Posterior Cramér-Rao bounds for estimation in graphical models

Blabla 31.03.2005

Justin Dauwels

dauwels@isi.ee.ethz.ch | www.dauwels.com

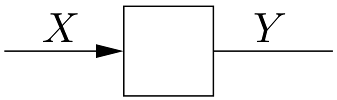
Signal and Information Processing Laboratory

Department of Information Technology and Electrical Engineering

ETH Zurich, Switzerland

Motivation

Estimation



noisy channel

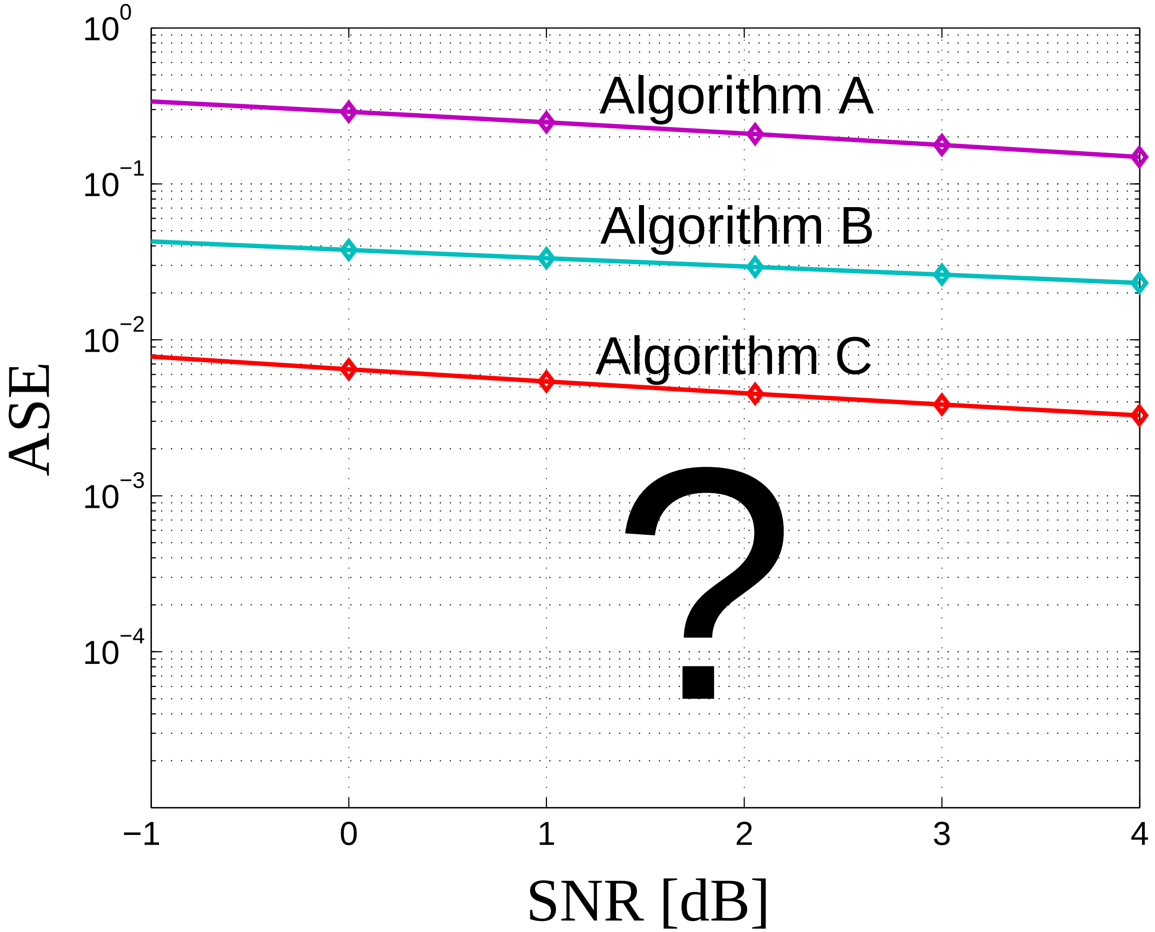
X and Y continuous random vectors

Estimators

- Block/Symbol MAP
- Block/Symbol ML
- Minimum mean squared error (MMSE)

Often infeasible!

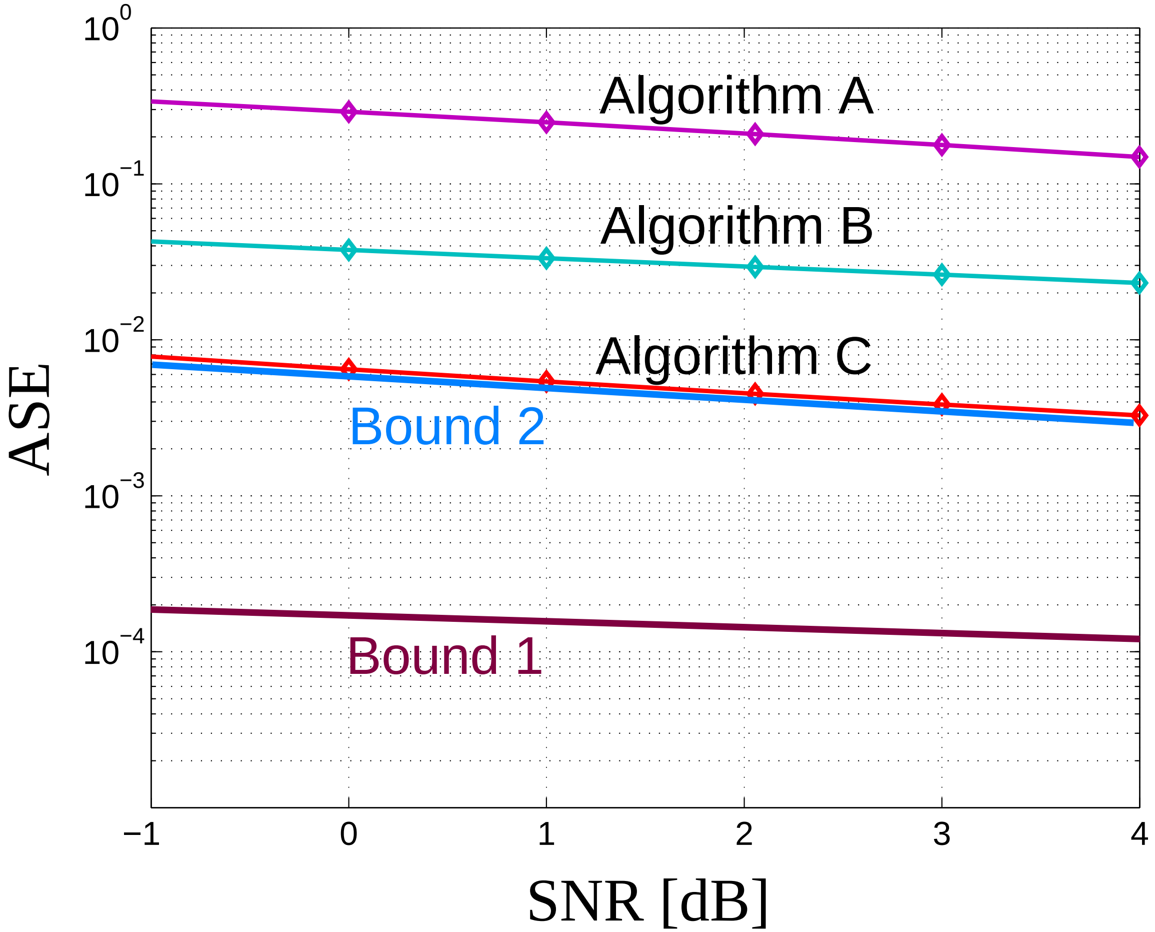
Motivation (2)



How close to MMSE estimator?

Partial answer: lower bound on MSE of "any" estimator

Motivation (3)



Can only confirm that algorithm is close to optimal!

Overview

- Posterior Cramér-Rao bound
- Message passing algorithm
- Applications
  - State space models
  - Joint decoding and channel estimation
  - Coupled hidden Markov models
- Conclusion
- Outlook

Notation/ Definitions

- $X \triangleq (X_1, X_2, \ldots, X_N)^T$, $Y \triangleq (Y_1, Y_2, \ldots, Y_M)^T$, and $X_i, Y_i \in \mathbb{R}$.
- $p(x, y)$ joint probability function of X and Y.
- $\hat{X}(Y)$ estimate of X based on observations Y.
- Error matrix $\mathbf{E} \triangleq \mathrm{E}_{XY}[(\hat{X}(Y) - X)(\hat{X}(Y) - X)^T]$.
- Posterior information matrix $\mathbf{J}$

$$J_{ij} \triangleq \mathrm{E}_{XY}\left[\nabla_{x_i} \log p(x, y) \nabla_{x_j}^T \log p(x, y)\right].$$

- Information matrix $\mathbf{I}(\theta)$

$$I_{ij}(\theta) \triangleq \mathrm{E}_{Y}\left[\nabla_{\theta_i} \log p(y|\theta) \nabla_{\theta_j}^T \log p(y|\theta)\right].$$

Posterior Cramér-Rao bound

Theorem (Van Trees '68)

If

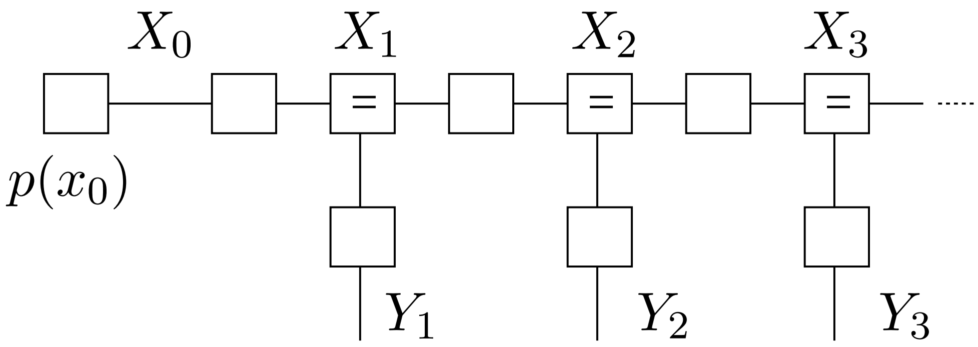
1. the posterior information matrix $\mathbf{J}$ exists and is non-singular,
2. $\int_x \nabla_{x_j} [B(x)p(x)]dx = 0$, where $B(x) = \int_y [\hat{x}(y) - x]p(y|x)dy$,

then $\mathbf{E} \succeq \mathbf{J}^{-1}$ (posterior Cramér-Rao bound).

