

Computing Cramér-Rao bounds in graphical models

Brain-Style Information Systems Research Group,
RIKEN BSI, Saitama, Japan

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

- Cramér-Rao bounds (CRBs) are **lower bounds** on MSE of estimators.
- In the case of **(deterministic) parameter** estimation, the CRBs bound the MSE (mean squared error) of **unbiased** estimators.
- In the case of **random variables**, the CRBs bound the MSE of both **unbiased** and **biased** estimators.
- CRBs have been computed for estimators of **fixed** parameters in various contexts, e.g., communications and image processing.
- CRBs have been computed for **filtering** the **state** in state space models with **freely** evolving states [Tichavský et al., 1998], e.g., tracking of slowly varying parameters.
- Extension to
 - **cycle-free graphical models**, e.g., **filtering/smoothing** of **input/state** in **general** state space models
 - Particular **cyclic** graphical models, i.e., joint decoding and channel estimation
- In this talk, all proofs are omitted.

Overview

- Introduction to CRBs
- Matrix inversion lemma
- CRBs for estimation in cycle-free graphical models
- State space model
- CRBs for filtering
- CRBs for smoothing
- Examples
- Conclusion

Introduction to CRBs

- Let $p(x, y)$ be the **joint probability function** of X and Y , where $X^T \triangleq (X_1, X_2, \dots, X_N)$ and $X_i \in \mathbb{R}^l$;
thus $X^T = (X_{11}, X_{12}, \dots, X_{1l}, X_{21}, \dots, X_{2l}, \dots, X_{Nl})$.
- $\hat{X}(Y)$ is the **estimate** of X based on the observations Y .
- **Error matrix** $\mathbf{E} \triangleq \mathbf{E}_{XY} [(\hat{X}(Y) - X)(\hat{X}(Y) - X)^T]$.
- **Posterior information matrix** $\mathbf{J}(X)$ defined as

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

under the assumption that $\mathbf{E}_{XY} \left[\nabla_{x_i} \log p(x, y) \nabla_{x_j}^T \log p(x, y) \right]$ exists,
 $\forall x$ and y .

Introduction to CRBs (2)

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

Let $p(x, y)$ be the **joint probability function** of X and Y and \mathbf{E} the **error matrix**. If

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

then $\mathbf{E} \succeq \mathbf{J}(X)^{-1}$.

In words

$\mathbf{D} \triangleq \mathbf{E} - \mathbf{J}(X)^{-1}$ is **positive semi-definite**, i.e., $v^T \mathbf{D} v \geq 0, \forall v \in \mathbb{R}^{lN}$.

Notice that the estimators are not necessarily **unbiased**!

The assumption (3) is **weaker**.

Introduction to CRBs (5)

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

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

Introduction to CRBs (4)

Corollaries

Since $\mathbf{D} \triangleq \mathbf{E} - \mathbf{J}^{-1}$ is **positive semi-definite**, i.e., $x^T \mathbf{D} x \geq 0, \forall x \in \mathbb{R}^{lN}$, it follows

- $\mathbf{E}_{ii} = \mathbf{E}_{p(x_i, y)}[(\hat{X}_i(Y) - X_i)(\hat{X}_i(Y) - X_i)^T] \succeq [\mathbf{J}(X)^{-1}]_{ii}$

Proof: Set $x = [0, \dots, 0, x_i, 0]^T$.

- $E_{ij, ij} = \mathbf{E}_{p(x_{ij}, y)}[(\hat{X}_{ij}(Y) - X_{ij})^2] \geq [\mathbf{J}(X)^{-1}]_{ij, ij}$.

Proof: Set $x = [0, \dots, 0, x_{ij}, 0]^T$.

- $\text{Tr}[\mathbf{E} - \mathbf{J}^{-1}] \geq 0$, hence $\text{Tr}[\mathbf{E}] \triangleq \sum_{i,j} \mathbf{E}[(\hat{X}_{ij} - X_{ij})^2] \geq \text{Tr}[\mathbf{J}(X)^{-1}]$.

- $\sum_i \mathbf{E}[(\hat{X}_{ij} - X_{ij})^2] \geq \sum_i [\mathbf{J}(X)^{-1}]_{ij}$.

Observation

We need to **invert $\mathbf{J}(X)$** ! However, we only need the **diagonal elements** of $\mathbf{J}(X)^{-1}$.

If $\mathbf{J}(X)$ is **sparse**, the inversion can be done **efficiently** by **matrix inversion lemma**.

This is can be viewed as **message passing**.

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

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

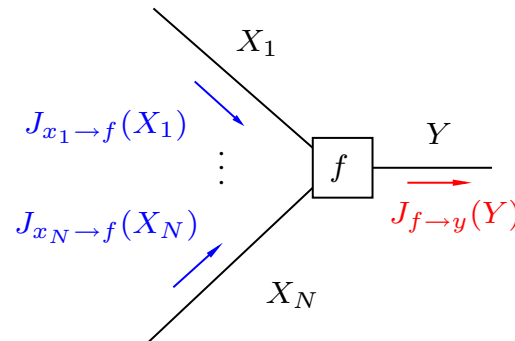
$$\begin{aligned} [\mathbf{A}^{-1}]_{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \\ &= -(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{aligned}$$

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

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

CRB for estimation in graphical models (2)

Update rule



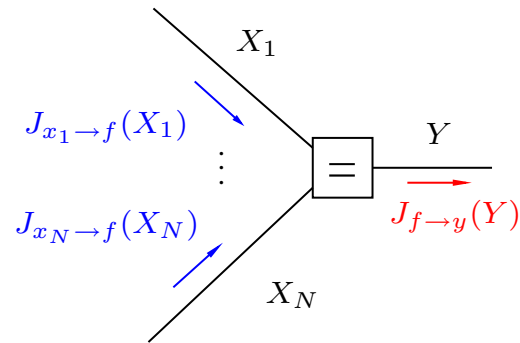
$$J_{f \rightarrow y}^{-1}(Y) = \left(\begin{bmatrix} J_{x_1 \rightarrow f}(X_1) + \mathbb{E}[-\Delta_{x_1}^{x_1} \log f] & \dots & \mathbb{E}[-\Delta_{x_1}^{x_N} \log f] & \mathbb{E}[-\Delta_{x_1}^y \log f] \\ \vdots & \dots & \dots & \vdots \\ \mathbb{E}[-\Delta_{x_N}^{x_1} \log f] & \dots & J_{x_N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

Remark:

- The expectations $\mathbb{E}[\Delta_{x_i}^{x_j} \log f]$ are supposed to be **computable**.
- One can exchange rows and corresponding columns in the above matrix, it leads to the same $J_{f \rightarrow y}(Y)$.

