

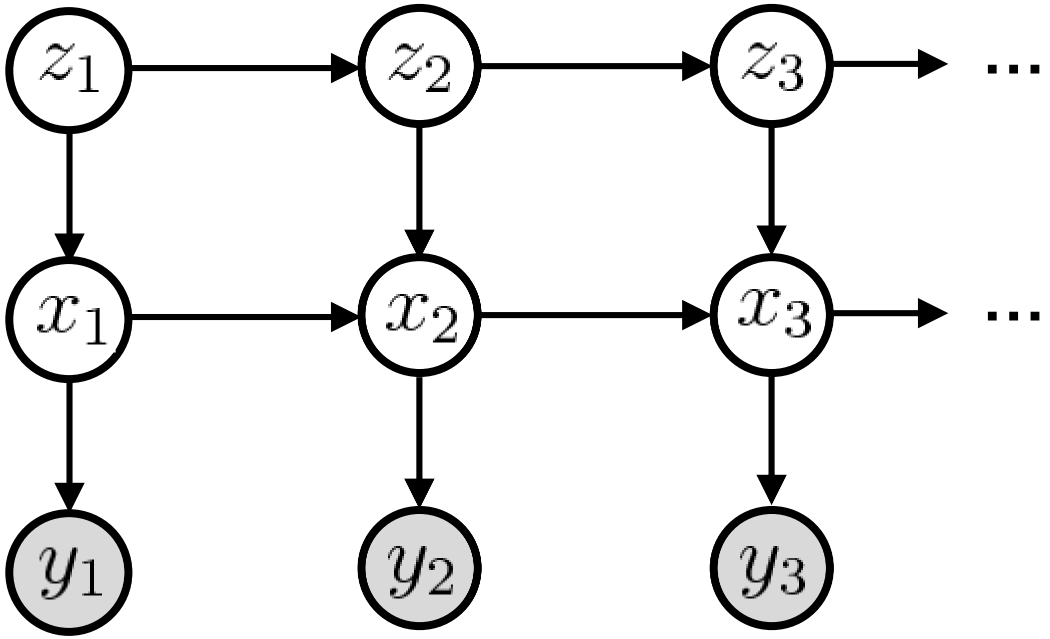
Dynamic Bayesian Networks

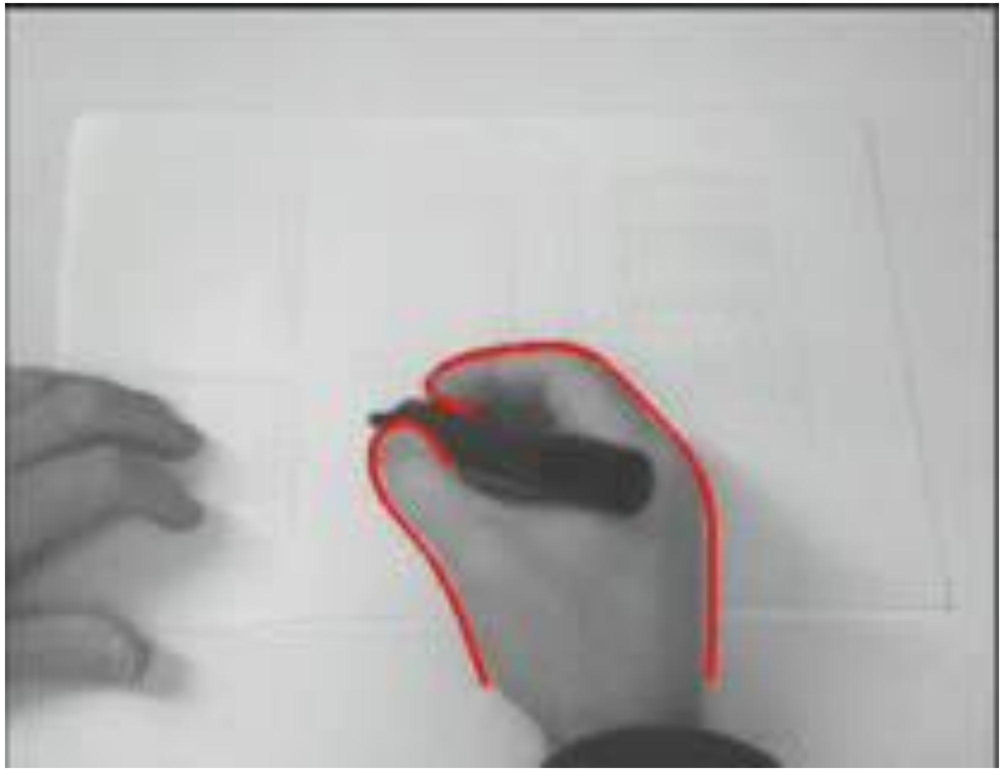


Caution: DBN terminology is somewhat vague and overloaded (e.g. Deep Belief Net)

- Multivariate latent state (e.g. $x \in \mathbb{R}^3$)
- Dynamics for each component within and across time
- Sometimes used as catch-all term for dynamical systems

Switching Dynamical System





Colors indicate 3 writing modes

[Video: Isard & Blake, ICCV 1998.]