In words:
$\mathbf{D} \triangleq \mathbf{E} - \mathbf{J}^{-1}$ is positive semi-definite, i.e., $v^T \mathbf{D} v \geq 0, \forall v \in \mathbb{R}^N$.

Holds for biased $\hat{x}(y)$!

Posterior Cramér-Rao bound (2)



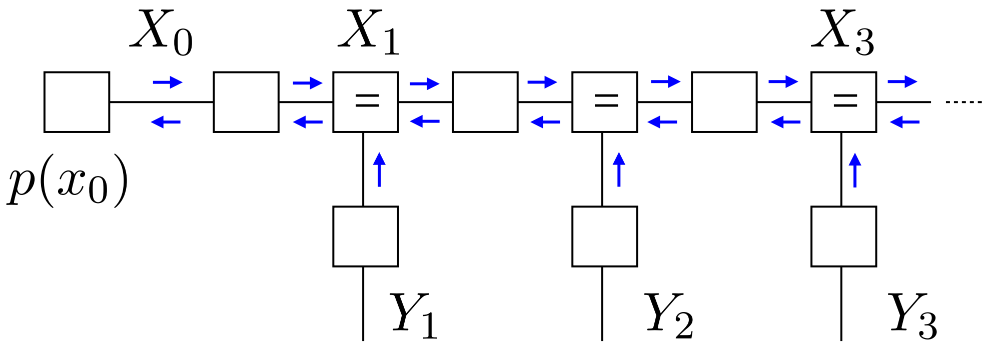
From PCRB, it follows

- $\mathrm{E}[(\hat{X}_i(Y) - X_i)^2] = E_{ii} \succeq [\mathbf{J}^{-1}]_{ii}$
- $\sum_i \mathrm{E}[(\hat{X}_i(Y) - X_i)^2] = \sum_i E_{ii} \geq \sum_i [\mathbf{J}^{-1}]_{ii}$.

Observations

- Only need the diagonal elements of $\mathbf{J}^{-1}$.
- $\mathbf{J}$ is often sparse.
- Inversion can be done efficiently by matrix inversion lemma.

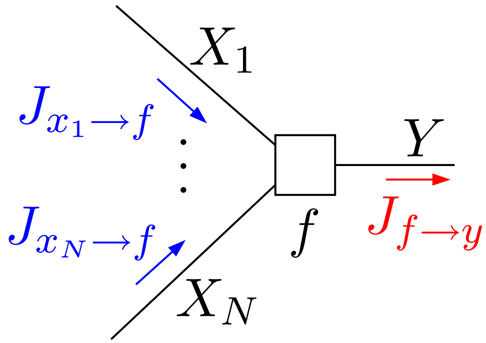
PCRB for estimation in graphical models



- The PCRBs of estimation in cycle-free graphical models can be computed efficiently by message passing.
- Messages are matrices.
- Messages are updated at each node according to specific update rule.
- The PCRBs are computed by combining those messages.

PCRB for estimation in graphical models (2)

Differentiable node function



$$J^{-1}_{f \to y}(Y) = \left(\begin{bmatrix} J_{x_1 \to f}(X_1) + \mathrm{E}[-\Delta^{x_1}_{x_1} \log f] & \dots & \mathrm{E}[-\Delta^{x_N}_{x_1} \log f] & \mathrm{E}[-\Delta^{y}_{x_1} \log f] \\ \vdots & \dots & \dots & \vdots \\ \mathrm{E}[-\Delta^{x_1}_{x_N} \log f] & \dots & J_{x_N \to f}(X_N) + \mathrm{E}[-\Delta^{x_N}_{x_N} \log f] & \mathrm{E}[-\Delta^{y}_{x_N} \log f] \\ \mathrm{E}[-\Delta^{x_1}_{y} \log f] & \dots & \mathrm{E}[-\Delta^{x_N}_{y} \log f] & \mathrm{E}[-\Delta^{y}_{y} \log f] \end{bmatrix}^{-1} \right)_{N+1,N+1}$$

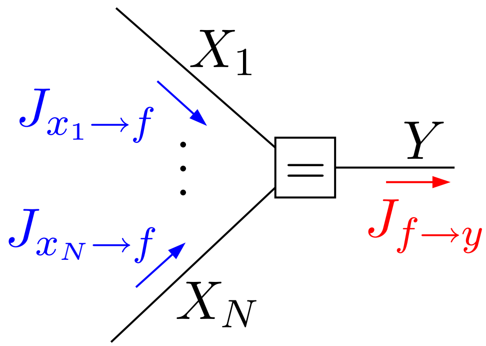
with $\Delta^{x_j}_{x_i} \triangleq \nabla_{x_i} \nabla^T_{x_j}$

Remarks

- Expectations $\mathrm{E}[\Delta^{x_j}_{x_i} \log f]$ supposed to be well-defined.
- They can easily be computed numerically.
- Rows and corresponding columns can be exchanged.

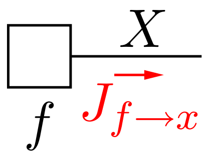
PCRB for estimation in graphical models (3)

Equality constraint node



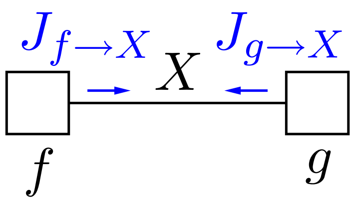
$$J_{f\to y} = \sum_{i=1}^{N} J_{x_i\to f}$$

Terminal node



$$J_{f\to x} = -\mathrm{E}[\Delta_x^x \log f]$$

PCRB



$$J_{\mathrm{tot}} = J_{f\to X} + J_{g\to X}$$

State space model: filtering

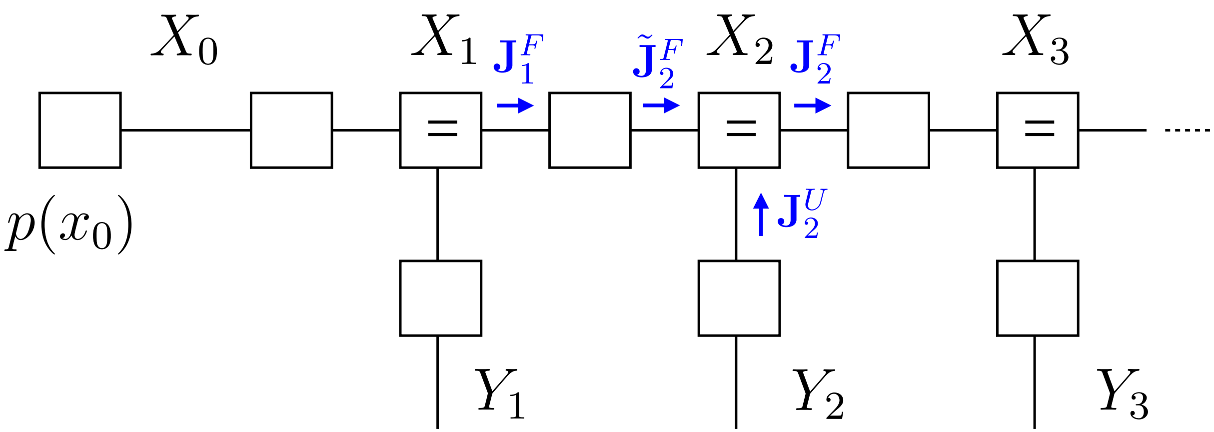
Forward sweep [Tichavský et al., 1998]

$$(\tilde{\mathbf{J}}^F_{k+1})^{-1} = \left(\begin{bmatrix} \mathbf{J}^F_k - \mathrm{E}[\Delta^{x_k}_{x_k} \log p(x_{k+1}|x_k)] & -\mathrm{E}[\Delta^{x_{k+1}}_{x_k} \log p(x_{k+1}|x_k)] \\ -\mathrm{E}[\Delta^{x_k}_{x_{k+1}} \log p(x_{k+1}|x_k)] & -\mathrm{E}[\Delta^{x_{k+1}}_{x_{k+1}} \log p(x_{k+1}|x_k)] \end{bmatrix}^{-1} \right)_{22}$$

$$\triangleq \left(\begin{bmatrix} \mathbf{J}^F_k + \mathbf{J}^{11}_k & \mathbf{J}^{12}_k \\ \mathbf{J}^{21}_k & \mathbf{J}^{22}_k \end{bmatrix}^{-1} \right)_{22}$$

$$\tilde{\mathbf{J}}^F_{k+1} = \mathbf{J}^{22}_k - \mathbf{J}^{21}_k (\mathbf{J}^F_k + \mathbf{J}^{11}_k)^{-1} \mathbf{J}^{12}_k$$

$$\mathbf{J}^F_{k+1} = -\mathrm{E}[\Delta^{x_{k+1}}_{x_{k+1}} \log p(y_{k+1}|x_{k+1})] + \tilde{\mathbf{J}}^F_{k+1} \triangleq \mathbf{J}^U_{k+1} + \tilde{\mathbf{J}}^F_{k+1}.$$