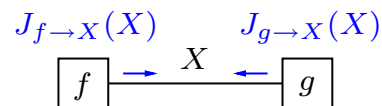
CRB for estimation in graphical models (3)

Update rule



$$J_{f \rightarrow y}(Y) = \sum_{i=1}^N J_{x_i \rightarrow f}(X_i)$$

Posterior CRB



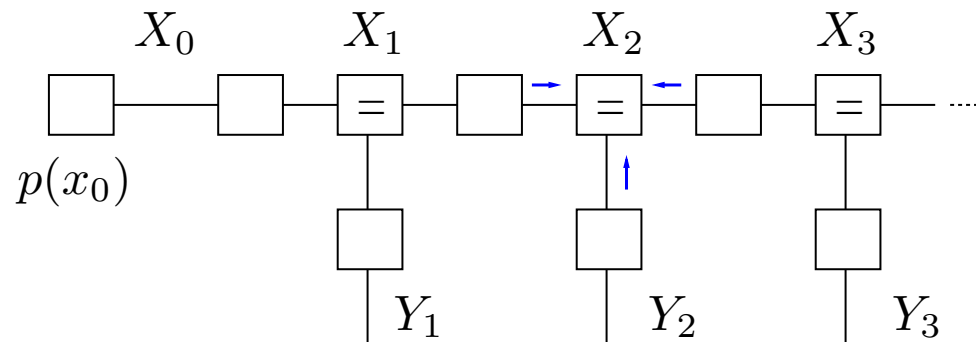
$$J(X) = J_{f \rightarrow X}(X) + J_{g \rightarrow X}(X)$$

State space model (I)

- Assume $p(x, y)$ factorizes as

$$p(x, y) = p(x_0) \prod_{k=1}^N p(x_k | x_{k-1}) p(y_k | x_k).$$

- Freely evolving state X .
- The marginals $p(x_k)$ can be computed by **message passing**.
- The **CRBs** can be computed similarly, i.e., by **forward and backward sweep!**



CRB for filtering

Theorem [Tichavský et al., 1998] (“forward sweep”)

$$\mathbf{J}(X_{k+1}) = \mathbf{J}^U(X_{k+1}) + \mathbf{D}_k^{22} - \mathbf{D}_k^{21}(\mathbf{J}(X_k) + \mathbf{D}_k^{11})^{-1}\mathbf{D}_k^{12},$$

where

$$\mathbf{D}_k^{11} = \mathbf{E}[-\Delta_{x_k}^{x_k} \log p(x_{k+1}|x_k)]$$

$$\mathbf{D}_k^{12} = [D_k^{21}]^T = \mathbf{E}[-\Delta_{x_k}^{x_{k+1}} \log p(x_{k+1}|x_k)]$$

$$\mathbf{D}_k^{22} = \mathbf{E}[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(x_{k+1}|x_k)]$$

$$\mathbf{J}^U(X_{k+1}) = \mathbf{E}[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(y_{k+1}|x_{k+1})]$$

$$\Delta_x^y = \nabla_x \nabla_y^T$$

$$J(X_k) \triangleq [J(X_1^k)^{-1}]_{kk}$$

$$X_s^t = (X_s, X_{s+1}, \dots, X_t) \quad (t > s; t, s \in \mathbb{N}).$$

CRB for smoothing

Forward Sweep:

$$\mathbf{J}^F(X_{k+1}) = \mathbf{J}^U(X_{k+1}) + \mathbf{D}_k^{22} - \mathbf{D}_k^{21}(\mathbf{J}^F(X_k) + \mathbf{D}_k^{11})^{-1}\mathbf{D}_k^{12} = \mathbf{J}^U(X_{k+1}) + \tilde{\mathbf{J}}^F(X_{k+1}).$$

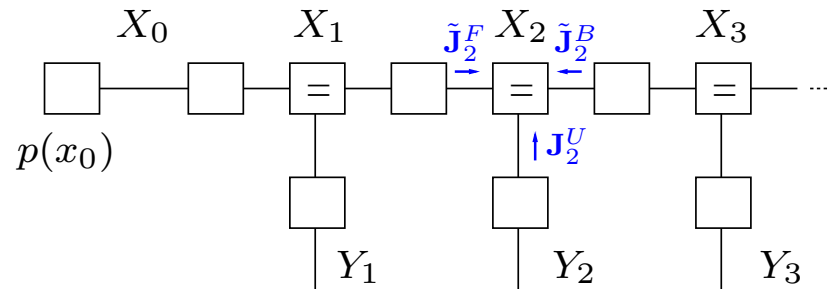
Initialization: $\mathbf{J}^F(X_0) = \mathbf{E}[-\Delta_{x_0}^{x_0} \log p(x_0)]$

Backward Sweep:

$$\mathbf{J}^B(X_{k-1}) = \mathbf{J}^U(X_{k-1}) + \mathbf{D}_{k-1}^{11} - \mathbf{D}_{k-1}^{12}(\mathbf{J}^B(X_k) + \mathbf{D}_{k-1}^{22})^{-1}\mathbf{D}_{k-1}^{21} = \mathbf{J}^U(X_{k-1}) + \tilde{\mathbf{J}}^B(X_{k-1}).$$

