Probabilistic Graphical Models

Brown University CSCI 2950-P, Spring 2013
Prof. Erik Sudderth

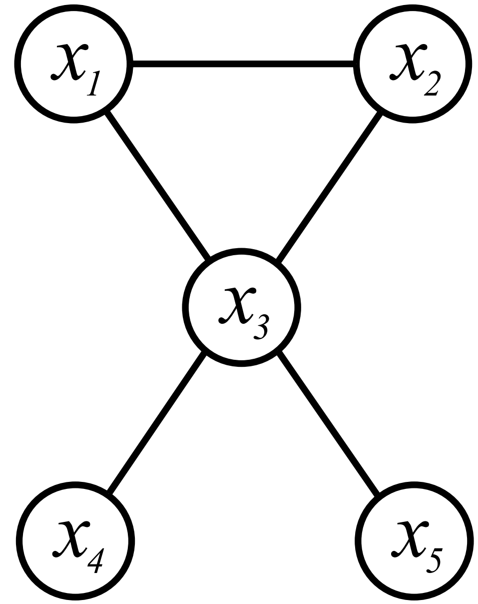
Lecture 12: Gaussian Belief Propagation, State Space Models and Kalman Filters

Guest Kalman Filter Lecture by Jason Pacheco

Some figures courtesy Michael Jordan's draft textbook,
An Introduction to Probabilistic Graphical Models

Pairwise Markov Random Fields

$$p(x) = \frac{1}{Z} \prod_{(s,t)\in\mathcal{E}} \psi_{st}(x_s, x_t) \prod_{s\in\mathcal{V}} \psi_s(x_s)$$



- Simple parameterization, but still expressive and widely used in practice
- Guaranteed Markov with respect to graph
- Any jointly Gaussian distribution can be represented by only *pairwise* potentials

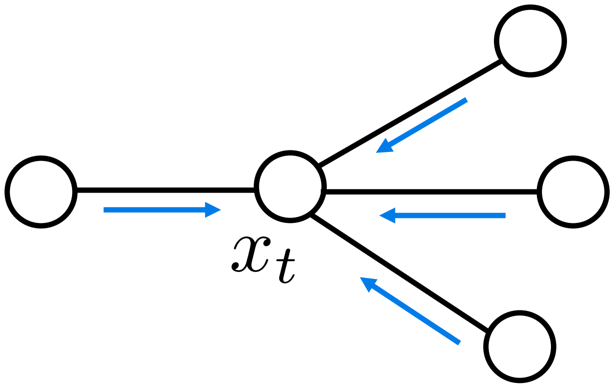
$\mathcal{E} \longrightarrow$ set of undirected edges (s,t) linking pairs of nodes

$\mathcal{V} \longrightarrow$ set of N nodes or vertices, $\{1, 2, \ldots, N\}$

$Z \longrightarrow$ normalization constant (partition function)

Belief Propagation (Integral-Product)

BELIEFS: Posterior marginals

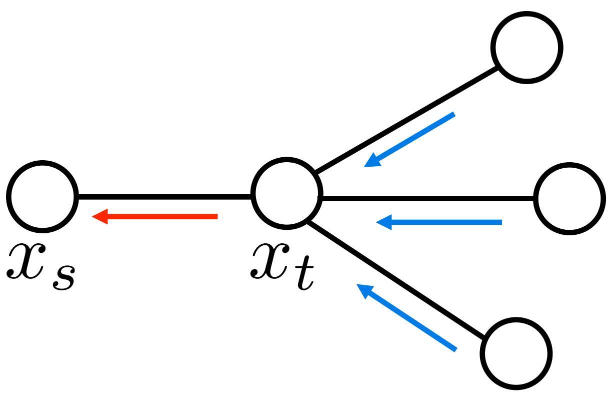


$$\hat{p}_t(x_t) \propto \psi_t(x_t) \prod_{u \in \Gamma(t)} m_{ut}(x_t)$$

$\Gamma(t) \longrightarrow$ neighborhood of node t
(adjacent nodes)

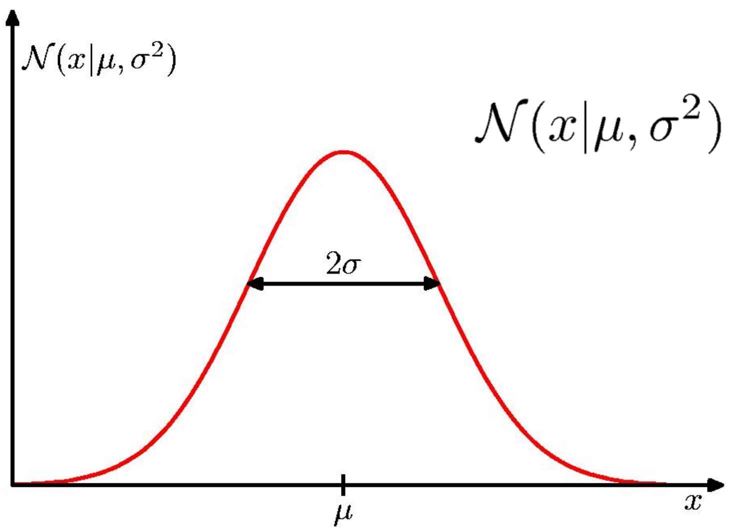
MESSAGES: Sufficient statistics

$$m_{ts}(x_s) \propto \int_{x_t} \psi_{st}(x_s, x_t) \psi_t(x_t) \prod_{u \in \Gamma(t) \setminus s} m_{ut}(x_t)$$

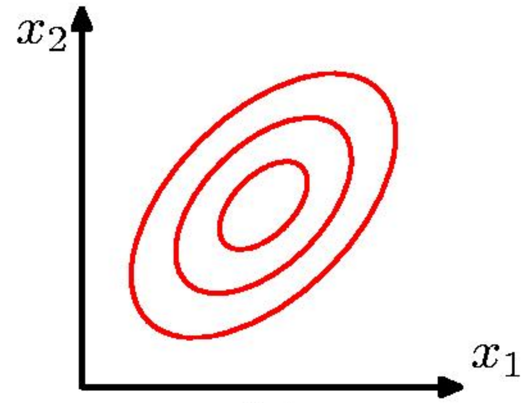


I) Message Product
II) Message Propagation

Gaussian Distributions



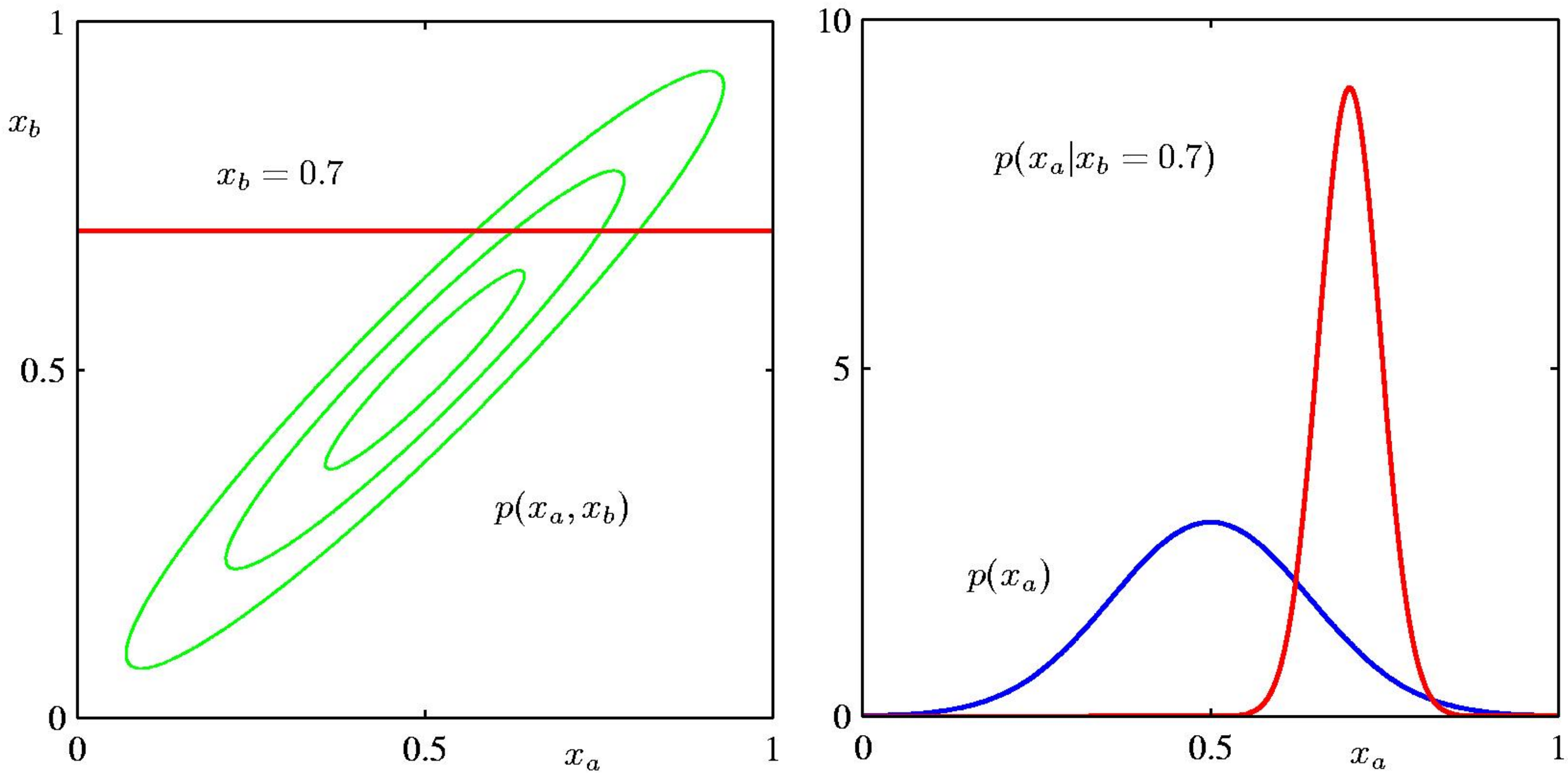
$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