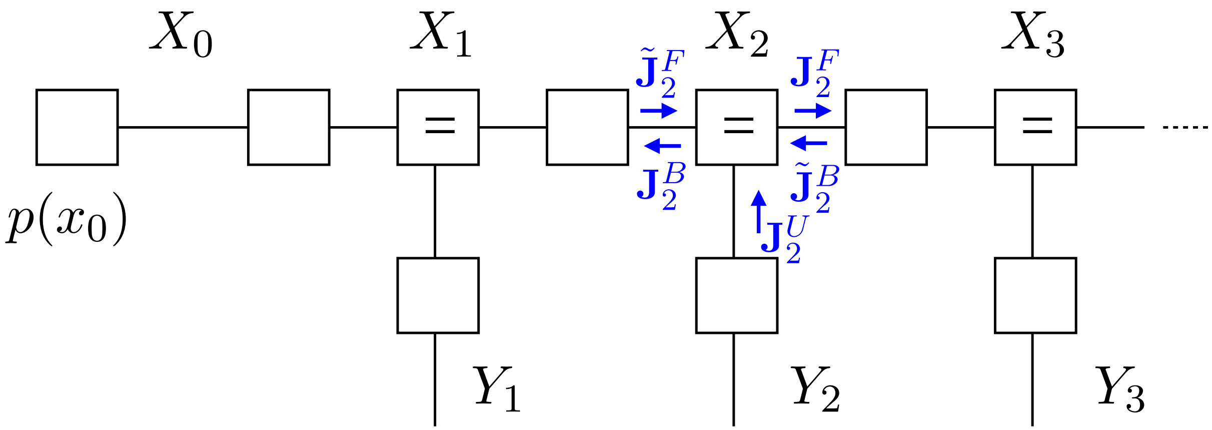


ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

State space model: smoothing

Forward + backward sweep



Posterior CRB

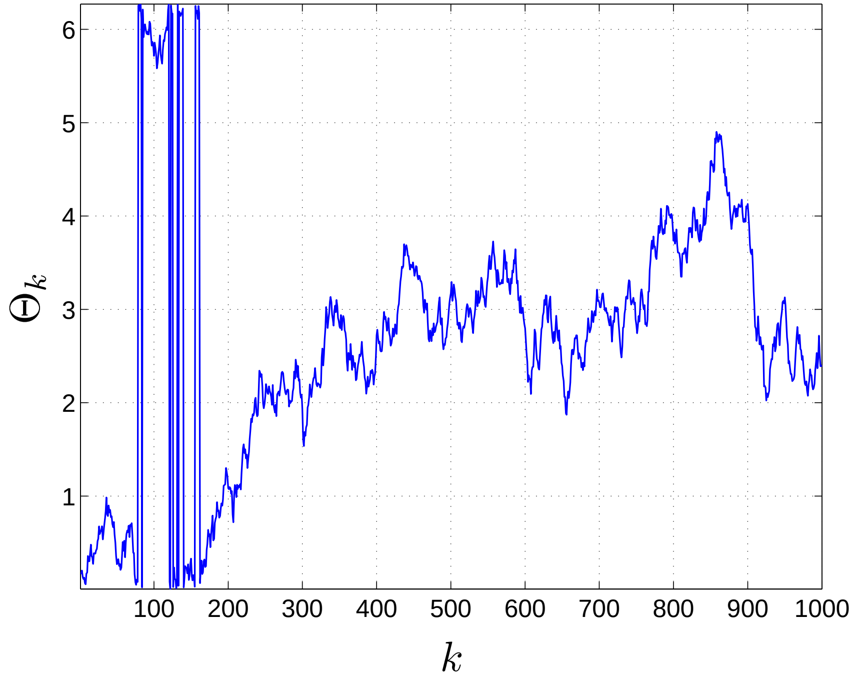
$$\mathbf{J}_k^{\text{tot}} = \tilde{\mathbf{J}}_k^F + \tilde{\mathbf{J}}_k^B + \mathbf{J}_k^U.$$

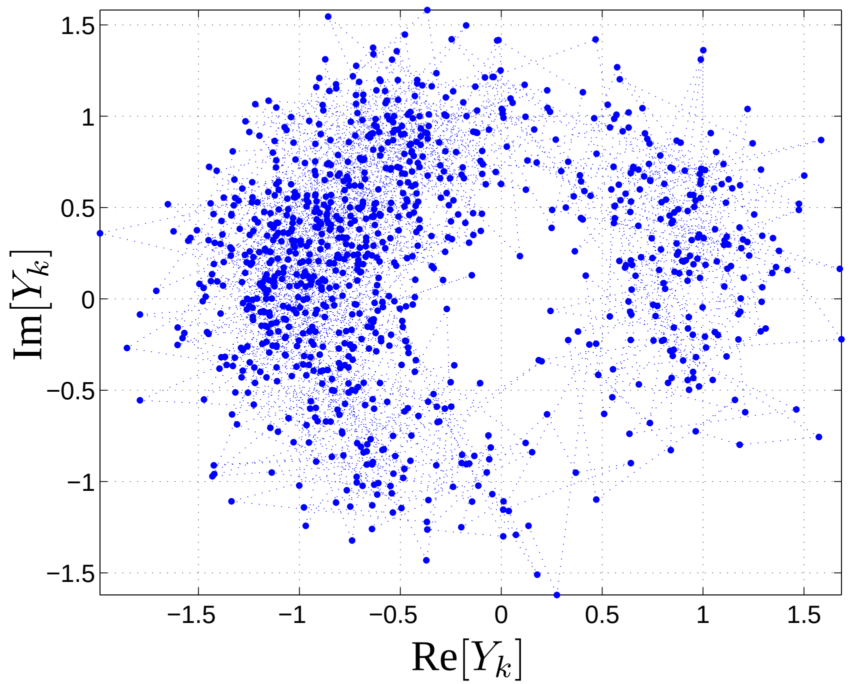
Example

Random walk phase model

$$\begin{aligned}\Theta_{k+1} &= (\Theta_k + W_k) \bmod 2\pi \\ Y_k &= \exp(j\Theta_k) + V_k,\end{aligned}$$

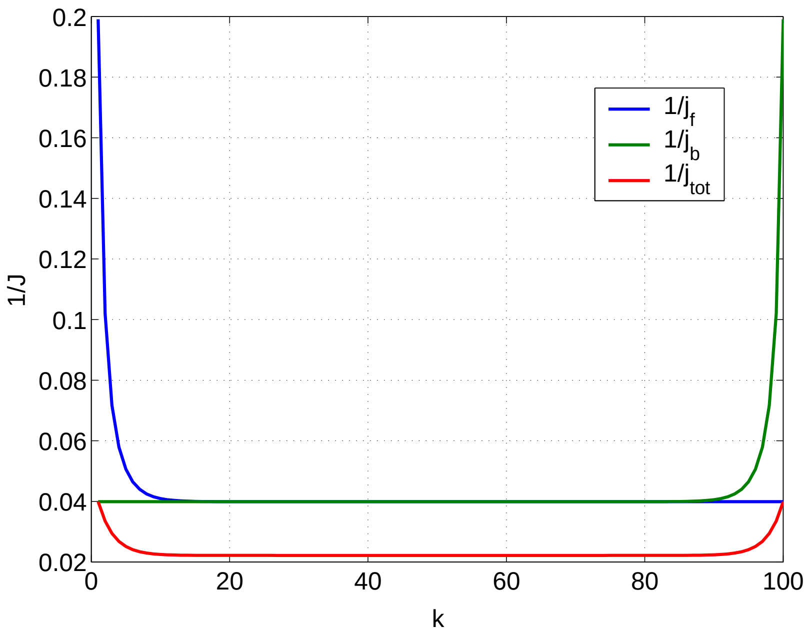
W_k and V_k: i.i.d. (mean free) Gaussian RVs with variance σ_θ^2 and $2\sigma_0^2$ resp., and $k = 1, \ldots, L$.