Initialization: $\mathbf{J}^B(X_N) = \mathbf{E}[-\Delta_{x_N}^{x_N} \log p(y_N|x_N)] = \mathbf{J}^U(X_N)$

Posterior CRB: $\mathbf{J}^{\text{tot}}(X_k) = \tilde{\mathbf{J}}^F(X_k) + \tilde{\mathbf{J}}^B(X_k) + \mathbf{J}^U(X_k).$

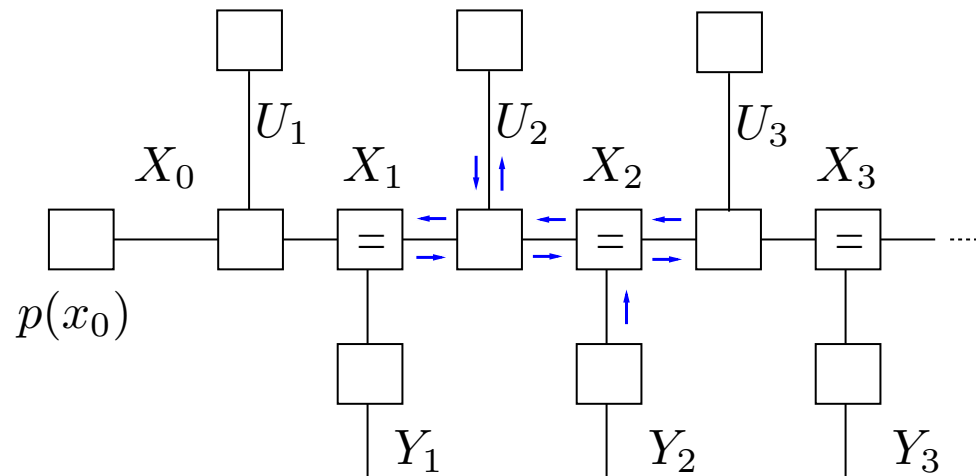


State space model (II)

- Assume $p(x, y, u)$ factorizes as

$$p(x, y, u) = p(x_0) \prod_{k=1}^N p(x_k | u_k, x_{k-1}) p(y_k | x_k) p(u_k).$$

- CRB for input U_k and/or state X_k .
- The marginals $p(u_k)$ and $p(x_k)$ can be computed by **message passing**.
- The **CRBs** can be computed similarly, i.e., by **forward and backward sweep**!



CRB for filtering

Theorem (“forward sweep”)

$\mathbf{J}(X_{k+1}) = \mathbf{J}^U(X_{k+1}) + \boldsymbol{\zeta}_k^{22} - \boldsymbol{\zeta}_k^{21}(\boldsymbol{\zeta}_k^{11})^{-1}\boldsymbol{\zeta}_k^{12}$, where

$$\boldsymbol{\zeta}_k^{11} = \mathbf{J}^D(U_{k+1}) + \mathbf{D}_k^{22} - \mathbf{D}_k^{21}(\mathbf{J}(X_k) + \mathbf{D}_k^{11})^{-1}\mathbf{D}_k^{21}$$

$$\boldsymbol{\zeta}_k^{12} = \mathbf{D}_k^{23} - \mathbf{D}_k^{21}(\mathbf{J}(X_k) + \mathbf{D}_k^{11})^{-1}\mathbf{D}_k^{31}$$

$$\boldsymbol{\zeta}_k^{21} = \mathbf{D}_k^{32} - \mathbf{D}_k^{31}(\mathbf{J}(X_k) + \mathbf{D}_k^{11})^{-1}\mathbf{D}_k^{21} = (\boldsymbol{\zeta}_k^{12})^T$$

$$\boldsymbol{\zeta}_k^{22} = \mathbf{D}_k^{33} - \mathbf{D}_k^{31}(\mathbf{J}(X_k) + \mathbf{D}_k^{11})^{-1}\mathbf{D}_k^{31}$$

$$\mathbf{D}_k^{11} = \mathbb{E}[-\Delta_{x_k}^{x_k} \log p(x_{k+1} | u_{k+1}, x_k)]$$

$$\mathbf{D}_k^{12} = \mathbb{E}[-\Delta_{x_k}^{u_{k+1}} \log p(x_{k+1} | u_{k+1}, x_k)] = (\mathbf{D}_k^{21})^T$$

$$\mathbf{D}_k^{13} = \mathbb{E}[-\Delta_{x_k}^{x_{k+1}} \log p(x_{k+1} | u_{k+1}, x_k)] = (\mathbf{D}_k^{31})^T$$

$$\mathbf{D}_k^{22} = \mathbb{E}[-\Delta_{u_{k+1}}^{u_{k+1}} \log p(x_{k+1} | u_{k+1}, x_k)]$$

$$\mathbf{J}^D(U_{k+1}) \triangleq \mathbb{E}[-\Delta_{u_{k+1}}^{u_{k+1}} \log p(u_{k+1})]$$

$$\mathbf{D}_k^{23} = \mathbb{E}[-\Delta_{u_{k+1}}^{x_{k+1}} \log p(x_{k+1} | u_{k+1}, x_k)] = (\mathbf{D}_k^{32})^T$$

$$\mathbf{D}_k^{33} = \mathbb{E}[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(x_{k+1} | u_{k+1}, x_k)]$$

$$\mathbf{J}^U(X_{k+1}) \triangleq \mathbb{E}[-\Delta_{x_{k+1}}^{x_{k+1}} \log p(y_{k+1} | x_{k+1})]$$

CRB for smoothing

Forward sweep

$$\mathbf{J}^F(X_{k+1}) = \mathbf{J}^U(X_{k+1}) + \boldsymbol{\zeta}_k^{22} - \boldsymbol{\zeta}_k^{21}(\boldsymbol{\zeta}_k^{11})^{-1}\boldsymbol{\zeta}_k^{12} = \mathbf{J}^U(X_{k+1}) + \tilde{\mathbf{J}}^F(X_{k+1})$$