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

- Simplest joint distribution that can capture arbitrary mean & covariance
- Justifications from *central limit theorem* and *maximum entropy* criterion
- Probability density above assumes covariance is *positive definite*
- ML parameter estimates are *sample mean* & *sample covariance*

Gaussian Conditionals & Marginals



For any joint multivariate Gaussian distribution, all marginal distributions are Gaussians, and all conditional distributions are Gaussians

Partitioned Gaussian Distributions

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix}$$

Marginals:

$$\begin{aligned} p(\mathbf{x}_1) &= \mathcal{N}(\mathbf{x}_1|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \\ p(\mathbf{x}_2) &= \mathcal{N}(\mathbf{x}_2|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}) \end{aligned}$$

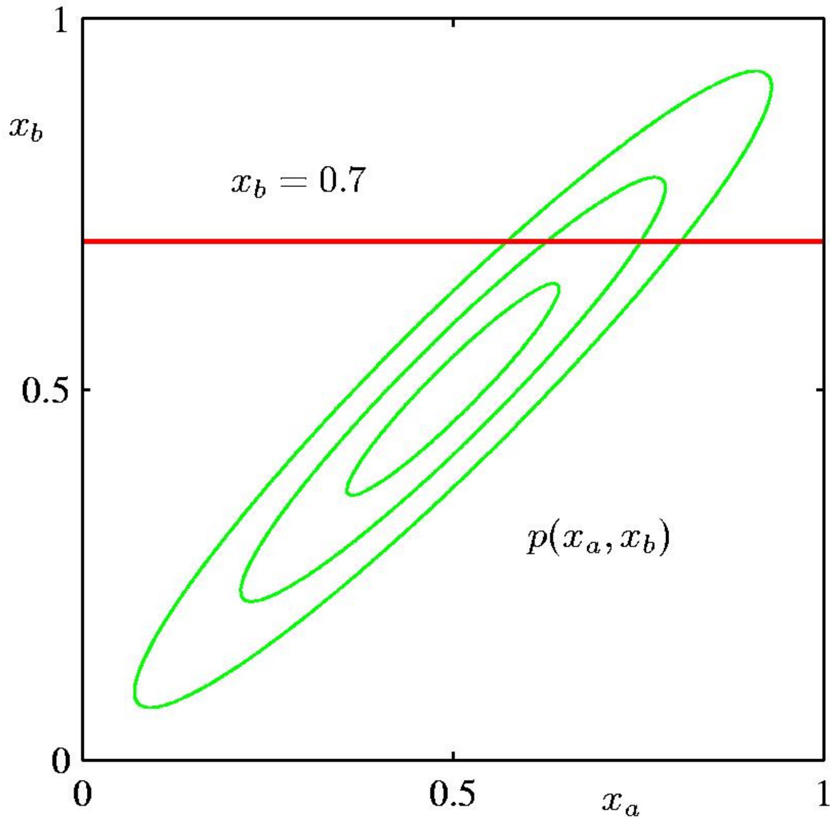
Conditionals:

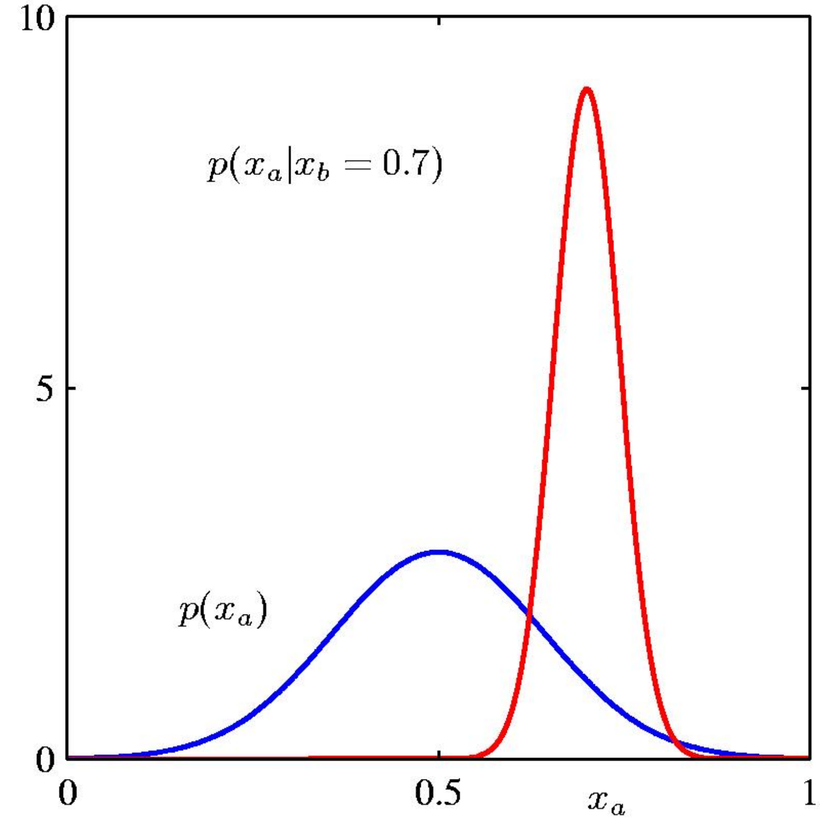
$$\begin{aligned} p(\mathbf{x}_1|\mathbf{x}_2) &= \mathcal{N}(\mathbf{x}_1|\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}) \\ \boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= \boldsymbol{\mu}_1 - \boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= \boldsymbol{\Sigma}_{1|2}\left(\boldsymbol{\Lambda}_{11}\boldsymbol{\mu}_1 - \boldsymbol{\Lambda}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right) \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \boldsymbol{\Lambda}_{11}^{-1} \end{aligned}$$

Intuition:

$$p(x_1 \mid x_2) = \frac{p(x_1, x_2)}{p(x_2)} \propto \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) + \frac{1}{2}(x_2-\mu_2)^T\Sigma_{22}^{-1}(x_2-\mu_2)\right\}$$

Gaussian Conditionals & Marginals



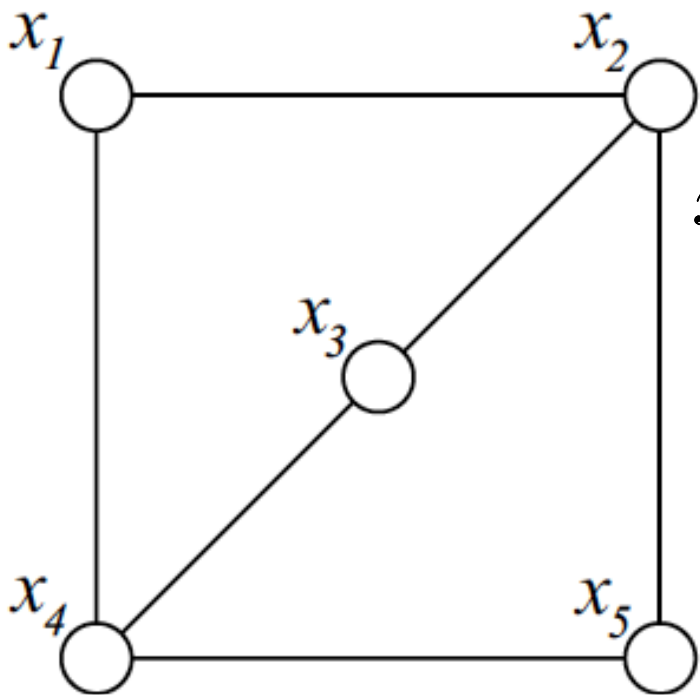


$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$p(x_1) = \mathcal{N}(x_1|\mu_1, \sigma_1^2)$$

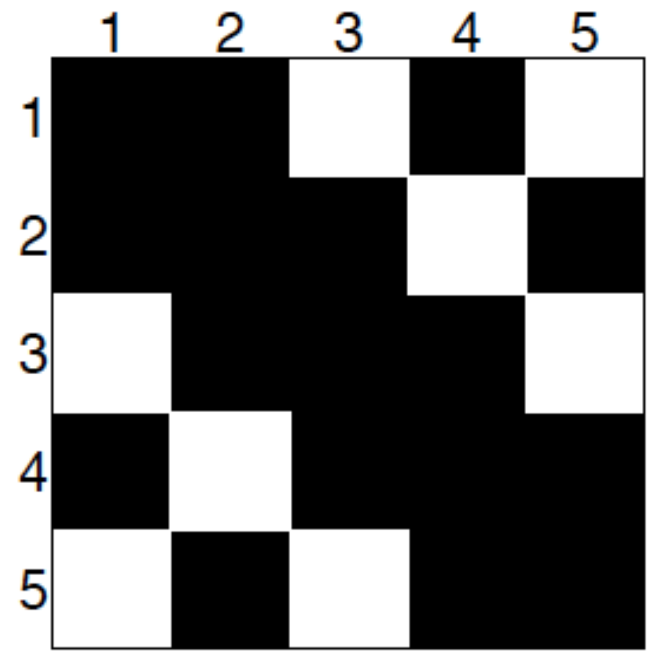
$$p(x_1 \mid x_2) = \mathcal{N}\left(x_1 \mid \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right)$$

Gaussian Graphical Models



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

$$J = \Sigma^{-1}$$



$$p(x) = \frac{1}{Z} \prod_{(s,t) \in \mathcal{E}} \psi_{s,t}(x_s, x_t)$$

$$\sum_{t \in N(s)} J_{s(t)} = J_{s,s}$$

$$\psi_{s,t}(x_s, x_t) = \exp\left\{ -\frac{1}{2} \begin{bmatrix} x_s^T & x_t^T \end{bmatrix} \begin{bmatrix} J_{s(t)} & J_{s,t} \\ J_{t,s} & J_{t(s)} \end{bmatrix} \begin{bmatrix} x_s \\ x_t \end{bmatrix} \right\}$$