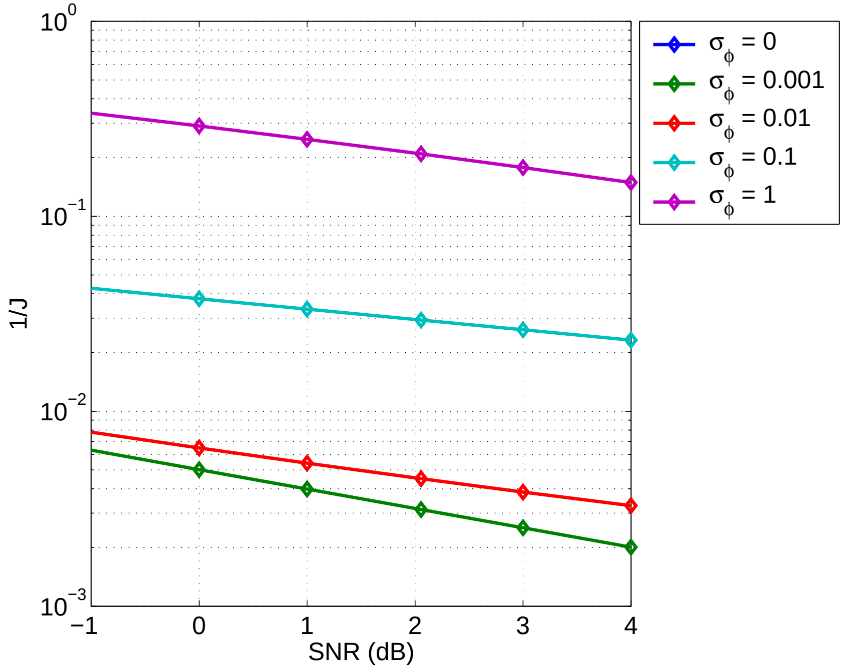




Example (2)

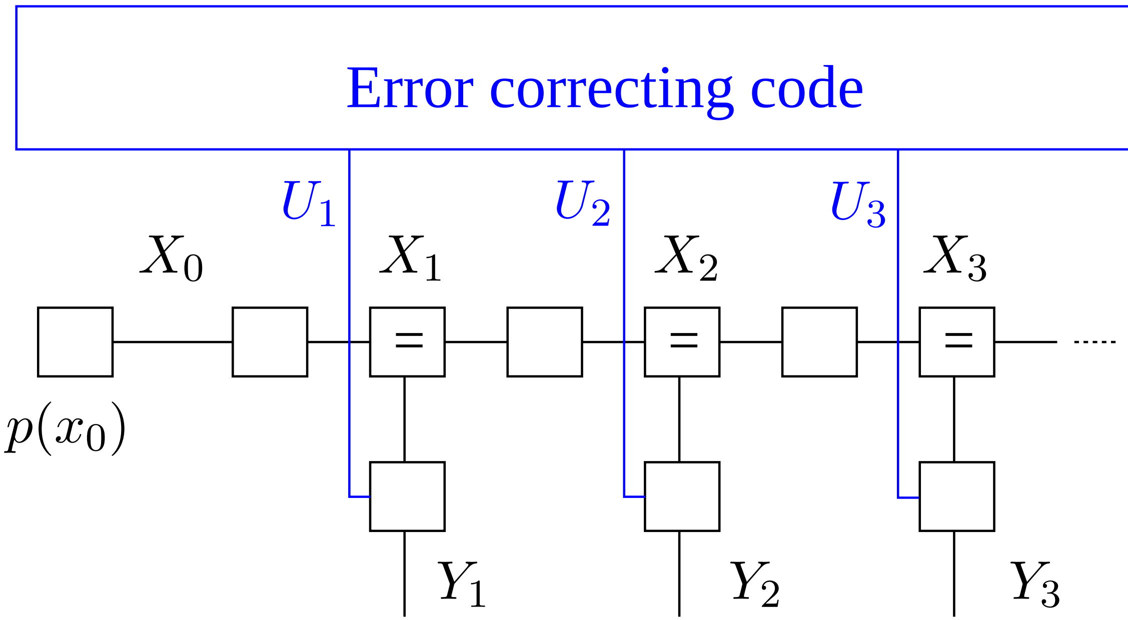
Random walk phase model ($L = 100$)

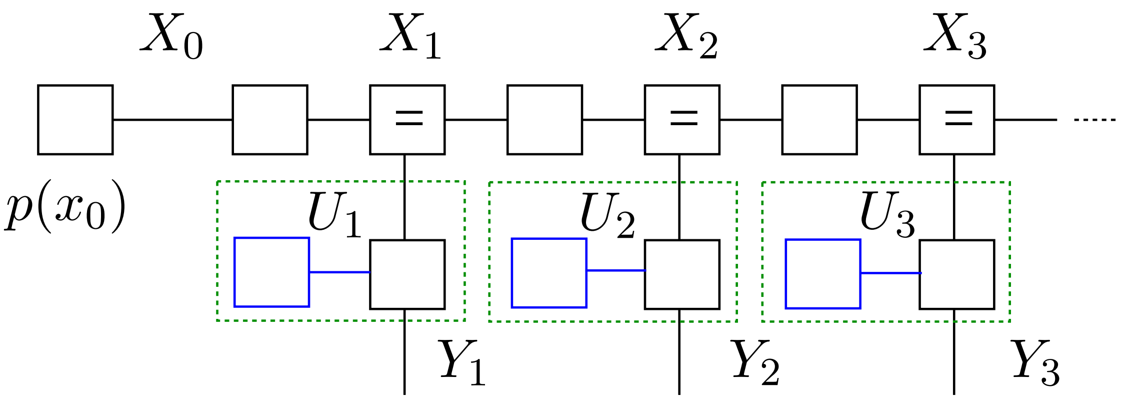




$$\sigma_0^2 = 0.1991, \sigma_\phi^2 = 0.01$$

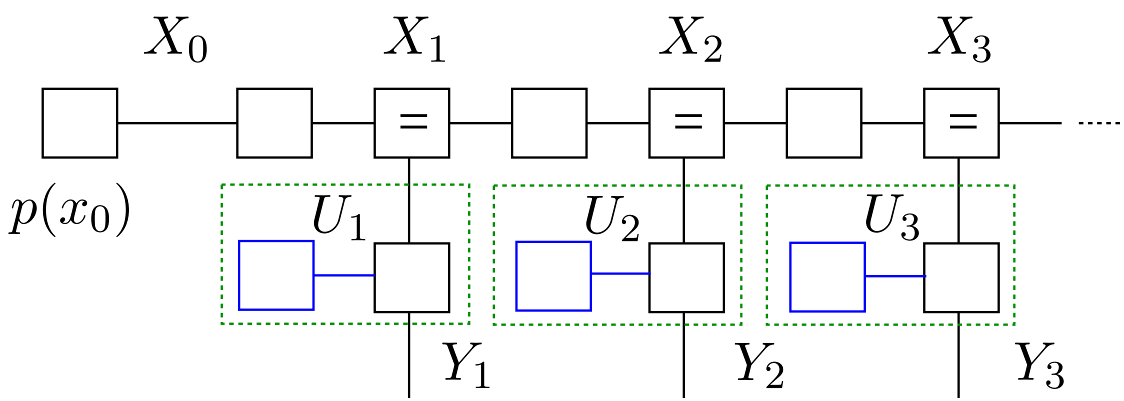
Channel with freely evolving state





Channel with freely evolving state (2)

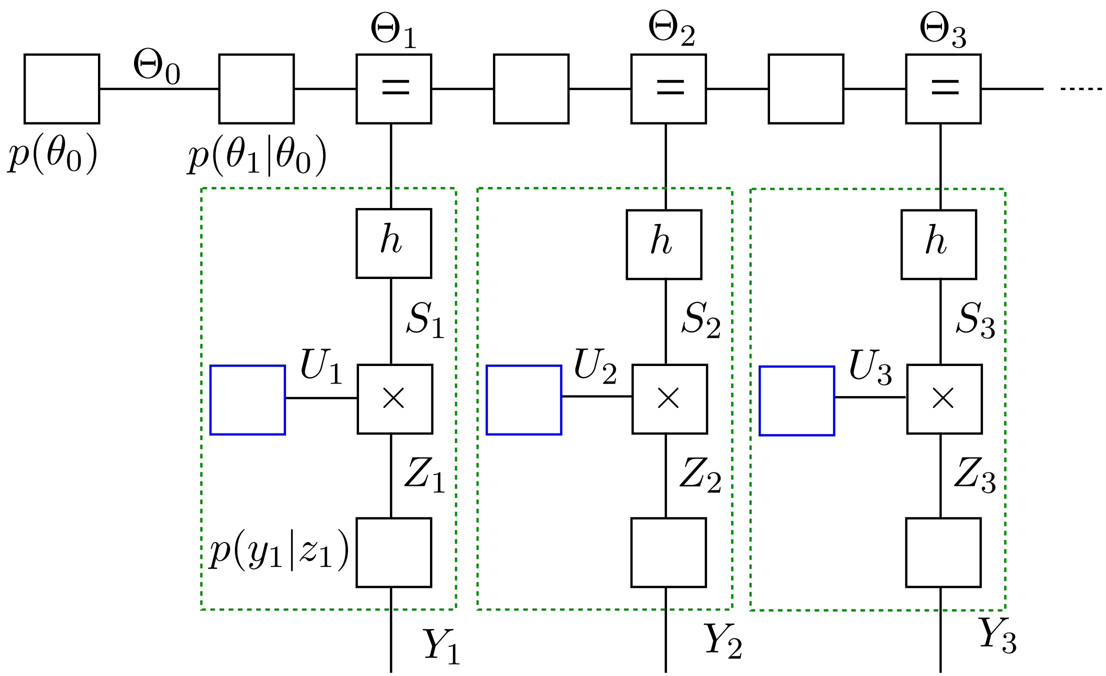
- Independence assumption + exact marginals of $U_k \Rightarrow$ exact PCRB
- Independence assumption + approximate marginals of $U_k \Rightarrow$?
- Known $U_k \Rightarrow$ lower bound on PCRB
- Unknown U_k with uniform prior $\Rightarrow$ upper bound on PCRB



Example

$$\begin{aligned}\Theta_{k+1} &= (\Theta_k + W_k) \bmod 2\pi \\ Y_k &= U_k \exp(j\Theta_k) + V_k,\end{aligned}$$