Discrete switching state:

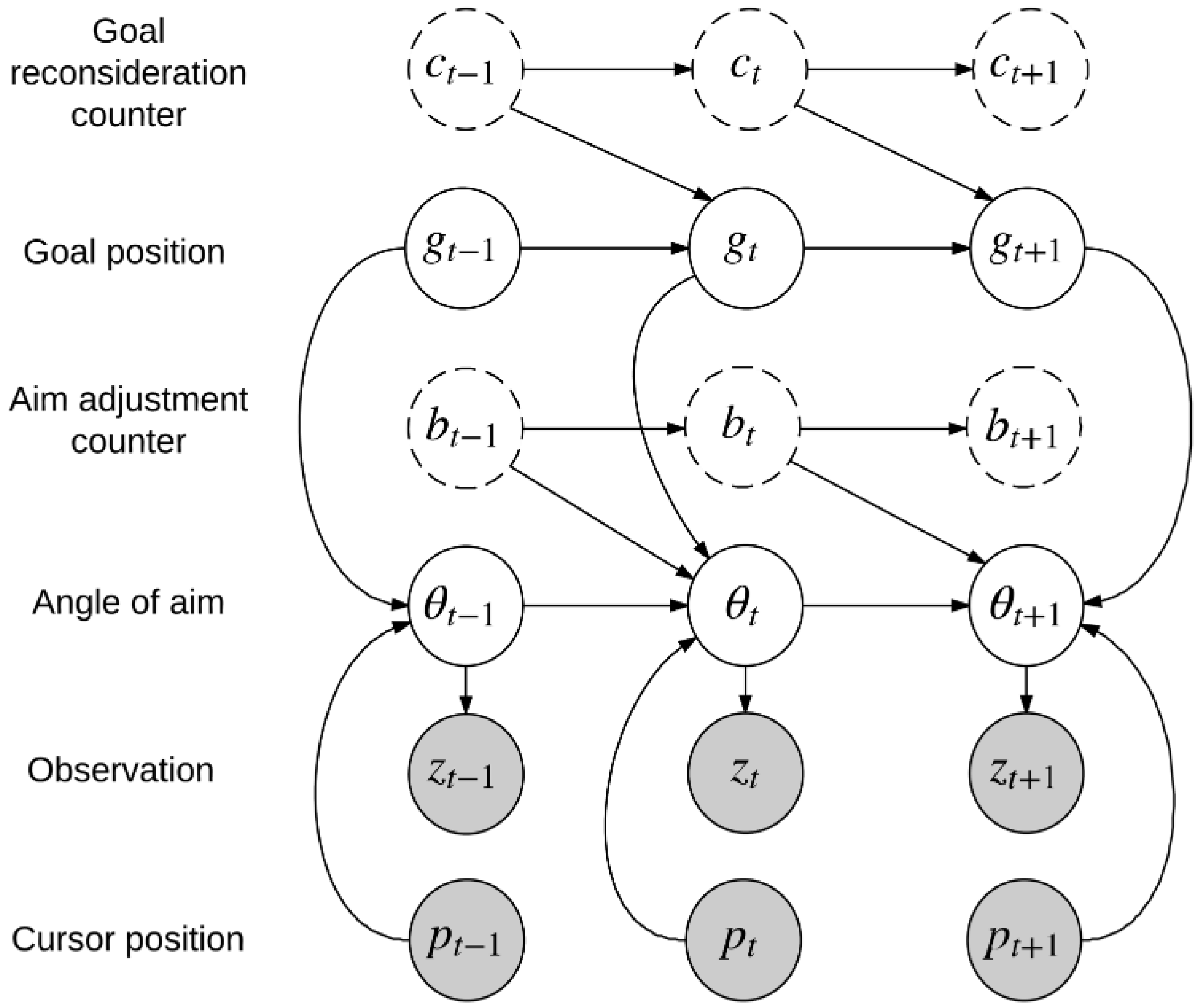
$z_t \mid z_{t-1} \sim \mathrm{Cat}(\pi(z_{t-1}))$ With stochastic transition matrix π

Switching state selects dynamics:

$x_t \mid x_{t-1} \sim \mathcal{N}(A_{z_t} x_{t-1}, \Sigma_{z_t})$ (e.g. Linear Gaussian)

Semi-Markov Process

Intracortical Brain-Computer Interface



Block 12: "Multiscale Semi-Markov Model"

- Counter decrements deterministically:
  $c_t = c_{t-1} - 1 \quad \text{if} \quad c_{t-1} > 0$
- Resample when exhausted:
  $c_t \sim \text{Some-PMF}(\cdot) \quad \text{if} \quad c_{t-1} = 0$
- Controls dynamics: $g_t \mid c_{t-1} \sim p_{c_{t-1}}(\cdot)$

Summary

- Definition of state-space model / dynamical system

- Analytic inference for Gaussian LDS possible via Kalman

- Gaussian nonlinear DS requires approximations (EKF / UKF)

- More complex dynamical structure:
  - Switching state-space model
  - Dynamic Bayesian Networks
  - Semi-Markov Processes