Initialization

$$\mathbf{J}^F(X_0) = \mathbf{E}[-\Delta_{x_0}^{x_0} \log p(x_0)]$$

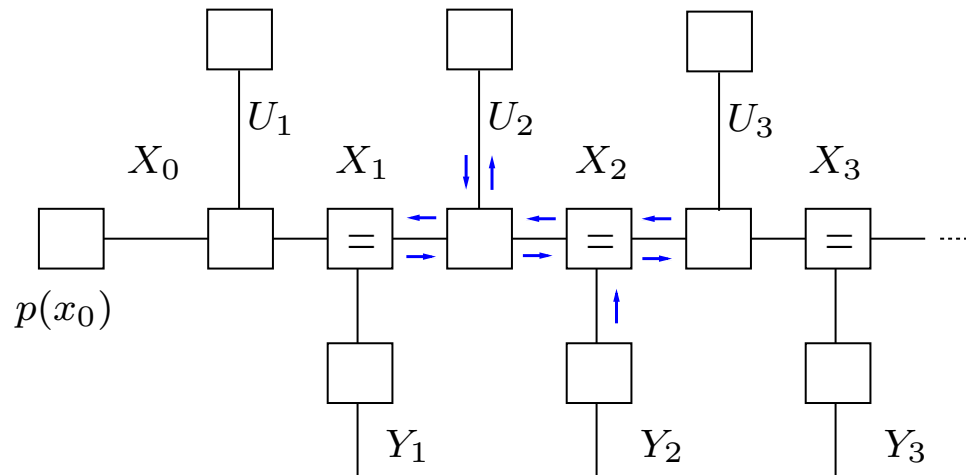
Backward sweep

$$\mathbf{J}^B(X_k) = \mathbf{J}^U(X_k) + \boldsymbol{\rho}_k^{11} - \boldsymbol{\rho}_k^{12}(\boldsymbol{\rho}_k^{22})^{-1}\boldsymbol{\rho}_k^{21} = \mathbf{J}^U(X_k) + \tilde{\mathbf{J}}^B(X_k)$$

Initialization

$$\mathbf{J}(X_N) = \mathbf{J}^U(X_N)$$

Posterior CRB: $\mathbf{J}^{\text{tot}}(X_k) = \tilde{\mathbf{J}}^F(X_k) + \tilde{\mathbf{J}}^B(X_k) + \mathbf{J}^U(X_k)$.



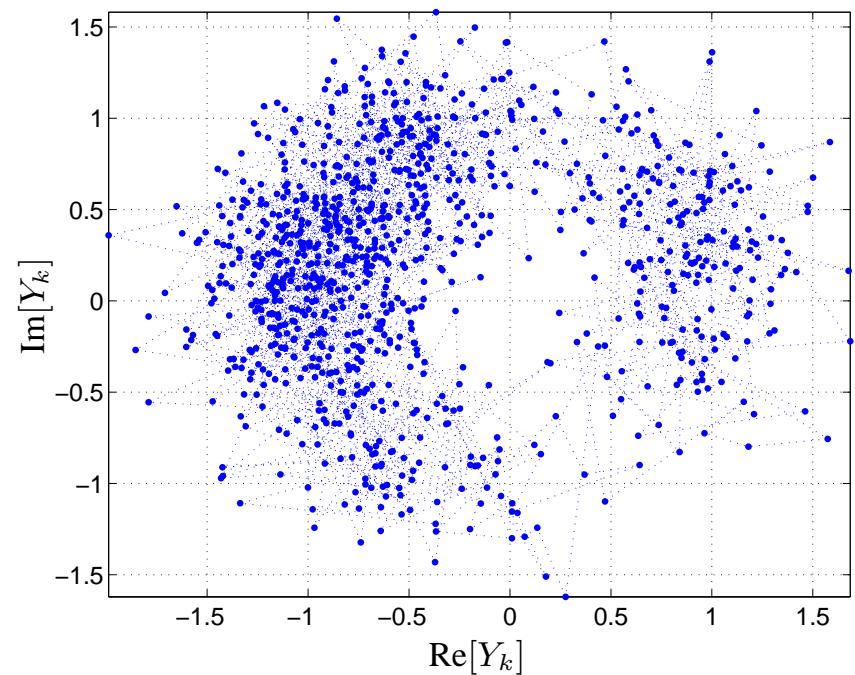
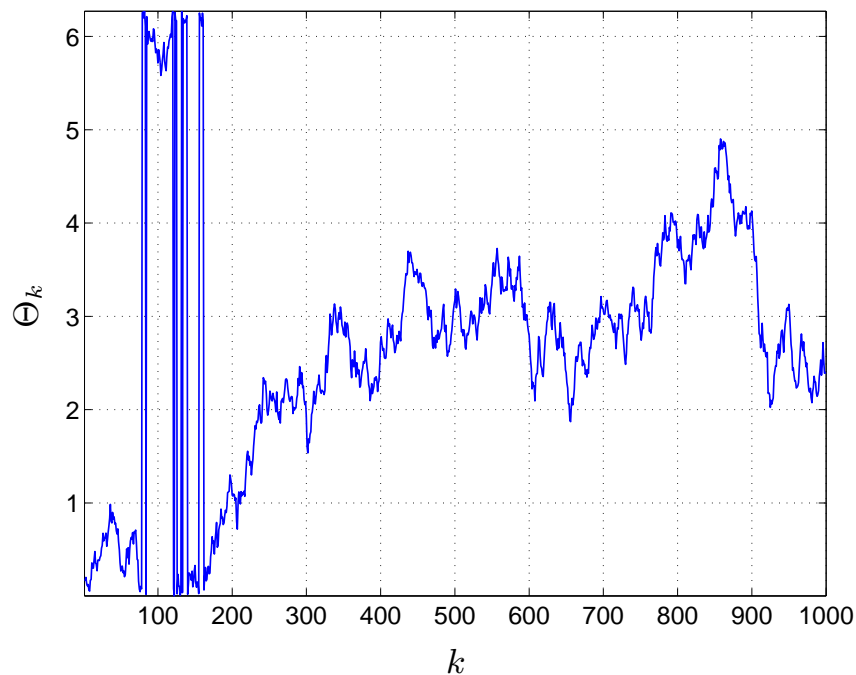
Example 1

Random walk phase model

$$\Theta_{k+1} = (\Theta_k + W_k) \bmod 2\pi$$

$$Y_k = \exp(j\Theta_k) + V_k,$$

where W_k and V_k are i.i.d. (mean free) Gaussian RVs with variance σ_θ^2 and $2\sigma_0^2$ resp.