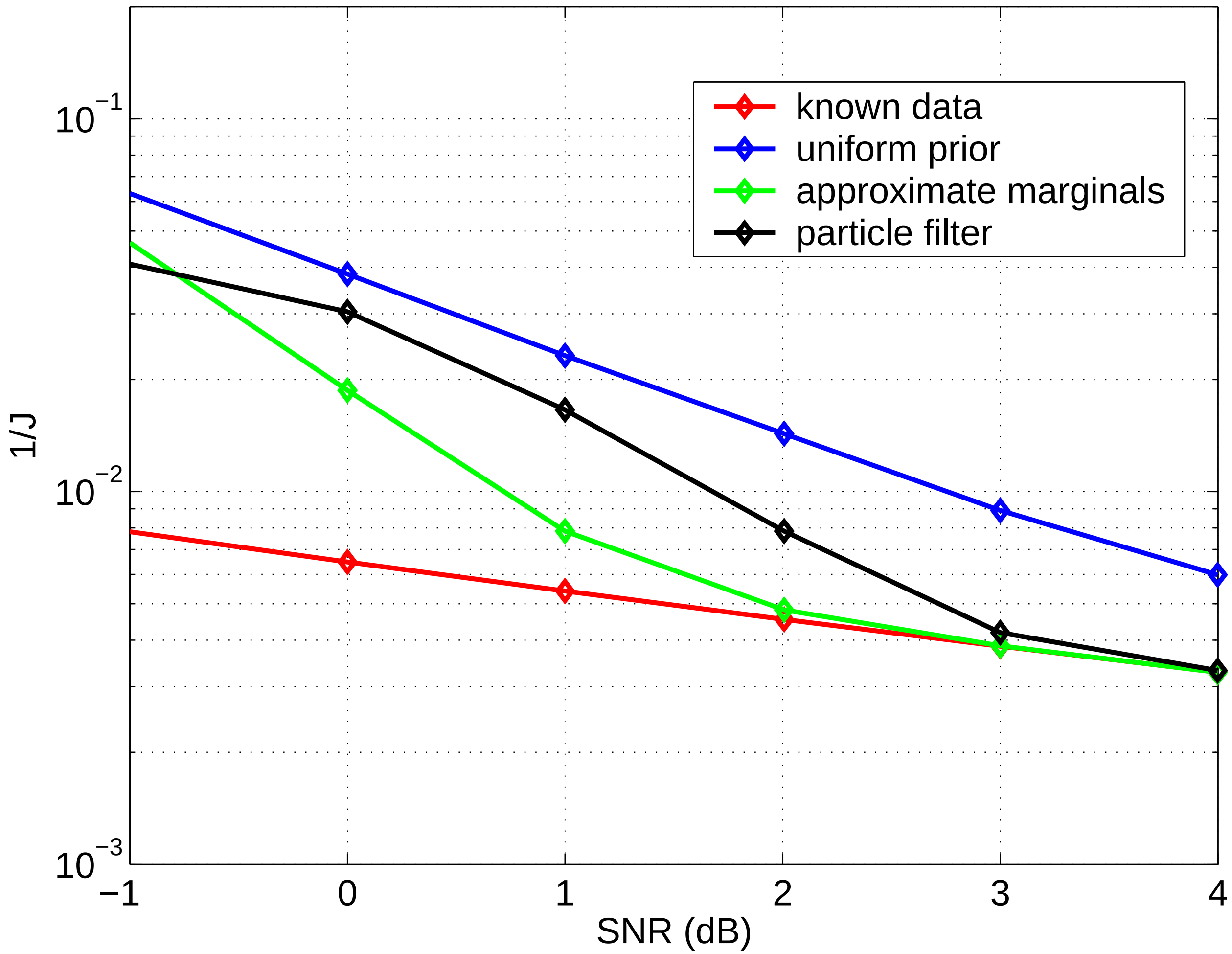
W_k and V_k: i.i.d. (mean free) Gaussian RVs with variance σ_θ^2 and $2\sigma_0^2$ resp.
U_k: M-ary symbols protected by an error correcting code



$S_i \triangleq \exp(j\Theta_i)$

Example (2)

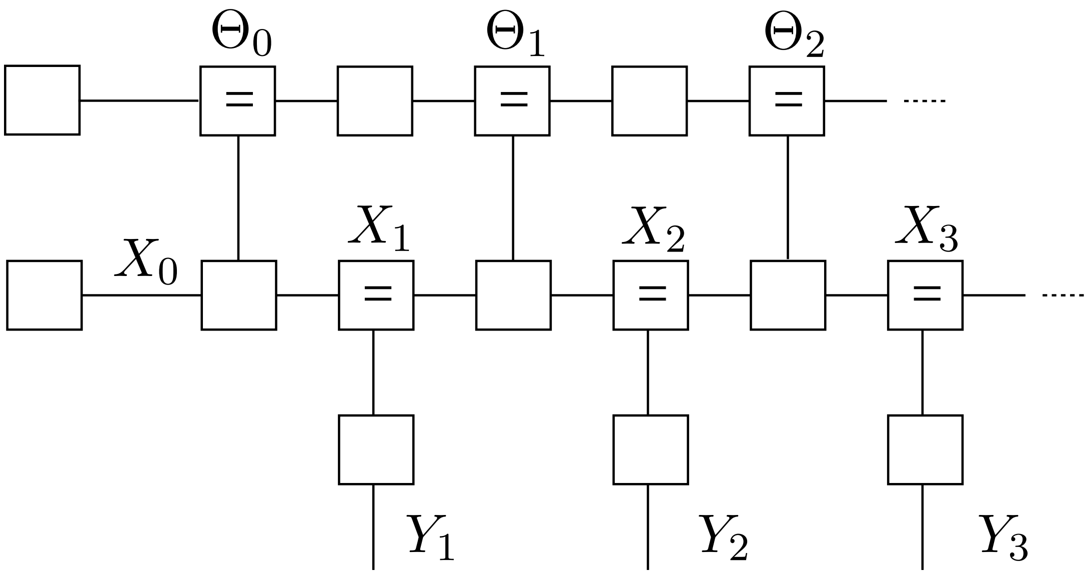
4-PSK; 3-6 LDPC of length 100; $\sigma_\phi = 0.01$

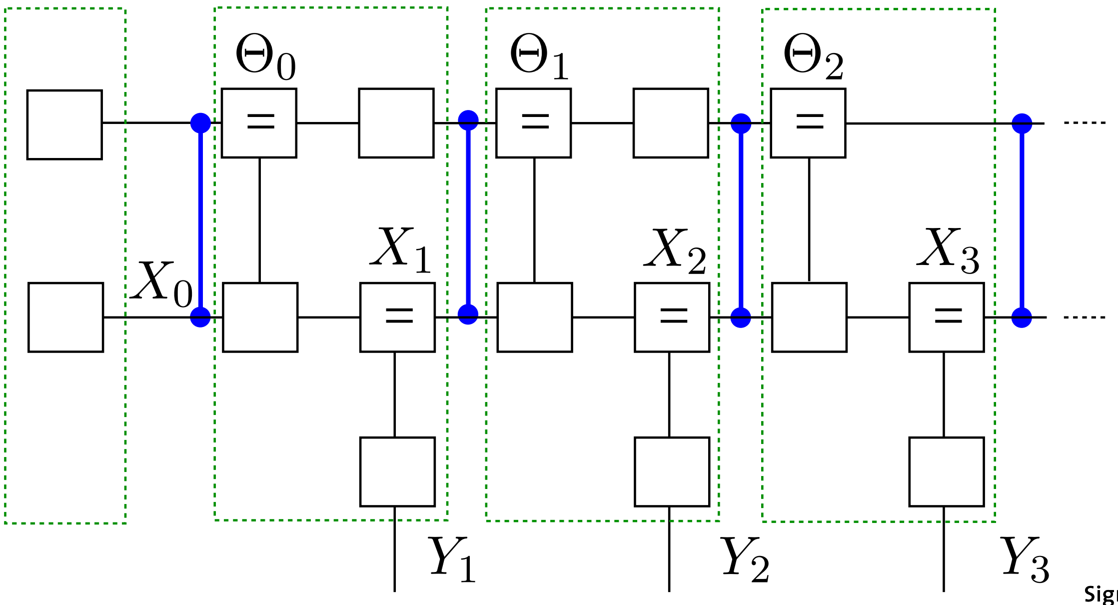


ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

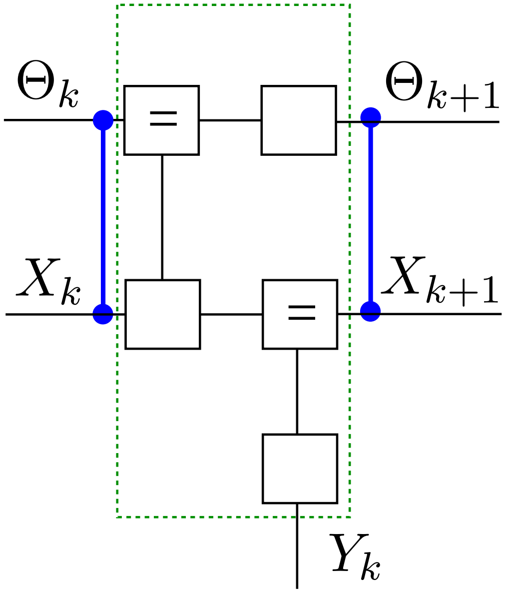
Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

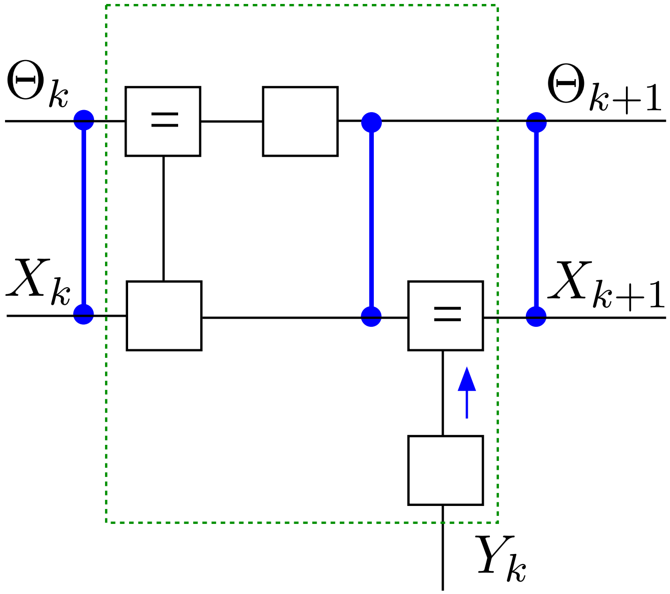
Coupled hidden Markov models





Coupled hidden Markov models (2)





- Message = 2 × 2 block matrix
- In general: N coupled HMMs ⇒ message $=$ $N \times N$ block matrix
- Updated by applying "PCRB rules".