Gaussian Potentials

$$p(x) = \frac{1}{Z}\exp\left\{-\frac{1}{2}x^T P^{-1} x\right\} = \frac{1}{Z}\prod_{s=1}^{N}\prod_{t=1}^{N}\exp\left\{-\frac{1}{2}x_s^T J_{s,t} x_t\right\} =$$

$$\frac{1}{Z}\prod_{(s,t)\in\mathcal{E}}\exp\left\{-\frac{1}{2}\begin{bmatrix}x_s^T & x_t^T\end{bmatrix}\begin{bmatrix}J_{s(t)} & J_{s,t}\\ J_{t,s} & J_{t(s)}\end{bmatrix}\begin{bmatrix}x_s\\ x_t\end{bmatrix}\right\} = \frac{1}{Z}\prod_{(s,t)\in\mathcal{E}}\psi_{s,t}(x_s, x_t)$$

$$Z = \left((2\pi)^N \det P\right)^{1/2}$$

$$p(x) = \frac{1}{Z}\prod_{(s,t)\in\mathcal{E}}\psi_{s,t}(x_s, x_t)$$

$$\sum_{t\in N(s)} J_{s(t)} = J_{s,s}$$

$$\psi_{s,t}(x_s, x_t) = \exp\left\{-\frac{1}{2}\begin{bmatrix}x_s^T & x_t^T\end{bmatrix}\begin{bmatrix}J_{s(t)} & J_{s,t}\\ J_{t,s} & J_{t(s)}\end{bmatrix}\begin{bmatrix}x_s\\ x_t\end{bmatrix}\right\}$$

Interpreting GMRF Parameters

$$p(x) = \frac{1}{Z}\exp\left\{-\frac{1}{2}x^T P^{-1} x\right\} = \frac{1}{Z}\prod_{s=1}^{N}\prod_{t=1}^{N}\exp\left\{-\frac{1}{2}x_s^T J_{s,t} x_t\right\} =$$

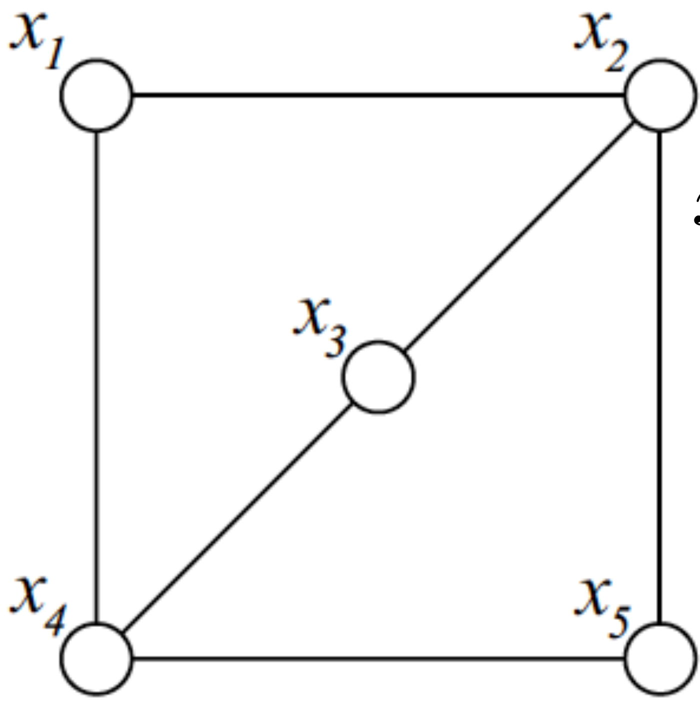
$$\frac{1}{Z}\prod_{(s,t)\in\mathcal{E}}\exp\left\{-\frac{1}{2}\begin{bmatrix} x_s^T & x_t^T \end{bmatrix}\begin{bmatrix} J_{s(t)} & J_{s,t} \\ J_{t,s} & J_{t(s)} \end{bmatrix}\begin{bmatrix} x_s \\ x_t \end{bmatrix}\right\} = \frac{1}{Z}\prod_{(s,t)\in\mathcal{E}}\psi_{s,t}\left(x_s, x_t\right)$$

$$Z = \left((2\pi)^N \det P\right)^{1/2}$$

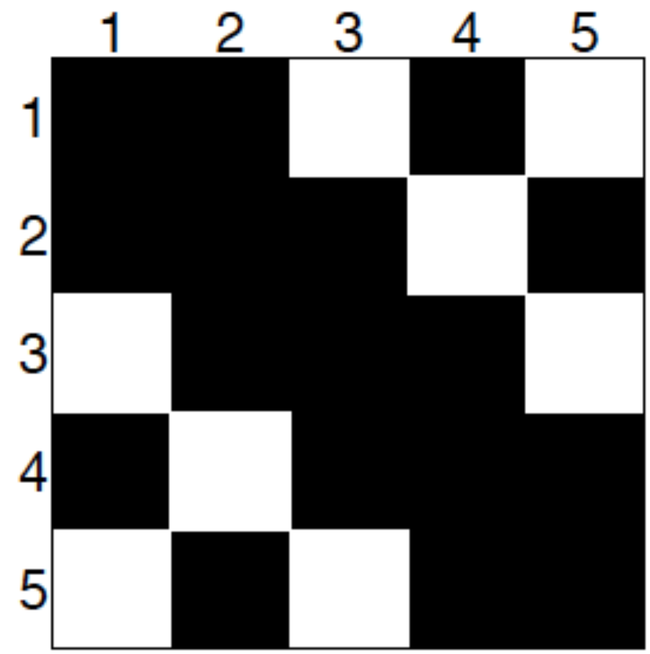
$$\mathrm{var}\left(x_s \left| x_{N(s)}\right.\right) = \left(J_{s,s}\right)^{-1}$$

$$\rho_{st|N(s,t)} \triangleq \frac{\mathrm{cov}\left(x_s, x_t \left| x_{N(s,t)}\right.\right)}{\sqrt{\mathrm{var}\left(x_s \left| x_{N(s,t)}\right.\right)\mathrm{var}\left(x_t \left| x_{N(s,t)}\right.\right)}} = \frac{-J_{s,t}}{\sqrt{J_{s,s}J_{t,t}}}$$

Gaussian Markov Properties

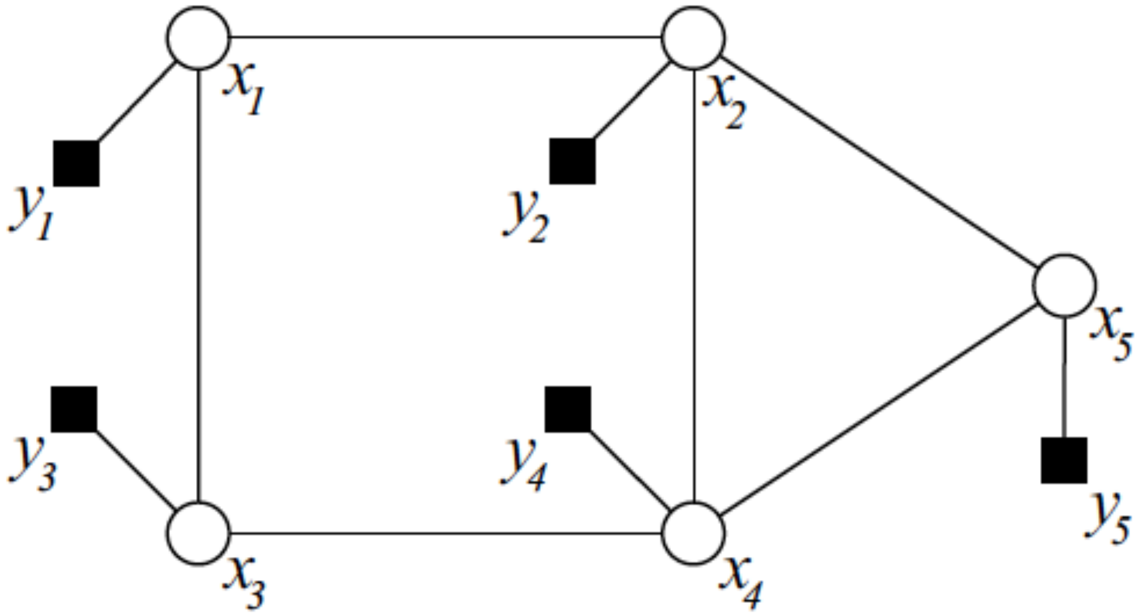


$$x \sim \mathcal{N}(\mu, \Sigma)$$
$$J = \Sigma^{-1}$$



Theorem 2.2. Let $x \sim \mathcal{N}(0, P)$ be a Gaussian stochastic process which is Markov with respect to an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Assume that x is *not* Markov with respect to any $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ such that $\mathcal{E}' \subsetneq \mathcal{E}$, and partition $J = P^{-1}$ into a $|\mathcal{V}| \times |\mathcal{V}|$ grid according to the dimensions of the node variables. Then for any $s, t \in \mathcal{V}$ such that $s \neq t$, $J_{s,t} = J_{t,s}^T$ will be nonzero if and only if $(s, t) \in \mathcal{E}$.