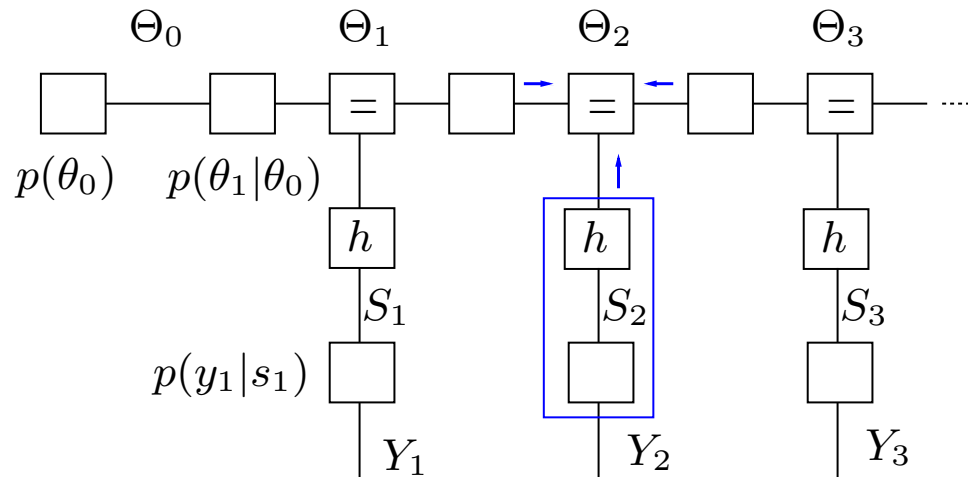
Example 1 (2)

Random walk phase model

$$\Theta_{k+1} = (\Theta_k + W_k) \bmod 2\pi$$

$$Y_k = \exp(j\Theta_k) + V_k,$$

where W_k and V_k are i.i.d. (mean free) Gaussian RVs with variance σ_θ^2 and $2\sigma_0^2$ resp.



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

$$h(\theta_k, s_k) \triangleq \delta(s_k - \exp(j\theta_k))$$

$$p(y_k | s_k) \triangleq (2\pi\sigma_N^2)^{-1} \exp(-\|y_k - s_k\|^2 / 2\sigma_N^2)$$

$$p(\theta_k | \theta_{k-1}) \triangleq (2\pi\sigma_W^2)^{-1/2} \sum_{n \in \mathbb{Z}} \exp(-((\theta_k - \theta_{k-1}) + n2\pi)^2 / 2\sigma_W^2).$$

Example 1 (3)

Random walk phase model

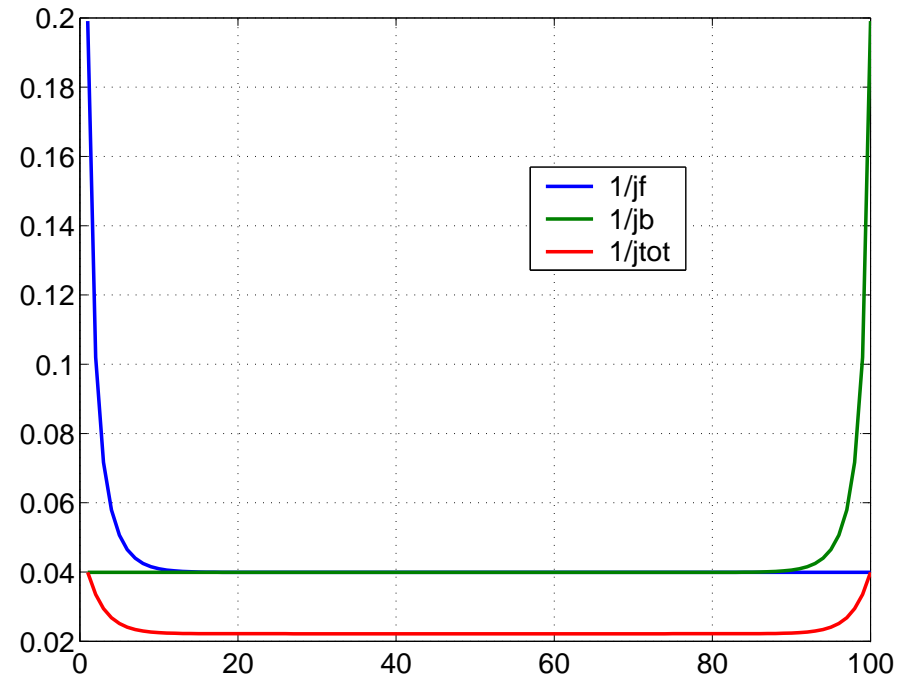
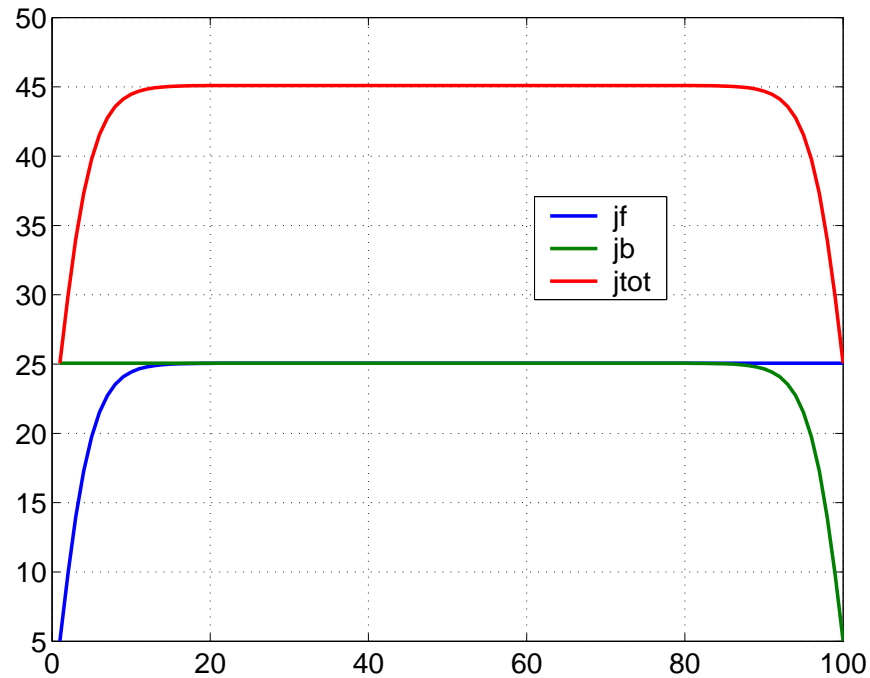
$$\begin{aligned}\Theta_{k+1} &= (\Theta_k + W_k) \bmod 2\pi \\ Y_k &= \exp(j\Theta_k) + V_k,\end{aligned}$$