Conclusion

- Posterior Cramér-Rao bound = lower bound on MSEE.
- Can confirm that practical algorithm is close to optimal.
- Efficiently/easily computed by message passing algorithm.
- Tight in interesting SNR region.
- Applications
  - Filtering/smoothing in state space models
  - Joint decoding + channel estimation
  - Coupled hidden Markov models.
- More information
  `http://www.dauwels.com/files/CRB_long.pdf`

Outlook

- PCRBs for some standard problems in signal processing.
- Design of pilot sequences.
- Extension to discrete variables.

Thank you for your attention!

Posterior Cramér-Rao bounds for estimation in graphical models

Blabla 31.03.2005

Justin Dauwels

dauwels@isi.ee.ethz.ch | www.dauwels.com

Signal and Information Processing Laboratory

Department of Information Technology and Electrical Engineering

ETH Zurich, Switzerland

Posterior information matrix

Posterior information matrix $\mathbf{J}(X)$ can be computed in several ways

$$
\begin{aligned}
\mathbf{J}_{ij}(X) &\triangleq \mathrm{E}_{XY}\left[\nabla_{x_i}\log p(x,y)\nabla^T_{x_j}\log p(x,y)\right] \\
&\overset{\text{(L1)}}{=} -\mathrm{E}_{XY}\left[\nabla_{x_i}\nabla^T_{x_j}\log p(x,y)\right] \\
&= -\mathrm{E}_{XY}\left[\nabla_{x_i}\nabla^T_{x_j}\log p(y|x)\right] - \mathrm{E}_{X}\left[\nabla_{x_i}\nabla^T_{x_j}\log p(x)\right] \\
&\overset{\text{(L2 \& L3)}}{=} \mathrm{E}_{XY}\left[\nabla_{x_i}\log p(y|x)\nabla^T_{x_j}\log p(y|x)\right] + \mathrm{E}_{X}\left[\nabla_{x_i}\log p(x)\nabla^T_{x_j}\log p(x)\right]
\end{aligned}
$$

Posterior information matrix (2)

Lemma 1

If

1. $\nabla_{x_j} p(x,y)$ and $\nabla_{x_i}\nabla_{x_j}^T p(x,y)$ exist $\forall x$ and y,
2. $\int_x \int_y p(x,y)\nabla_{x_i} \log p(x,y)\nabla_{x_j}^T \log p(x,y) dx dy$ and $\int_x \int_y p(x,y)\nabla_{x_i}\nabla_{x_j}^T \log p(x,y) dx dy$ exist,
3. $\int_{x,y} \nabla_{x_i}\nabla_{x_j}^T p(x,y) dx dy = 0$,

then $\mathrm{E}_{XY}\left[-\nabla_{x_i}\nabla_{x_j}^T \log p(x,y)\right] = \mathrm{E}_{XY}\left[\nabla_{x_i} \log p(x,y)\nabla_{x_j}^T \log p(x,y)\right]$.

Proof:

Since (1) holds, $\nabla_{x_j}^T p(x,y) = p(x,y)\nabla_{x_j}^T \log p(x,y)$ and

$$\begin{aligned}\nabla_{x_i}\nabla_{x_j}^T p(x,y) &= \nabla_{x_i}\left(p(x,y)\nabla_{x_j}^T \log p(x,y)\right)\\ &= p(x,y)\nabla_{x_i} \log p(x,y)\nabla_{x_j}^T \log p(x,y) + p(x,y)\nabla_{x_i}\nabla_{x_j}^T \log p(x,y).\end{aligned}$$

Integrating both sides over x and y and using (2) and (3), one obtains

$$\int_x \int_y p(x,y)\nabla_{x_i} \log p(x,y)\nabla_{x_j}^T \log p(x,y) dx dy + \int_x \int_y p(x,y)\nabla_{x_i}\nabla_{x_j}^T \log p(x,y) dx dy = 0.$$

Thus

$$\mathrm{E}_{XY}[\nabla_{x_i}\nabla_{x_j}^T \log p(x,y)] = -\mathrm{E}_{XY}[\nabla_{x_i} \log p(x,y)\nabla_{x_j}^T \log p(x,y)].$$

Posterior information matrix (3)

Lemma 2

If

1. $\nabla_{x_j} p(y|x)$ and $\nabla_{x_i} \nabla_{x_j}^T p(y|x)$ exist $\forall x$ and y,
2. $\int_x \int_y p(x,y) \nabla_{x_i} \log p(y|x) \nabla_{x_j}^T \log p(y|x) dx dy$ and $\int_x \int_y p(x,y) \nabla_{x_i} \nabla_{x_j}^T \log p(y|x) dx dy$ exist,

then $\mathrm{E}_{XY}\left[-\nabla_{x_i} \nabla_{x_j}^T \log p(y|x)\right] = \mathrm{E}_{XY}\left[\nabla_{x_i} \log p(y|x) \nabla_{x_j}^T \log p(y|x)\right]$.

Proof:

Differentiate both sides of the equation $\int_y p(y|x) dy = 1$ wrp to x_j. Since (1) holds, one can differentiate under the integral sign (Leibniz's formula). As a consequence

$$\int_y \nabla_{x_j}^T p(y|x) dy = \int_y p(y|x) \nabla_{x_j}^T \log p(y|x) dy = 0.$$

Differentiate a second time (under the integral sign), it follows

$$\int_y p(y|x) \nabla_{x_i} \nabla_{x_j}^T \log p(y|x) dy + \int_y \nabla_{x_i} p(y|x) \nabla_{x_j}^T \log p(y|x) dy = 0.$$

Multiply both sides with $p(x)$ and integrate over x using (2). As a consequence

$$\mathrm{E}_{XY}[\nabla_{x_i} \nabla_{x_j}^T \log p(y|x)] = -\mathrm{E}_{XY}[\nabla_{x_i} \log p(y|x) \nabla_{x_j}^T \log p(y|x)].$$

Posterior information matrix (4)

Lemma 3

If

1. $\nabla_{x_j} p(x)$ and $\nabla_{x_i} \nabla_{x_j}^T p(x)$ exist $\forall x$,
2. $\int_x p(x) \nabla_{x_i} \log p(x) \nabla_{x_j}^T \log p(x) dx$ and $\int_x p(x) \nabla_{x_i} \nabla_{x_j}^T \log p(x) dx$ exist,
3. $\int_x \nabla_{x_i} \nabla_{x_j}^T p(x) dx = 0$,

then $\mathrm{E}_X \left[-\nabla_{x_i} \nabla_{x_j}^T \log p(x) \right] = \mathrm{E}_X \left[\nabla_{x_i} \log p(x) \nabla_{x_j}^T \log p(x) \right]$.

Proof:

Since (1) holds, $\nabla_{x_j}^T p(x) = p(x) \nabla_{x_j}^T \log p(x)$ and

$$\begin{aligned} \nabla_{x_i} \nabla_{x_j}^T p(x) &= \nabla_{x_i} \left(p(x) \nabla_{x_j}^T \log p(x) \right) \\ &= p(x) \nabla_{x_i} \log p(x) \nabla_{x_j}^T \log p(x) + p(x) \nabla_{x_i} \nabla_{x_j}^T \log p(x). \end{aligned}$$

Integrating both sides over x and using (2) and (3), one obtains

$$\int_x p(x) \nabla_{x_i} \log p(x) \nabla_{x_j}^T \log p(x) dx + \int_x \int_y p(x) \nabla_{x_i} \nabla_{x_j}^T \log p(x) dx = 0.$$

Thus

$$\mathrm{E}_X [\nabla_{x_i} \nabla_{x_j}^T \log p(x)] = -\mathrm{E}_X [\nabla_{x_i} \log p(x, y) \nabla_{x_j}^T \log p(x)].$$

PCRB

Proof

$$\begin{aligned}
\nabla_{x_j}[B_i(x)p(x)] &= \nabla_{x_j}\int_y [\hat{x}_i(y) - x_i]p(y|x)p(x)dy \\
&= \nabla_{x_j}\int_y [\hat{x}_i(y) - x_i]p(x,y)dy \\
&\overset{(1)\&(2)}{=} -\delta_{ij}\int_y p(x,y)dy + \int_y [\hat{x}_i(y) - x_i]\nabla_{x_j}p(x,y)dy
\end{aligned}$$

Integrate wrp to x

$$\begin{aligned}
\int_x \nabla_{x_j}[B_i(x)p(x)]dx &= -\delta_{ij}\int_{x,y} p(x,y)dxdy + \int_{x,y}[\hat{x}_i(y) - x_i]\nabla_{x_j}p(x,y)dxdy \\
&\overset{(3)}{\Leftrightarrow} \\
0 &= -\delta_{ij} + \int_{x,y}[\hat{x}_i(y) - x_i]\nabla_{x_j}p(x,y)dxdy \\
&= -\delta_{ij} + \int_{x,y}[\hat{x}_i(y) - x_i]p(x,y)\nabla_{x_j}\log p(x,y)dxdy
\end{aligned}$$

Define vector $v = [\hat{X}_1(Y) - X_1, \ldots, \hat{X}_N(Y) - X_N, \nabla_{x_1}\log p(x,y), \ldots, \nabla_{x_N}\log p(x,y)]^T$ and

$\mathbf{C}_v \triangleq \mathrm{E}[vv^T] = \begin{bmatrix} \mathbf{E} & \mathbf{I} \\ \mathbf{I} & \mathbf{J}(X) \end{bmatrix}$, where $\mathbf{I}$ is the $N \times N$ unity matrix;

$\mathbf{C}_v \succeq 0$, since $u^T\mathbf{C}_v u = u^T\mathrm{E}[vv^T]u = \mathrm{E}[u^Tvv^Tu] = \mathrm{E}[(u^Tv)^2] \geq 0, \forall u$.

