Workshop report

1. Daniels report is on website
2. Don’t expect to write it based on listening to one project (we had 6 only 2 was sufficient quality)
3. I suggest writing it on one presentation.
4. Include figures (from a related paper or their presentation)
5. Include references

Update: We are all set to have your students attend. We will not register them, so they can come and go as needed. food is for the registered participants and please allow them to eat first. Currently we have 70 registered participants and plant to order food for ~100.

May 22, Dictionary learning, Mike Bianco (half class), Bishop Ch 13

May 24, **Class HW** Bishop Ch 8/13

MAY 30 CODY

May 31, No Class. Workshop, Big Data and The Earth Sciences: Grand Challenges Workshop

June 5, Discuss workshop, Discuss final project. **Spiess Hall open for project discussion 11am-7pm.**

June 7, **Workshop report. No class**

June 12 Spiess Hall open for project discussion 9-11:30am and 2-7pm

June 16 Final report delivered. Beer time

For final project discussion **every** afternoon Mark and I will be available

Final Report

In class on July 5 a status report from each group is mandatory. Maximum 2min/person, (i.e. a 5-member group have 10min), shorter is fine. Have presentation on memory stick or email Mark. Class might run longer, so we could start earlier.

For the Final project (Due 16 June 5Pm). Delivery Dropbox request <2GB (details to follow).:

A) Deliver a code:

1) Assume we have reasonable compilers installed (we use Mac OsX)
2) Give instructions if any additional software should be installed.
3) You can ask us to download a dataset. Or include it in this submission
4) Don't include all developed codes. Just key elements.
5) We should not have to reprogram your code.

B) Report

1) The report should include all the following sections: Summary -> Introduction->Physical and Mathematical framework->Results.
2) Summary is a combination of an abstract and conclusion.
3) Plagiarism is not acceptable! When citing use " " for quotes and citations for relevant papers.
4) Don't write anything you don't understand.
5) Everyone in the group should understand everything that is written. If we do not understand a section during grading we should be able to ask any member of the group to clarify. You can delegate the writing, but not the understanding.
6) Use citations. Any concepts which are not fully explained should have a citation with an explanation.
7) Please be concise. Equations are good. Figures essential. Write as though your report is to be published in a scientific journal.
8) I have attached a sample report from Mark, though shorter is preferred.

Discrete Variables (1)

General joint distribution: K^2-1 parameters

$$p(\mathbf{x}_1, \mathbf{x}_2 | \boldsymbol{\mu}) = \prod_{k=1}^{K} \prod_{l=1}^{K} \mu_{kl}^{x_{1k} x_{2l}}$$

Independent joint distribution: 2(K-1) parameters

$$\hat{p}(\mathbf{x}_1, \mathbf{x}_2 | \boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_{1k}^{x_{1k}} \prod_{l=1}^{K} \mu_{2l}^{x_{2l}}$$

General joint distribution over M variables: K^M -1parameters

M-node Markov chain: K-1 + (M-1) K(K-1) parameters

Joint Distribution

Where $\psi_C(\mathbf{x}_C)$ is the potential over clique C and

$$p(\mathbf{x}) = \frac{1}{Z} \prod_C \psi_C(\mathbf{x}_C)$$

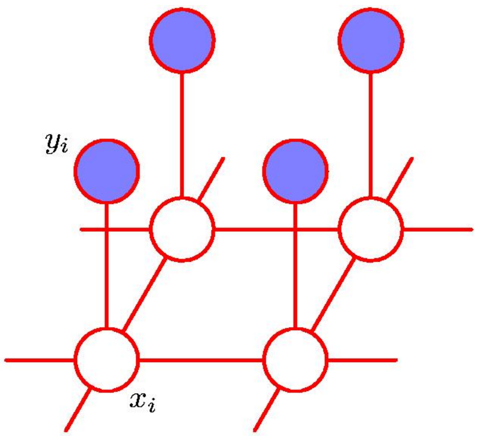
is the normalization coefficient; note: M K-state variables $\rightarrow$ K^M terms in Z.

$$Z = \sum_{\mathbf{x}} \prod_C \psi_C(\mathbf{x}_C)$$

Energies and the Boltzmann distribution

$$\psi_C(\mathbf{x}_C) = \exp\left\{-E(\mathbf{x}_C)\right\}$$

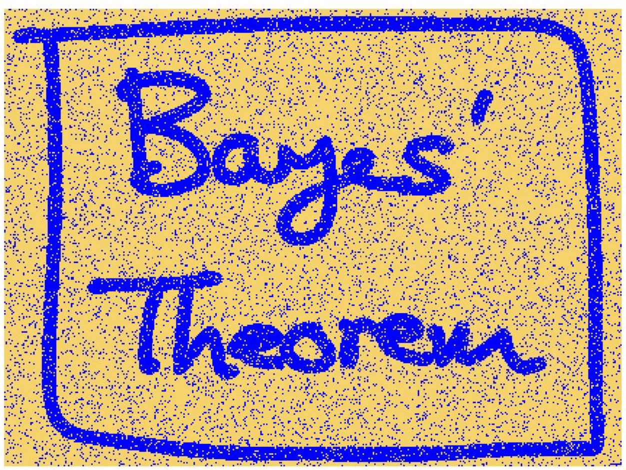
Illustration: Image De-Noising



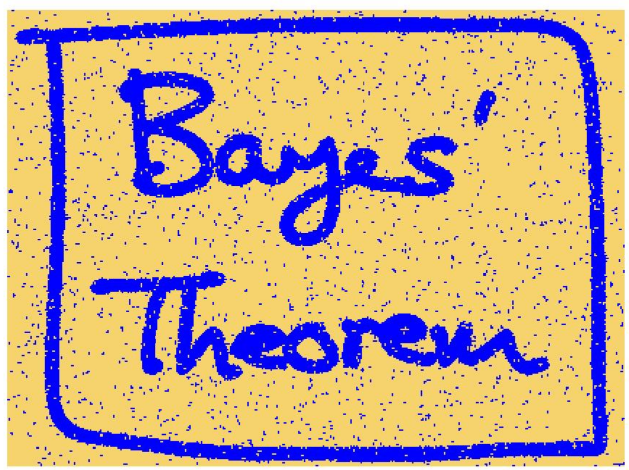
$$E(\mathbf{x}, \mathbf{y}) = h \sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \exp\{-E(\mathbf{x}, \mathbf{y})\}$$