Inference with Gaussian Observations



$$y_s = C_s x_s + v_s$$

$$v_s \sim \mathcal{N}(0, R_s)$$

$$p(y \mid x) = \prod_{s=1}^{N} p(y_s \mid x_s)$$

$$C = \mathrm{diag}(C_1, C_2, \ldots, C_N)$$

$$R = \mathrm{diag}(R_1, R_2, \ldots, R_N)$$

$$p(x_s \mid y) \sim \mathcal{N}\left(\widehat{x}_s, \widehat{P}_s\right)$$

Linear Gaussian Systems

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \qquad p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \boldsymbol{\Sigma}_y)$$

Marginal Likelihood:

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu}_x + \mathbf{b}, \boldsymbol{\Sigma}_y + \mathbf{A}\boldsymbol{\Sigma}_x\mathbf{A}^T)$$

Posterior Distribution:

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{x|y}, \boldsymbol{\Sigma}_{x|y})$$

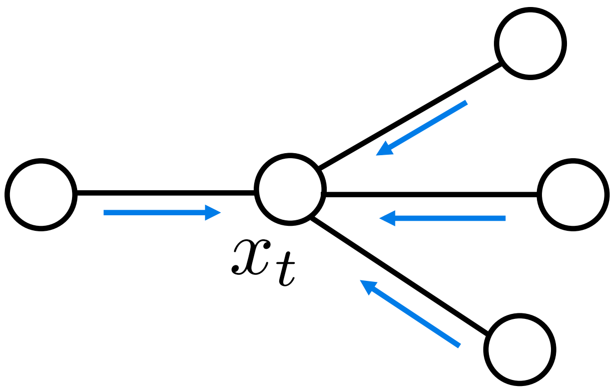
$$\boldsymbol{\Sigma}_{x|y}^{-1} = \boldsymbol{\Sigma}_x^{-1} + \mathbf{A}^T\boldsymbol{\Sigma}_y^{-1}\mathbf{A}$$

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\Sigma}_{x|y}[\mathbf{A}^T\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Sigma}_x^{-1}\boldsymbol{\mu}_x]$$

Gaussian BP: Complexity linear, not cubic, in number of nodes

Belief Propagation (Integral-Product)

BELIEFS: Posterior marginals

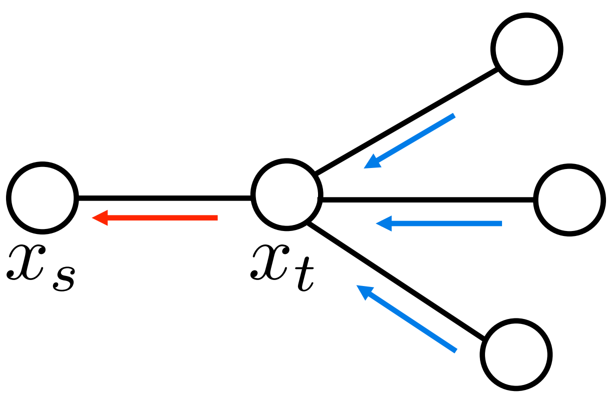


$$\hat{p}_t(x_t) \propto \psi_t(x_t) \prod_{u \in \Gamma(t)} m_{ut}(x_t)$$

$\Gamma(t) \longrightarrow$ neighborhood of node t (adjacent nodes)

MESSAGES: Sufficient statistics

$$m_{ts}(x_s) \propto \int_{x_t} \psi_{st}(x_s, x_t) \psi_t(x_t) \prod_{u \in \Gamma(t) \setminus s} m_{ut}(x_t)$$



I) Message Product

II) Message Propagation

Gaussian Belief Propagation

- The natural, canonical, or information parameterization of a Gaussian distribution arises from quadratic form:

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

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

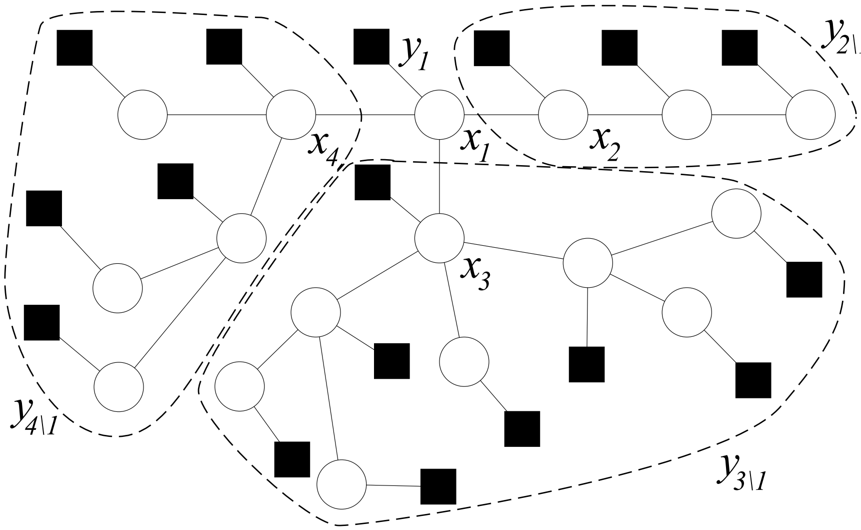
$$\vartheta = \Sigma^{-1}\mu$$
$$\Lambda = \Sigma^{-1}$$

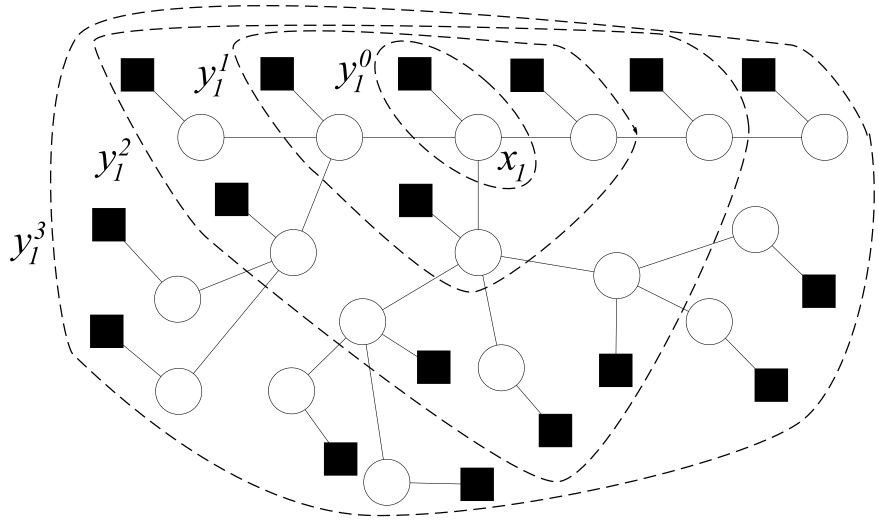
- Gaussian BP represents messages and marginals as:

$$m_{ts}(x_s) = \alpha \mathcal{N}^{-1}(\vartheta_{ts}, \Lambda_{ts})$$

$$p(x_s \mid y) = \mathcal{N}^{-1}(\vartheta_s, \Lambda_s)$$

Gaussian Belief Propagation





- Gaussian BP represents messages and marginals as:

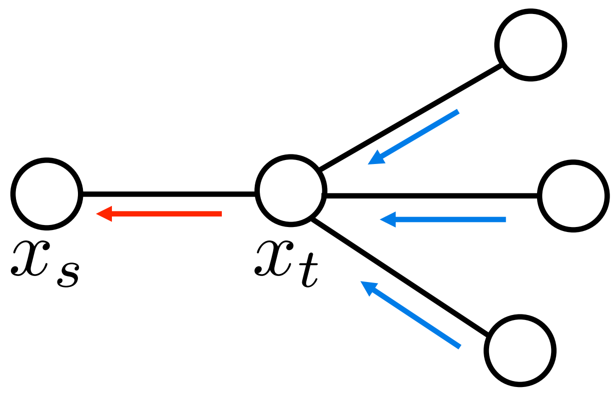
$$m_{ts}(x_s) = \alpha \mathcal{N}^{-1}(\vartheta_{ts}, \Lambda_{ts}) \qquad p(x_s \mid y) = \mathcal{N}^{-1}(\vartheta_s, \Lambda_s)$$

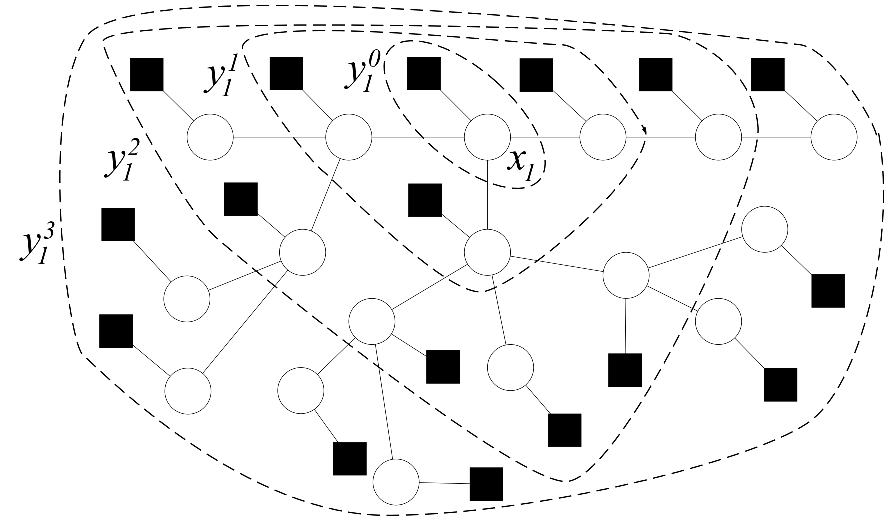
- Gaussian BP belief updates then have a simple form:

$$p(x_s \mid y_s^n) = \alpha p(y_s \mid x_s) \prod_{t \in N(s)} m_{ts}^n(x_s) \qquad \vartheta_s^n = C_s^T R_s^{-1} y_s + \sum_{t \in N(s)} \vartheta_{ts}^n$$

$$p(y_s \mid x_s) = \alpha \mathcal{N}^{-1}\left(C_s^T R_s^{-1} y_s, C_s^T R_s^{-1} C_s\right) \qquad \Lambda_s^n = C_s^T R_s^{-1} C_s + \sum_{t \in N(s)} \Lambda_{ts}^n$$