As a consequence of Lemma 4, $\mathbf{E} \succeq \mathbf{J}(X)^{-1}$.

PCRB (2)

Lemma 4

Let $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \succeq 0$, where $\mathbf{A}_{22}$ is nonsingular.
Then $(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) \succeq 0$.

Proof:
$\mathbf{A} \succeq 0$, which by definition means that $v^T\mathbf{A}v \geq 0, \forall v$.
Let $v = [v_1 v_2]^T$, then $v^T\mathbf{A}v = v_1^T\mathbf{A}_{11}v_1 + v_1^T\mathbf{A}_{12}v_2 + v_2^T\mathbf{A}_{21}v_1 + v_2^T\mathbf{A}_{22}v_2 \geq 0$.
Let $v_2^T \triangleq -v_1^T\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$, then

$$\begin{aligned} v^T\mathbf{A}v &= v_1^T\mathbf{A}_{11}v_1 + v_1^T\mathbf{A}_{12}v_2 + v_2^T\mathbf{A}_{21}v_1 + v_2^T\mathbf{A}_{22}v_2 \\ &= v_1^T\mathbf{A}_{11}v_1 + v_1^T\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{22}v_2 + v_2^T\mathbf{A}_{21}v_1 + v_2^T\mathbf{A}_{22}v_2 \quad (\mathbf{A}_{22} \text{ is invertible}) \\ &= v_1^T\mathbf{A}_{11}v_1 - v_2^T\mathbf{A}_{22}v_2 + v_2^T\mathbf{A}_{21}v_1 + v_2^T\mathbf{A}_{22}v_2 \\ &= v_1^T\mathbf{A}_{11}v_1 - v_1^T\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}v_1 \geq 0, \forall v_1. \end{aligned}$$

As a consequence, $(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) \succeq 0$.

PCRB and the Kalman filter

Nonlinear time-varying dynamical system

$$\begin{aligned} X_{k+1} &= f_k(X_k) + W_k \\ Y_k &= h_k(X_k) + V_k, \end{aligned}$$

where

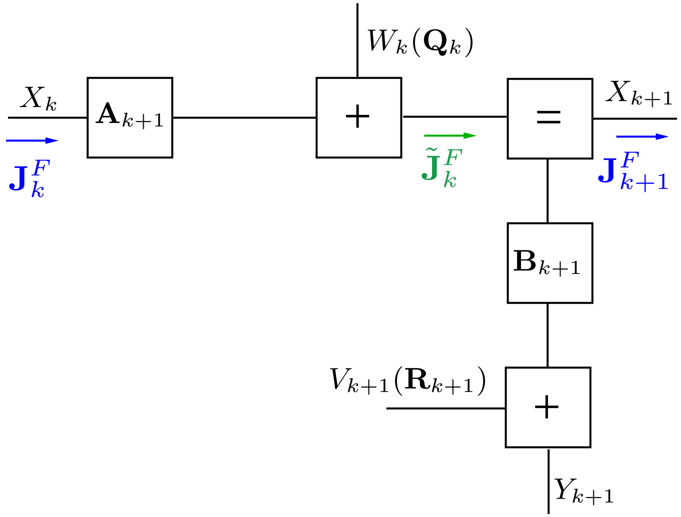
- f_k and h_k are in general non-linear functions
- W_k and V_k are i.i.d. (mean free) Gaussian random variables with covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ resp.

PCRB and the Kalman filter (2)

Nonlinear time-varying dynamical system (more specific case)

$$\begin{aligned}\mathbf{J}^F(X_{k+1}) &= \mathbf{Q}_k^{-1} - \mathbf{Q}_k^{-1}\mathrm{E}[\mathbf{A}_{k+1}](\mathbf{J}^F(X_k) + \mathrm{E}[\mathbf{A}_{k+1}^T\mathbf{Q}_k^{-1}\mathbf{A}_{k+1}])^{-1}\mathrm{E}[\mathbf{A}_{k+1}^T]\mathbf{Q}_k^{-1} \\ &\quad + \mathrm{E}[\mathbf{B}_{k+1}^T\mathbf{R}_{k+1}^{-1}\mathbf{B}_{k+1}] \\ &= \tilde{\mathbf{J}}_k^F + \mathrm{E}[\mathbf{B}_{k+1}^T\mathbf{R}_{k+1}^{-1}\mathbf{B}_{k+1}]\end{aligned}$$

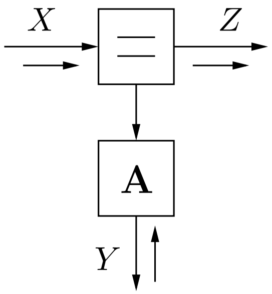
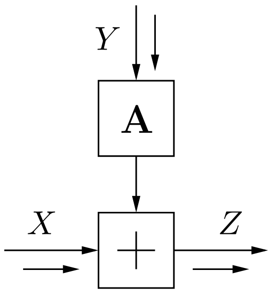
with $\mathbf{A}_{k+1}^T = \nabla_{x_k} f_k^T(x_k)$ and $\mathbf{B}_{k+1}^T = \nabla_{x_{k+1}} h_{k+1}^T(x_{k+1})$



Update equation for $\mathbf{J}_k^{F/B}$ similar to Kalman update of inverse covariance matrix

$\Rightarrow$ Kalman filter/smoother achieves PCRB for estimating the state of linear systems with Gaussian noise sources

PCRB and the Kalman filter (3)

Node	Update rule
	$m_Z = m_X + V_X A^T G (m_Y - A m_X)$ $V_Z = V_X - V_X A^T G A V_X$ $W_Z = W_X + A^T W_Y A$ $\xi_Z = \xi_X + A^T \xi_Y$ with $G = (V_Y + A V_X A^T)^{-1}$
	$m_Z = m_X + A m_Y$ $V_Z = V_X + A V_Y A^T$ $W_Z = W_X - W_X A H A^T W_X$ $\xi_Z = \xi_X + W_X A H (\xi_Y - A^T \xi_X)$ with $H = (W_Y + A^T W_X A)^{-1}$

Matrix inversion lemma

Let $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$, where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are nonsingular, such that $(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})$ and $(\mathbf{A}_{11} - \mathbf{A}_{22}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})$ are also nonsingular.

Then $\mathbf{A}$ is also nonsingular with

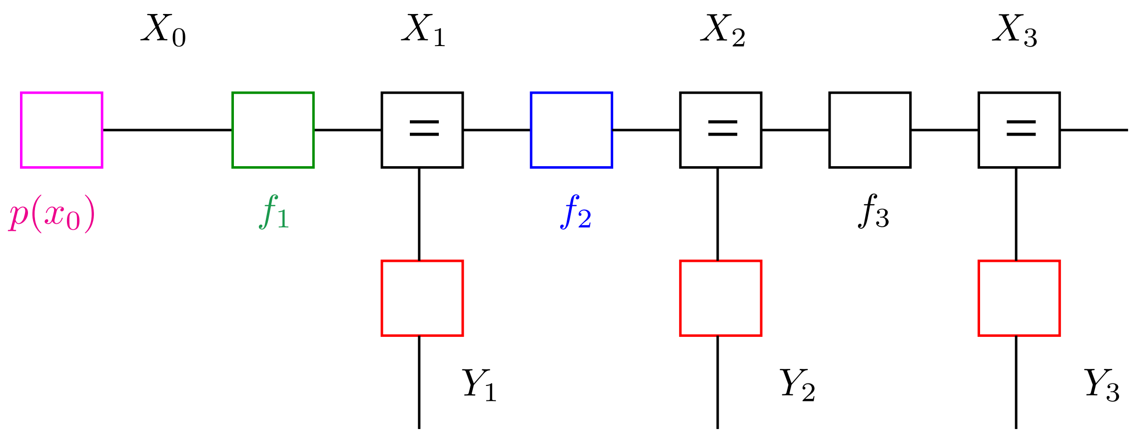
$$
\begin{aligned}
[\mathbf{A}^{-1}]_{11} &= \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \\
&= (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}
\end{aligned}
$$

$$
\begin{aligned}
[\mathbf{A}^{-1}]_{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \\
&= -(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}
\end{aligned}
$$

$$
\begin{aligned}
[\mathbf{A}^{-1}]_{21} &= -(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \\
&= -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}
\end{aligned}
$$

$$
\begin{aligned}
[\mathbf{A}^{-1}]_{22} &= \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\
&= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}.
\end{aligned}
$$