where W_k and V_k are i.i.d. (mean free) Gaussian RVs with variance σ_θ^2 and $2\sigma_0^2$ resp.

$$\begin{aligned}J^F(\Theta_{k+1}) &= J^U(\Theta_{k+1}) + D_k^{22} - D_k^{21}(J^F(\Theta_k) + D_k^{11})^{-1}D_k^{12} \\ &= \frac{1}{\sigma_0^2} + \frac{1}{\sigma_\phi^2} - \frac{1}{\sigma_\phi^4} \left(J^F(\Theta_k) + \frac{1}{\sigma_\phi^2} \right)^{-1} \triangleq \tilde{J}^F(\Theta_{k+1}) + \frac{1}{\sigma_0^2} \\ J^B(\Theta_k) &= J^U(\Theta_k) + D_k^{11} - D_k^{12}(J^B(\Theta_{k+1}) + D_k^{11})^{-1}D_k^{21} \\ &= \frac{1}{\sigma_0^2} + \frac{1}{\sigma_\phi^2} - \frac{1}{\sigma_\phi^4} \left(J^B(\Theta_{k+1}) + \frac{1}{\sigma_\phi^2} \right)^{-1} \triangleq \tilde{J}^B(\Theta_{k+1}) + \frac{1}{\sigma_0^2}.\end{aligned}$$

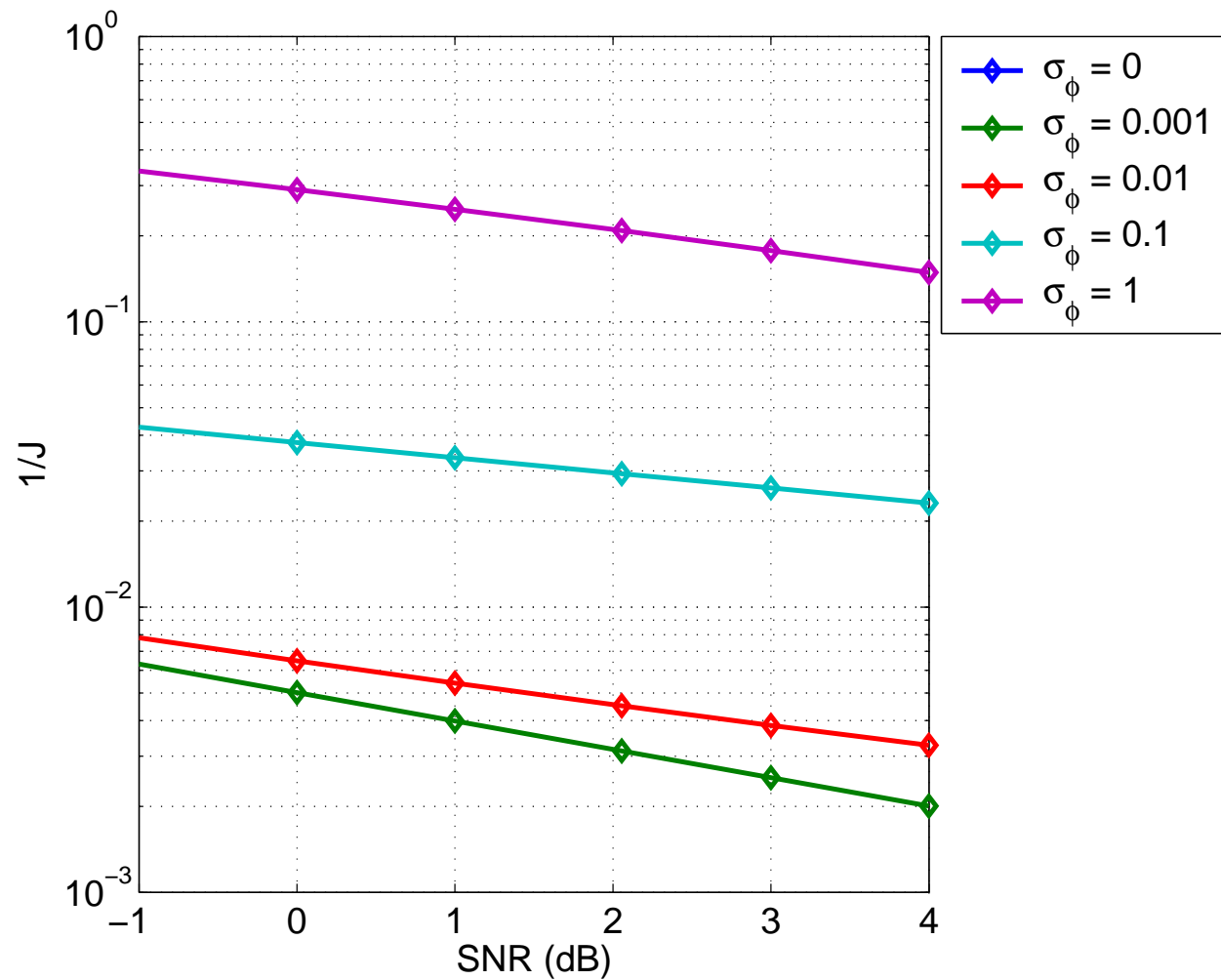
Example 1 (4)