Gaussian Belief Propagation



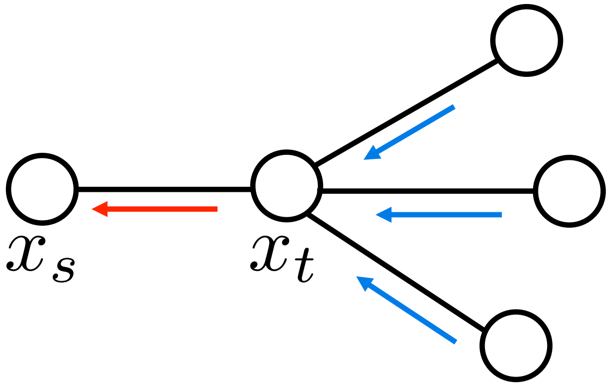


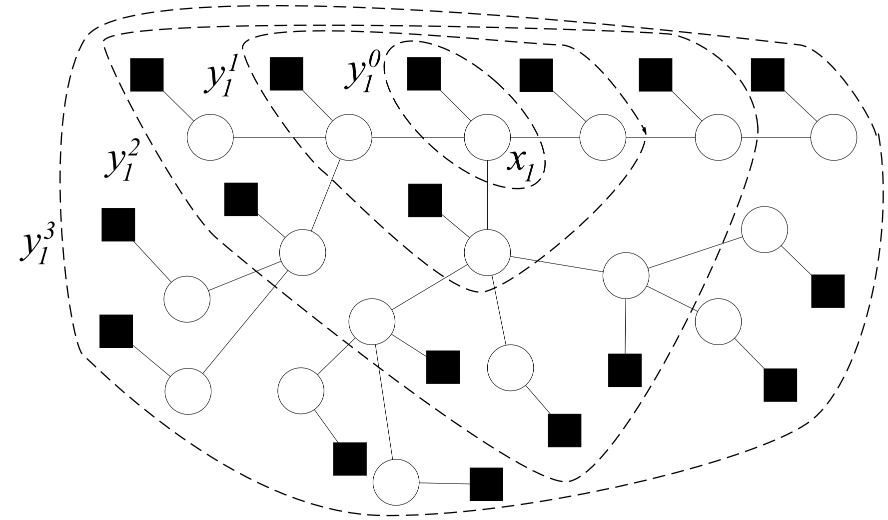
$$m_{ts}^{n}(x_s) = \alpha \int_{x_t} \psi_{s,t}(x_s, x_t)\, p(y_t \mid x_t) \prod_{u \in N(t) \setminus s} m_{ut}^{n-1}(x_t)\, dx_t$$

$$\psi_{s,t}(x_s, x_t)\, p(y_t \mid x_t) \prod_{u \in N(t) \setminus s} m_{ut}^{n-1}(x_t) \propto \mathcal{N}^{-1}(\overline{\vartheta}, \overline{\Lambda})$$

$$\overline{\vartheta} = \begin{bmatrix} \mathbf{0} \\ C_t^T R_t^{-1} y_t + \sum_{u \in N(t) \setminus s} \vartheta_{ut}^{n-1} \end{bmatrix} \quad \overline{\Lambda} = \begin{bmatrix} J_{s(t)} & J_{s,t} \\ J_{t,s} & J_{t(s)} + C_t^T R_t^{-1} C_t + \sum_{u \in N(t) \setminus s} \Lambda_{ut}^{n-1} \end{bmatrix}$$

Gaussian Belief Propagation





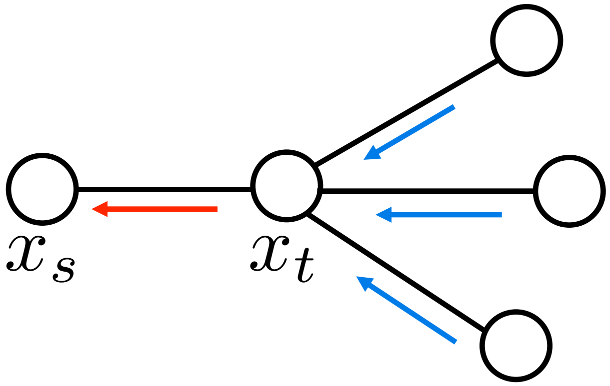
Matrix Inversion Lemma:

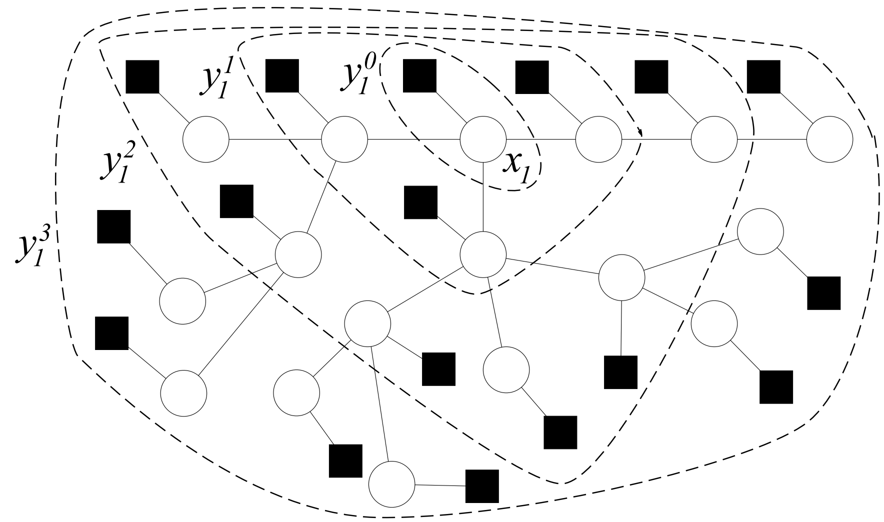
$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

$$M^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1} BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1} BD^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1} CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1} CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

Gaussian Belief Propagation



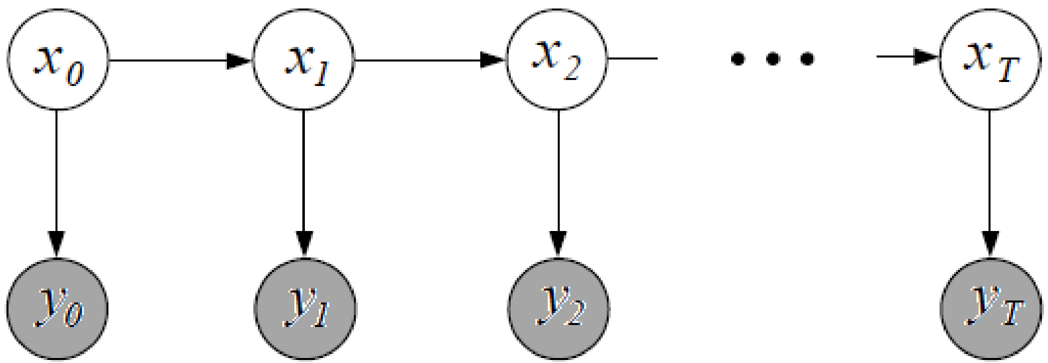


$$m_{ts}^{n}(x_s) = \alpha \int_{x_t} \psi_{s,t}(x_s, x_t)\, p(y_t \mid x_t) \prod_{u \in N(t) \setminus s} m_{ut}^{n-1}(x_t)\, dx_t$$

$$\vartheta_{ts}^{n} = -J_{s,t} \left(J_{t(s)} + C_t^T R_t^{-1} C_t + \sum_{u \in N(t) \setminus s} \Lambda_{ut}^{n-1} \right)^{-1} \left(C_t^T R_t^{-1} y_t + \sum_{u \in N(t) \setminus s} \vartheta_{ut}^{n-1} \right)$$

$$\Lambda_{ts}^{n} = J_{s(t)} - J_{s,t} \left(J_{t(s)} + C_t^T R_t^{-1} C_t + \sum_{u \in N(t) \setminus s} \Lambda_{ut}^{n-1} \right)^{-1} J_{t,s}$$

State Space Models



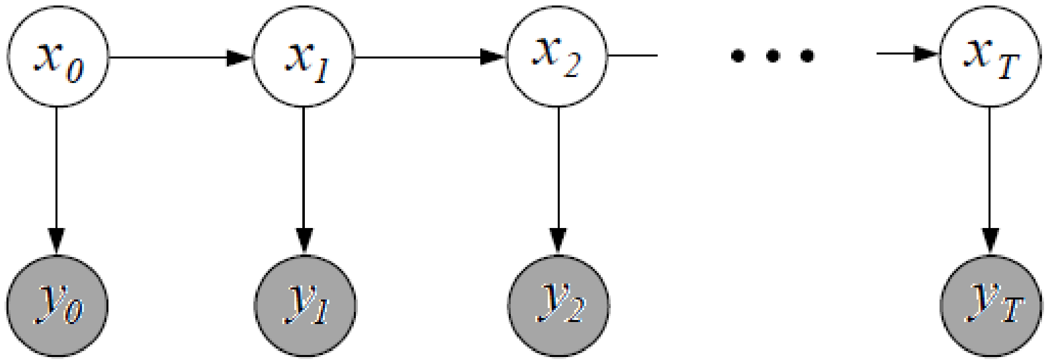
Like HMM, but over continuous variables...

Plant Equations:

$$x_{t+1} = Ax_t + Gw_t$$

$$y_t = Cx_t + v_t$$

State Space Models



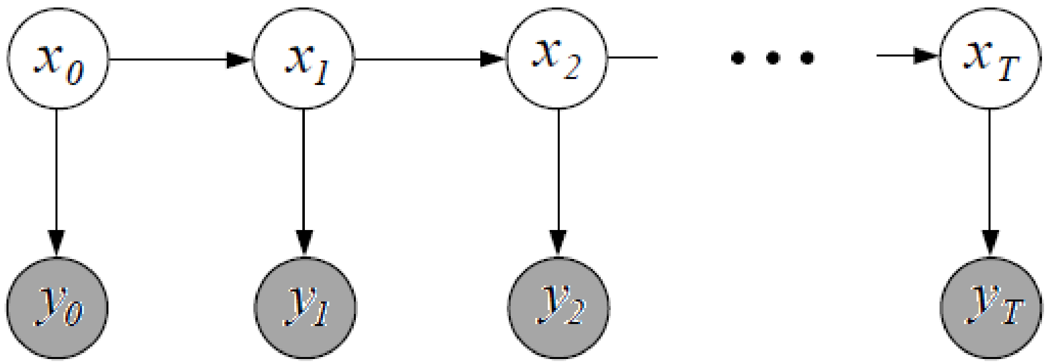
Dynamics: $x_{t+1} = Ax_t + Gw_t$

$$w_t \sim N(0, Q)$$

Process Noise

Same as, $p(x_{t+1} \mid x_t) = N(x_{t+1} \mid Ax_t, GQG^T)$

State Space Models



Measurement: $$y_t = Cx_t + v_t$$

$$v_t \sim N(0, R)$$

Measurement Noise

Same as, $p(y_{t+1} \mid x_{t+1}) = N(y_{t+1} \mid Cx_{t+1}, R)$

This is known as the ***Linear Gaussian*** assumption.