Markov model

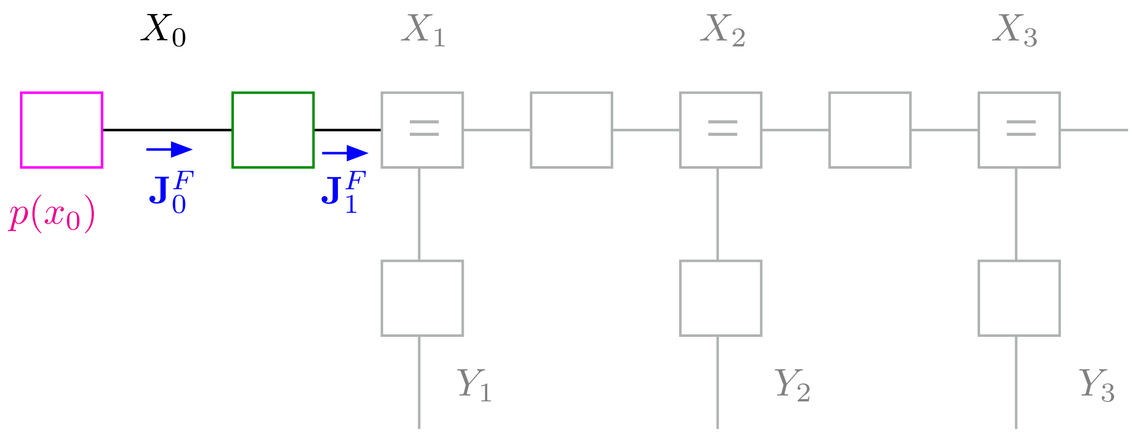


$$\mathrm{E}[(\hat{X}_2(Y) - X_2)^2] \succeq [\mathbf{J}^{-1}]_{22}$$

$$[\mathbf{J}^{-1}]_{22} = \left(\begin{bmatrix} f_0^{00} + f_1^{00} & f_1^{01} & 0 & 0 \\ f_1^{10} & f_1^{11} + f_2^{11} + f_{Y_1}^{11} & f_2^{21} & 0 \\ 0 & f_2^{21} & f_2^{22} + f_3^{22} + f_{Y_2}^{22} & f_3^{23} \\ 0 & 0 & f_3^{32} & f_3^{33} + f_{Y_3}^{33} \end{bmatrix}^{-1} \right)_{22}$$

with $f^{ij} \triangleq -\mathrm{E}[\nabla_{x_i} \nabla_{x_j}^T \log f]$

Markov model (2)

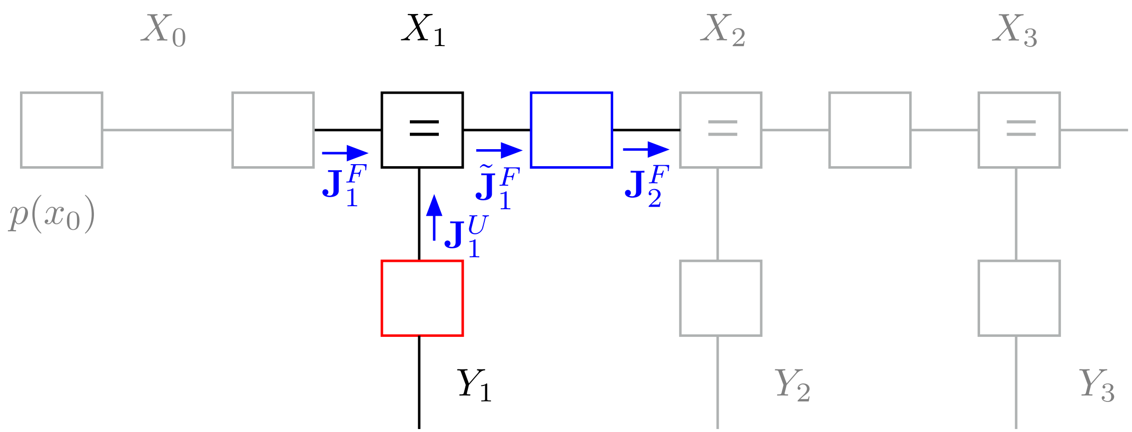


$$[\mathbf{J}^{-1}]_{22} = \left(\begin{bmatrix} f_0^{00} + f_1^{00} & f_1^{01} & 0 & 0 \\ f_1^{10} & f_1^{11} + f_2^{11} + f_{Y_1}^{11} & f_2^{21} & 0 \\ 0 & f_2^{21} & f_2^{22} + f_3^{22} + f_{Y_2}^{22} & f_3^{23} \\ 0 & 0 & f_3^{32} & f_3^{33} + f_{Y_3}^{33} \end{bmatrix}^{-1} \right)_{22}$$

$$= \left(\begin{bmatrix} f_2^{11} + f_{Y_1}^{11} + \mathbf{J}_1^F & f_2^{21} & 0 \\ f_2^{21} & f_2^{22} + f_3^{22} + f_{Y_2}^{22} & f_3^{23} \\ 0 & f_3^{32} & f_3^{33} + f_{Y_3}^{33} \end{bmatrix}^{-1} \right)_{11}$$

with $\mathbf{J}_1^F = \left(\left(\begin{bmatrix} f_0^{00} + f_1^{00} & f_1^{01} \\ f_1^{10} & f_1^{11} \end{bmatrix}^{-1} \right)_{11} \right)^{-1} = \left(\left(\begin{bmatrix} \mathbf{J}_0^F + f_1^{00} & f_1^{01} \\ f_1^{10} & f_1^{11} \end{bmatrix}^{-1} \right)_{11} \right)^{-1}$

Markov model (3)



$$[\mathbf{J}^{-1}]_{22} = \left(\begin{bmatrix} f_2^{11} + f_{Y_1}^{11} + \mathbf{J}_1^F & f_2^{21} & 0 \\ f_2^{21} & f_2^{22} + f_3^{22} + f_{Y_2}^{22} & f_3^{23} \\ 0 & f_3^{32} & f_3^{33} + f_{Y_3}^{33} \end{bmatrix}^{-1} \right)_{11}$$

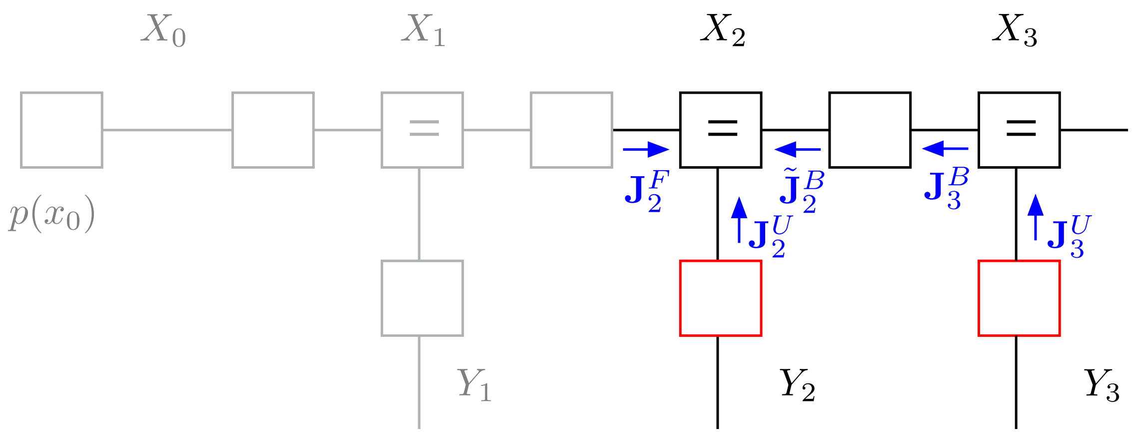
$$= \left(\begin{bmatrix} f_3^{22} + f_{Y_2}^{22} + \mathbf{J}_2^F & f_3^{23} \\ f_3^{32} & f_3^{33} + f_{Y_3}^{33} \end{bmatrix}^{-1} \right)_{00}$$

with $\mathbf{J}_2^F = \left(\left(\begin{bmatrix} f_2^{11} + \tilde{\mathbf{J}}_1^F & f_2^{12} \\ f_2^{21} & f_2^{22} \end{bmatrix}^{-1} \right)_{11} \right)^{-1}$ and $\tilde{\mathbf{J}}_1^F \triangleq \mathbf{J}_1^F + \mathbf{J}_1^U \triangleq \mathbf{J}_1^F + f_{Y_1}^{11}$

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

Markov model (4)



$$[\mathbf{J}^{-1}]_{22} = (\mathbf{J}_2^F + \mathbf{J}_2^U + \tilde{\mathbf{J}}_2^B)^{-1}$$

with $\tilde{\mathbf{J}}_2^B = \left(\left(\begin{bmatrix} f_3^{22} & f_2^{23} \\ f_3^{32} & f_3^{33} + \mathbf{J}_3^B \end{bmatrix}^{-1} \right)_{00} \right)^{-1}$

$\mathbf{J}_3^B \triangleq \mathbf{J}_3^U \triangleq f_{Y_3}^{33}$

$\mathbf{J}_2^U \triangleq f_{Y_2}^{22}$

Cramér-Rao bound

Information matrix $\mathbf{I}(\theta)$

$$I_{ij}(\theta) \triangleq \mathrm{E}_Y \left[\nabla_{\theta_i} \log p(y|\theta) \nabla^T_{\theta_j} \log p(y|\theta) \right].$$

Error matrix $\mathbf{E}(\theta)$

$$E_{ij}(\theta) \triangleq \mathrm{E}_Y \left[(\theta(y) - \theta)(\theta(y) - \theta)^T \right].$$

Theorem (Fisher, '22; Dugué, '37; Rao, '45; Cramér, '46)
If

1. the information matrix $\mathbf{I}(\theta)$ exists and is non-singular,
2. $\int_y (\hat{\theta}(y) - \theta) p(y|\theta) dy = 0$, i.e., the estimator $\hat{\theta}(y)$ is unbiased

then $\mathbf{E}(\theta) \succeq \mathbf{I}^{-1}(\theta)$ (Cramér-Rao bound).