Random walk phase model ($N = 100, \sigma_0^2 = 0.1991, \sigma_\phi^2 = 0.01$)



Example 1 (5)

Random walk phase model



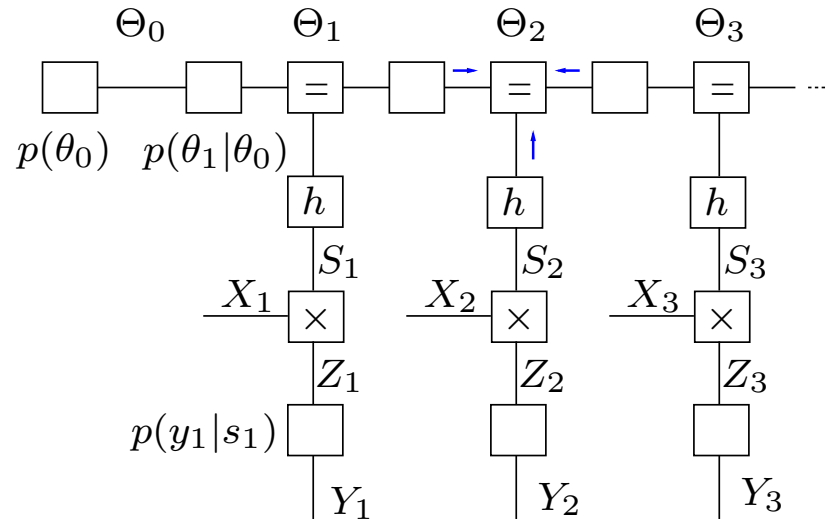
Example 2

Code + Random walk phase model

$$\Theta_{k+1} = (\Theta_k + W_k) \bmod 2\pi$$

$$Y_k = X_k \exp(j\Theta_k) + V_k,$$

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



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

$$h(\theta_k, s_k) \triangleq \delta(s_k - \exp(j\theta_k))$$

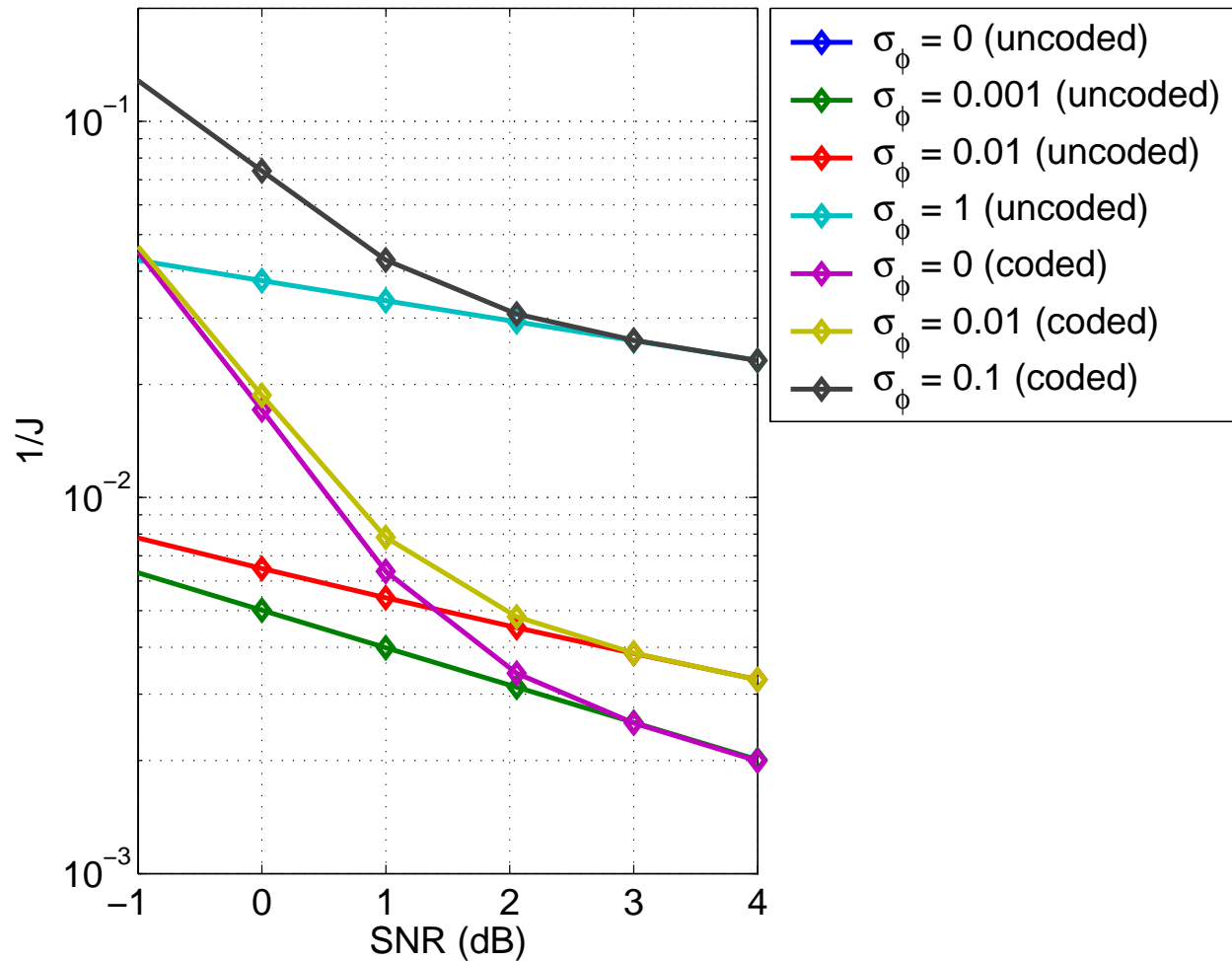
$$p(y_k | s_k) \triangleq (2\pi\sigma_N^2)^{-1} \exp(-\|y_k - s_k\|^2 / 2\sigma_N^2)$$