Example: Target Tracking

Target State

Measurement

Position

$$x_t = \begin{pmatrix} p_t \\ \dot{p}_t \end{pmatrix}$$

Velocity

$$y_t = p_t$$

- Constant velocity dynamics (1st order diffeq)

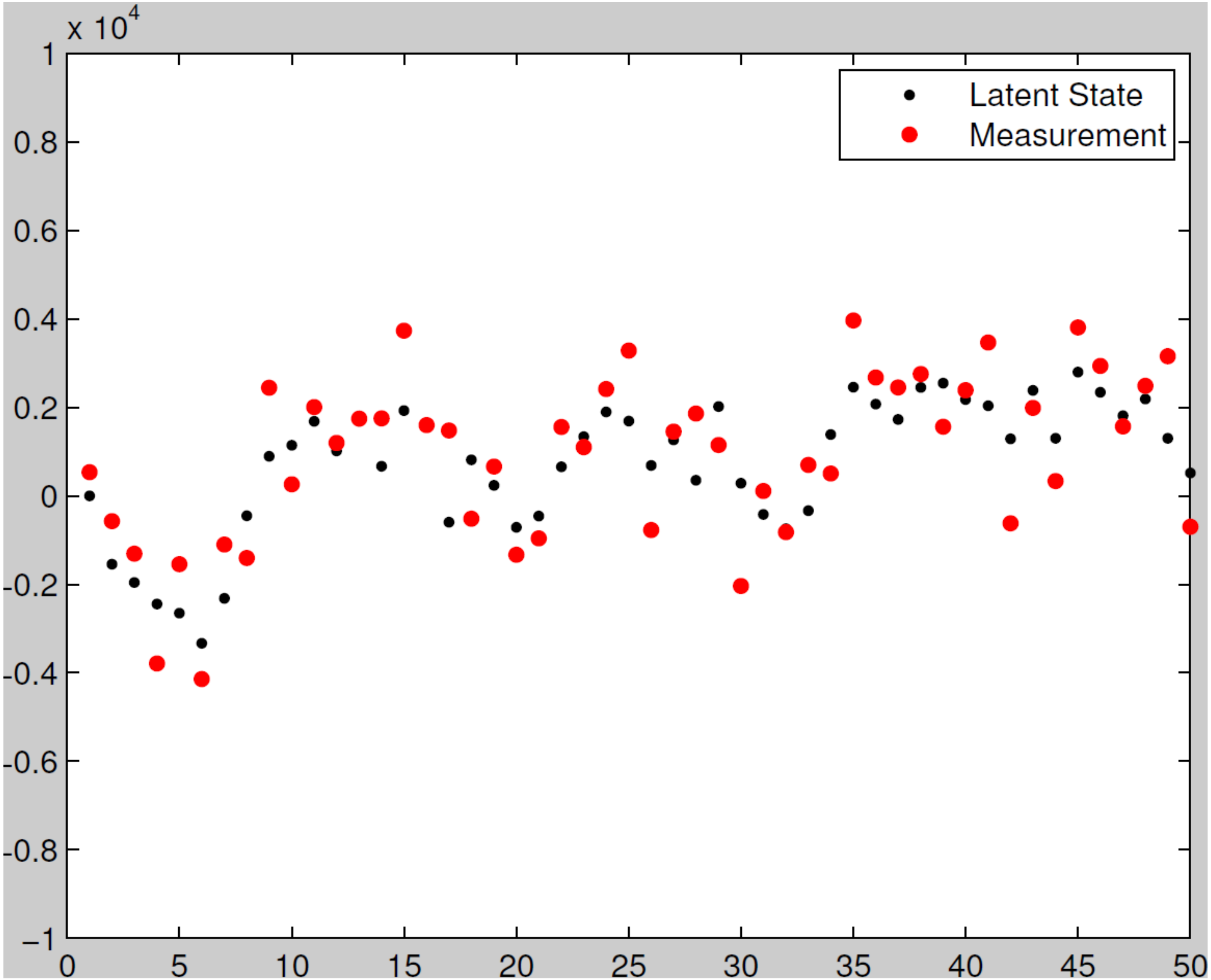
$$p_t = p_{t-1} + \dot{p}_t \qquad \dot{p}_t = \dot{p}_{t-1}$$

$$x_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} x_{t-1} + w_{t-1}$$

- Lower-dimensional measurement

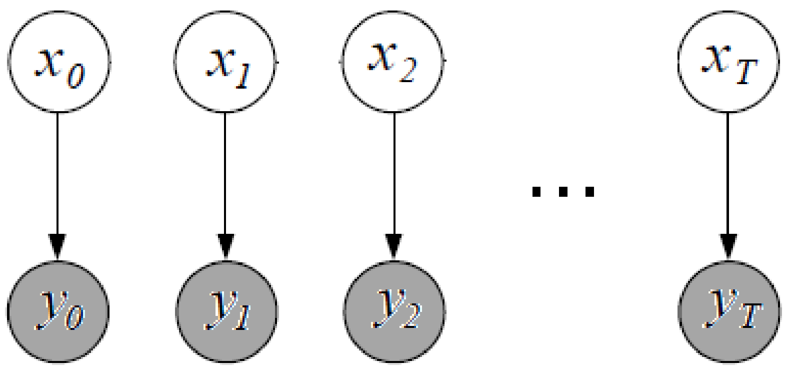
$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} x_t + v_t$$

Example: Target Tracking



Correspondence With Factor Analysis

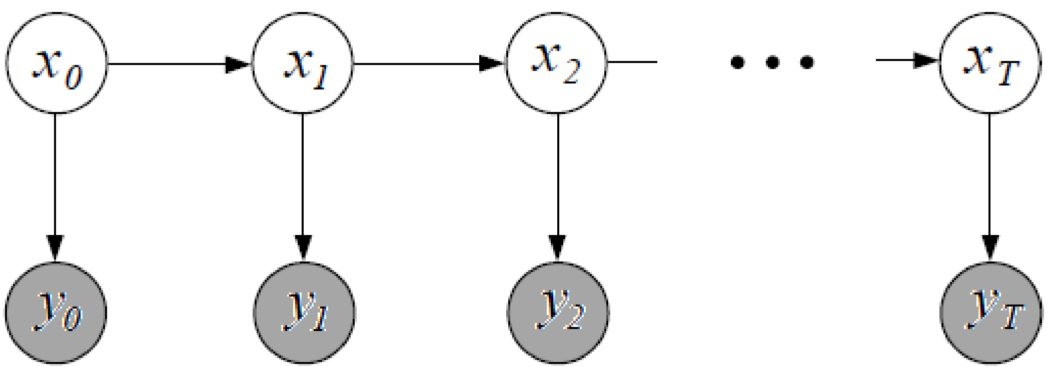
Factor Analysis:



$$x_t \sim N(0, I)$$

$$y_t \mid x_t \sim N(Fx_t, \Phi)$$

State Space:



$$x_t \mid x_{t-1} \sim N(Ax_{t-1}, GQG^T)$$

$$y_t \mid x_t \sim N(Cx_t, R)$$

Inference for State Space Model

- Broken into 2 parts
  - Filtering (Forward pass):

$$p(x_t \mid y_0, y_1, \ldots, y_t)$$

  - Smoothing (Backward pass)

$$p(x_t \mid y_0, y_1, \ldots, y_T)$$

- Why “filter”?
  - From signal processing
  - “Filters” out system noise to produce an estimate

Conditional Moments

- Everything is Gaussian

$$p(x_t \mid y_0, \ldots, y_t) = N(x_t \mid \hat{x}_{t|t}, P_{t|t})$$

- Can focus on mean / variance computations

Conditional Mean

$$\hat{x}_{t|t} \triangleq E[x_t | y_0, \ldots, y_t]$$

$$P_{t|t} \triangleq E[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T | y_0, \ldots, y_t]$$

Conditional Variance

Kalman Filter

<u>Two recursive updates:</u>

1) Time update:

- best guess *before* seeing measurement)

$$P(x_t | y_0, \ldots, y_t) \rightarrow P(x_{t+1} | y_0, \ldots, y_t)$$

2) Measurement Update:

- after measurement

$$P(x_{t+1} | y_0, \ldots, y_t) \rightarrow P(x_{t+1} | y_0, \ldots, y_{t+1})$$

Kalman Filter

1) Time update:

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t}$$

$$P_{t+1|t} = AP_{t|t}A^T + GQG^T$$

2) Measurement Update:

$$K_{t+1} \triangleq P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}$$

$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + K_{t+1}(y_{t+1} - C\hat{x}_{t+1|t})$$

$$P_{t+1|t+1} = P_{t+1|t} - K_{t+1}CP_{t+1|t}$$

Kalman Filter: Time Update

Recall that,

$$\hat{x}_{t+1|t} \triangleq E[x_{t+1} \mid y_0, \ldots, y_t]$$

so the conditional mean recursion is,

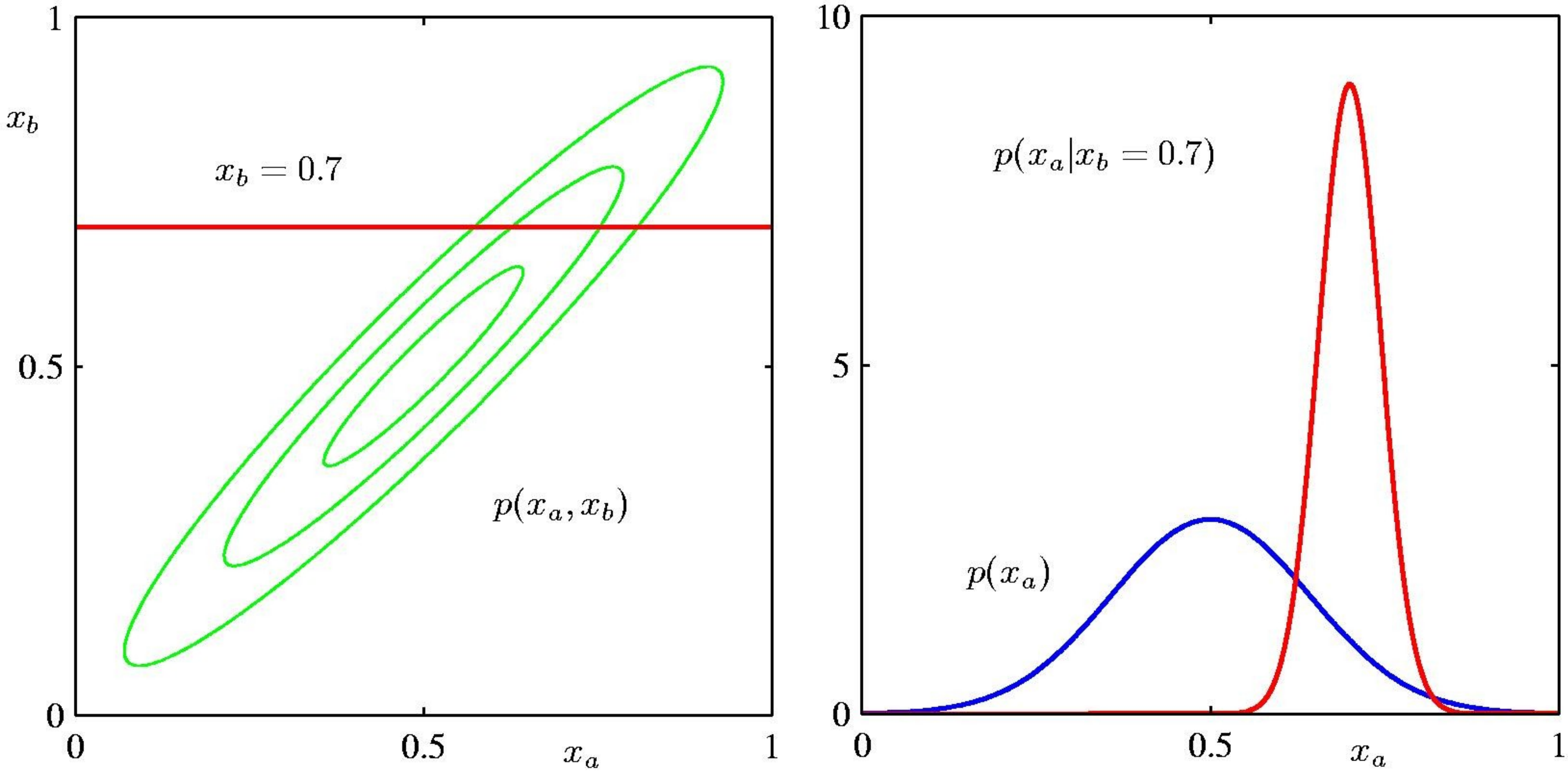
$$\begin{aligned} \hat{x}_{t+1} &= E[Ax_t + Gw_t] \\ &= A\hat{x}_{t|t} \end{aligned}$$

Zero Noise
In Expectation

Similar for covariance,

$$\begin{aligned} P_{t+1|t} &= E\left[(x_{t+1} - \hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T | y_0, \ldots, y_t\right] \\ &= E\left[(Ax_t + Gw_t - A\hat{x}_{t|t})(Ax_t + Gw_t - A\hat{x}_{t|t})^T | y_0, \ldots, y_t\right] \\ &= AP_{t|t}A^T + GQG^T, \end{aligned}$$

Gaussian Conditionals



For any joint multivariate Gaussian distribution, all conditional distributions are Gaussians

Gaussian Conditionals

Gaussian joint distributi$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

Inverse Covariance

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \quad \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix} \equiv \boldsymbol{\Sigma}^{-1}$$

Conditional Covariance

Conditional is Gaussian with parameters,

$$p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Lambda}_{aa}^{-1})$$

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b)$$

Kalman Filter: Measurement Update

Form the joint over x_{t+1}, y_{t+1} as,

$$p(x_{t+1} \mid y_0, \ldots, y_t) p(y_{t+1} \mid x_{t+1})$$

Time Update

Measurement Equation

Compute conditional,

$$p(x_{t+1} \mid y_0, \ldots, y_{t+1})$$

$$= \; N(x_{t+1} \mid \hat{x}_{t+1|t+1}, P_{t+1|t+1})$$

Kalman Filter: Measurement Update

$$K_{t+1} \triangleq P_{t+1|t} C^T (C P_{t+1|t} C^T + R)^{-1}$$
$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + K_{t+1}(y_{t+1} - C\hat{x}_{t+1|t})$$
$$P_{t+1|t+1} = P_{t+1|t} - K_{t+1} C P_{t+1|t}$$

- Quantity K_{t+1} called the *Kalman Gain Matrix*
- Because it multiplies observation, i.e. produces “gain”
- Update takes linear combination of predicted mean and observation, weighted by predicted covariance

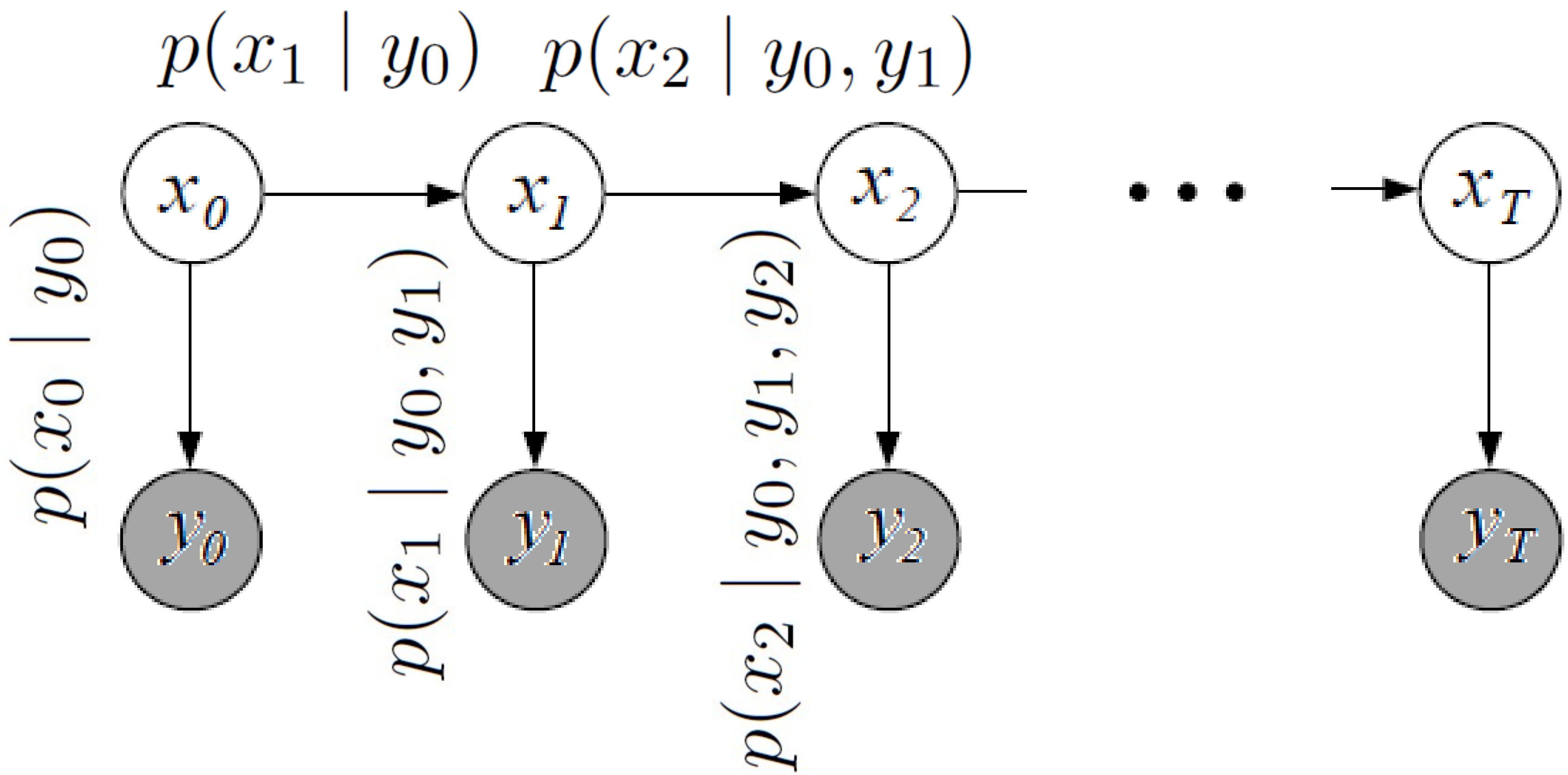
Kalman Filter

Consider the covariance updates,

$$
\begin{aligned}
P_{t+1|t} &= AP_{t|t}A^T + GQG^T \\
P_{t+1|t+1} &= P_{t+1|t} - K_{t+1}CP_{t+1|t}
\end{aligned}
$$

- Independent of observed measurements
- This is a property of Gaussians in general
- Only depend on process and measurement noise
- Can be computed offline

Kalman Filter



Information Filter

- Recall Gaussian has equivalent *canonical* parameterization
- Sometimes called *Information form*

Inverse Covariance

$$S_{t+1|t} = P^{-1}_{t+1|t}$$

$$S_{t+1|t+1} = P^{-1}_{t+1|t+1}$$

Canonical Mean

$$\hat{\xi}_{t+1|t} = P^{-1}_{t+1|t}\hat{x}_{t+1|t}$$

$$\hat{\xi}_{t+1|t+1} = P^{-1}_{t+1|t+1}\hat{x}_{t+1|t+1}$$

- Recursive updates follow definitions
- Matrix condition is reciprocal of condition of its inverse

Information Filter

- Define $H \triangleq GQG^T$, focus on precision update

$$\begin{aligned} S_{t+1|t} &= P_{t+1|t}^{-1} \\ &= (AP_{t|t}A^T + H)^{-1} \\ &= H^{-1} - H^{-1}A(S_{t|t} + A^T H^{-1} A)^{-1} A^T H^{-1} \end{aligned}$$

Matrix Inversion Lemma

- Measurement update

$$S_{t+1|t+1} = C^T R^{-1} C + S_{t+1|t}$$

- Wait a minute...this looks familiar...