$$p(\theta_k | \theta_{k-1}) \triangleq (2\pi\sigma_W^2)^{-1/2} \sum_{n \in \mathbb{Z}} \exp(-((\theta_k - \theta_{k-1}) + n2\pi)^2 / 2\sigma_W^2)$$

$$h_\times(s_k, z_k) = \delta(z_k - x_k s_k).$$

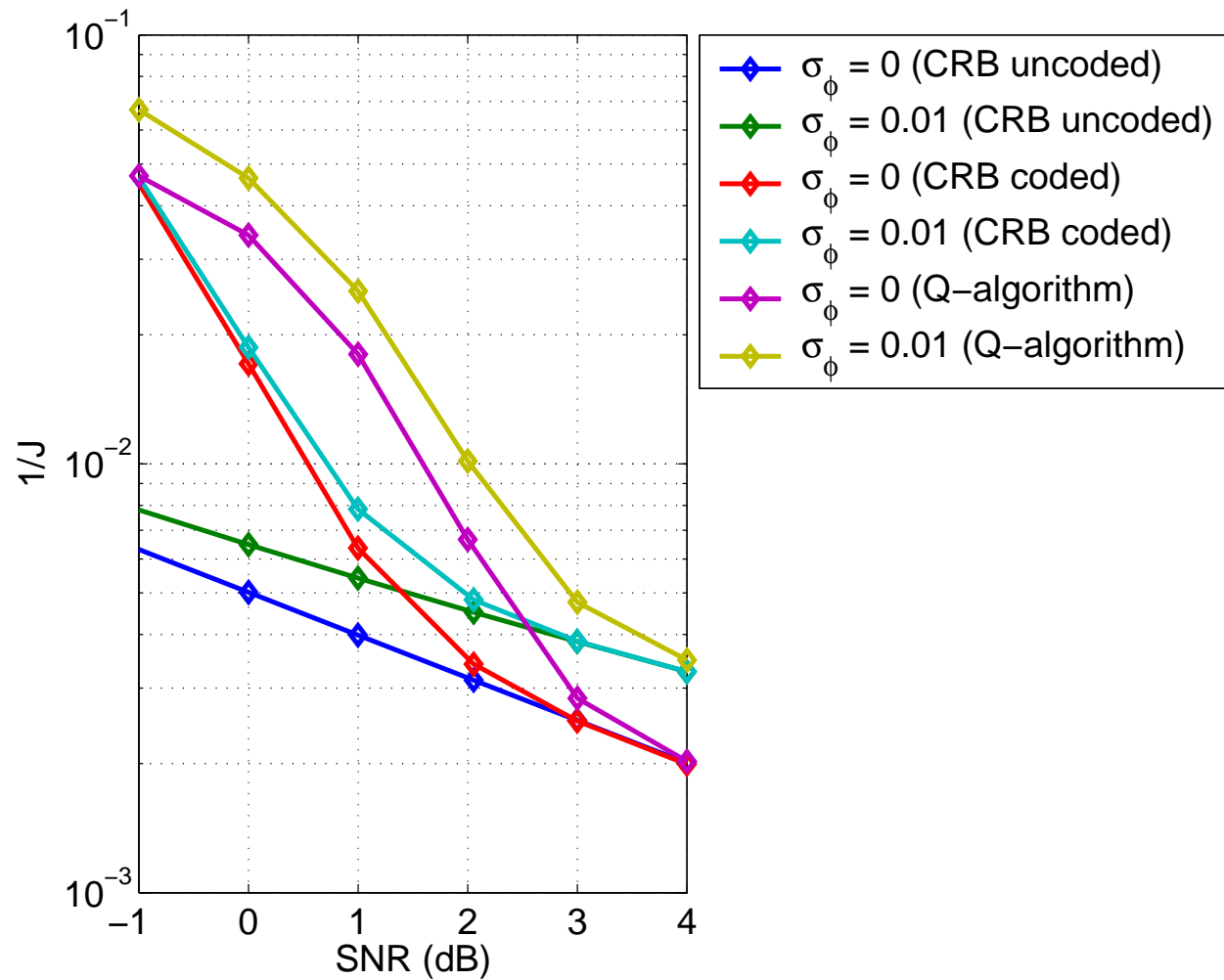
Example 2 (2)

Code + Random walk phase model (3-6 LDPC of length 100)



Example 2 (3)

Code + Random walk phase model (3-6 LDPC of length 100)



Conclusion

- CRBs have previously been computed for
 - parameter estimation
 - filtering (of states).
- We extended this to
 - estimation on trees
e.g., filtering and smoothing of input and/or state.
 - estimation on certain cyclic graphs, i.e., code + channel.
e.g., code + random walk phase model.

Thank you for your attention!

Computing Cramér-Rao bounds in graphical models

Brain-Style Information Systems Research Group,
RIKEN BSI, Saitama, Japan

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