Gaussian Belief Propagation

Message Update:

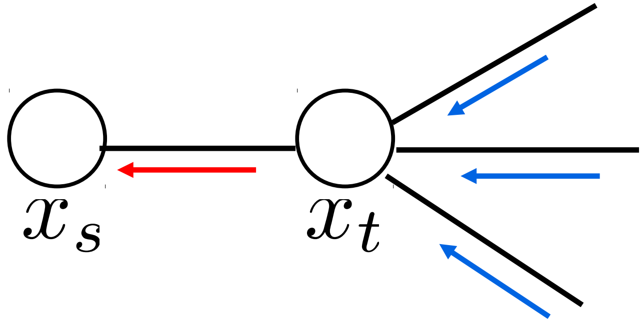
$$m_{ts}^{n}(x_s) = \alpha \mathcal{N}^{-1}(\vartheta_{ts}^{n}, \Lambda_{ts}^{n})$$

$$\Lambda_{ts}^{n} = J_{s(t)} - J_{s,t}(J_{t(s)} + C_t^T R_t^{-1} C_t + \sum_{u \in N(t) \setminus s} \Lambda_{ut}^{n-1})^{-1} J_{t,s}$$

Belief Update:

$$p(x_s \mid y) = \mathcal{N}^{-1}(\vartheta_s, \Lambda_s)$$

$$\Lambda_s^n = C_s^T R_s^{-1} C_s + \sum_{t \in N(s)} \Lambda_{ts}^n$$



Gaussian BP => Kalman Filter

Time Update:

$$S_{t+1|t} = H^{-1} - H^{-1}A(S_{t|t} + A^T H^{-1} A)^{-1} A^T H^{-1}$$

$$\Lambda_{ts}^{n} = J_{s(t)} - J_{s,t}(J_{t(s)} + C_t^T R_t^{-1} C_t + \sum_{u \in N(t) \setminus s} \Lambda_{ut}^{n-1})^{-1} J_{t,s}$$

Measurement Update:

$$S_{t+1|t+1} = C^T R^{-1} C + S_{t+1|t}$$

$$\Lambda_s^n \quad = \quad C_s^T R_s^{-1} C_s + \sum_{t \in N(s)} \Lambda_{ts}^n$$

Smoother

- Combines *forward* and backward probabilities

$$p(x_t \mid y_1, \ldots, y_t) p(x_t \mid y_{t+1}, \ldots, y_T)$$
$$\propto p(x_t \mid y_1, \ldots, y_T)$$

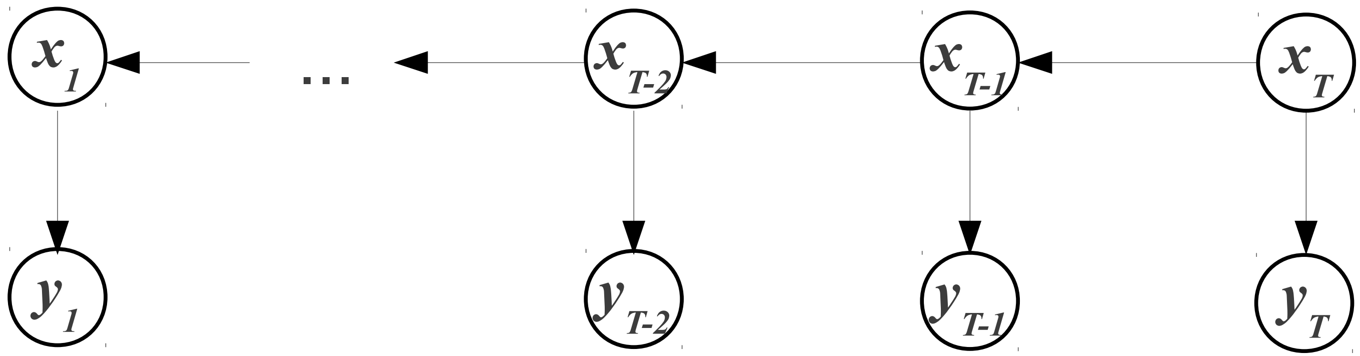
  to produce full marginal posterior.
- Similar to inference on an HMM (forward-backward algorithm)

Smoother

- Can we just invert the dynamics,

$$x_t = A^{-1}x_{t+1} - A^{-1}Gw_t$$

 and run Kalman filter backwards?



- No, w_t is no longer independent of the "past" state (e.g. x_{t+1}, ..., x_T)

Unconditional Distribution

- Marginal distributions of a Gaussian are Gaussian

$$p(x_t) = N(0, \Sigma_t)$$

From zero-mean noise assumption

Where,

$$\Sigma_t = A\Sigma_{t-1}A^T + GQG^T$$

- Covariance computed recursively
- Does not depend on means

Smoother

Given the unconditional marginal,

$$p(x_t) = N(0, \Sigma_t)$$

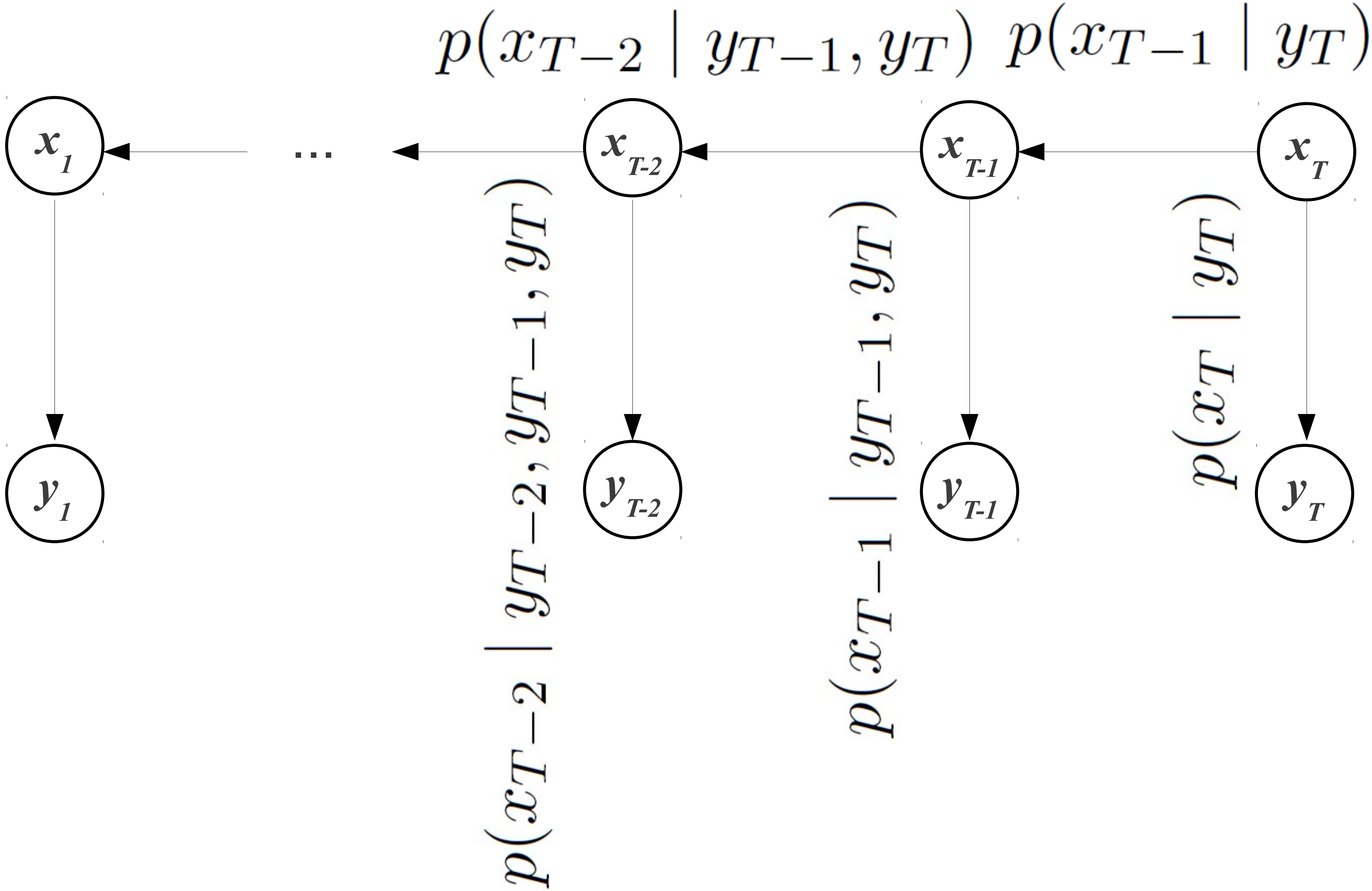
$$\Sigma_{t+1} = A\Sigma_t A^T + GQG^T$$

- Form the unconditional over x_t, x_{t+1}
- Solve for reverse dynamics

$$x_t = \widetilde{A} x_{t+1} + \widetilde{G} \widetilde{w}_{t+1}$$

- Run filter backwards with new dynamics

Smoother


$p(x_{T-2} \mid y_{T-1}, y_T)$
$p(x_{T-1} \mid y_T)$
x_1
...
x_{T-2}
x_{T-1}
x_T
$p(x_{T-2} \mid y_{T-2}, y_{T-1}, y_T)$
$p(x_{T-1} \mid y_{T-1}, y_T)$
$p(x_T \mid y_T)$
y_1
y_{T-2}
y_{T-1}
y_T

Smoother

Forward Conditional　　　　Backward Conditional

$$N(x_t \mid \hat{x}_{t|t}, P_{t|t})\, N(x_t \mid \hat{\mu}_{t|t}, S_{t|t})$$

- Gaussian closed under multiplication
- Multiply to produce full "smoothed" marginal

$$N(x_t \mid \hat{x}_t, P_t)$$

- Kalman filter + smoother equivalent to Gaussian BP

Summary

- Kalman filter is *optimal filter* for Linear Gaussian State Space model
- Smoother provides full marginal inference
- Gaussian BP produces equivalent algorithm
- Correspondence clearly shown in Information form of the filter