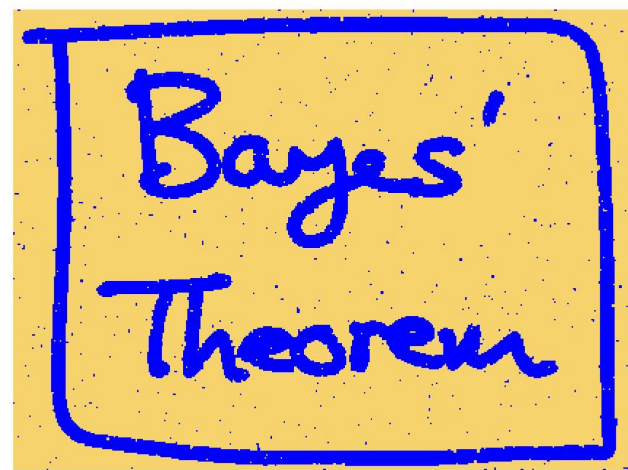
Noisy Image



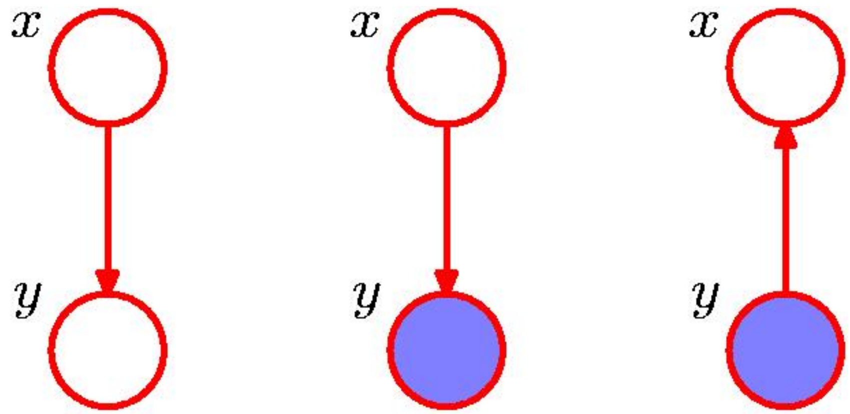
Restored Image (ICM)



Restored Image (Graph cuts)



Inference in Graphical Models



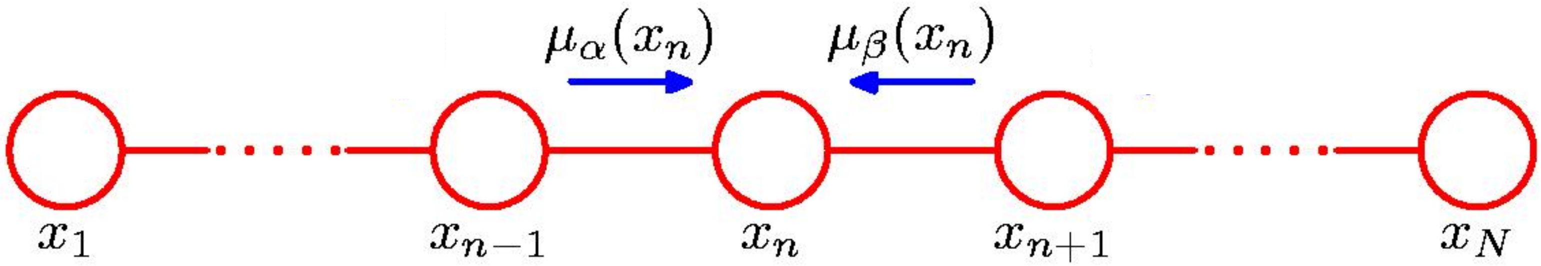
$$p(y) = \sum_{x'} p(y|x')p(x') \qquad\qquad p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Inference on a Chain

$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N)$$

$$p(x_n) = \sum_{x_1}\cdots\sum_{x_{n-1}}\sum_{x_{n+1}}\cdots\sum_{x_N} p(\mathbf{x})$$

Inference on a Chain



$$p(x_n) = \frac{1}{Z} \underbrace{\left[\sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \cdots \left[\sum_{x_1} \psi_{1,2}(x_1, x_2) \right] \cdots \right]}_{\mu_\alpha(x_n)}$$

$$\underbrace{\left[\sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N) \right] \cdots \right]}_{\mu_\beta(x_n)}$$

Inference on a Chain

To compute local marginals:

- Compute and store all forward messages, $\mu_\alpha(x_n)$
- Compute and store all backward messages, $\mu_\beta(x_n)$
- Compute Z at any node x_m
- Compute

$$p(x_n) = \frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n)$$

  for all variables required.

The Sum-Product Algorithm (1)

Objective:

i. to obtain an efficient, exact inference algorithm for finding marginals;

ii. in situations where several marginals are required, to allow computations to be shared efficiently.

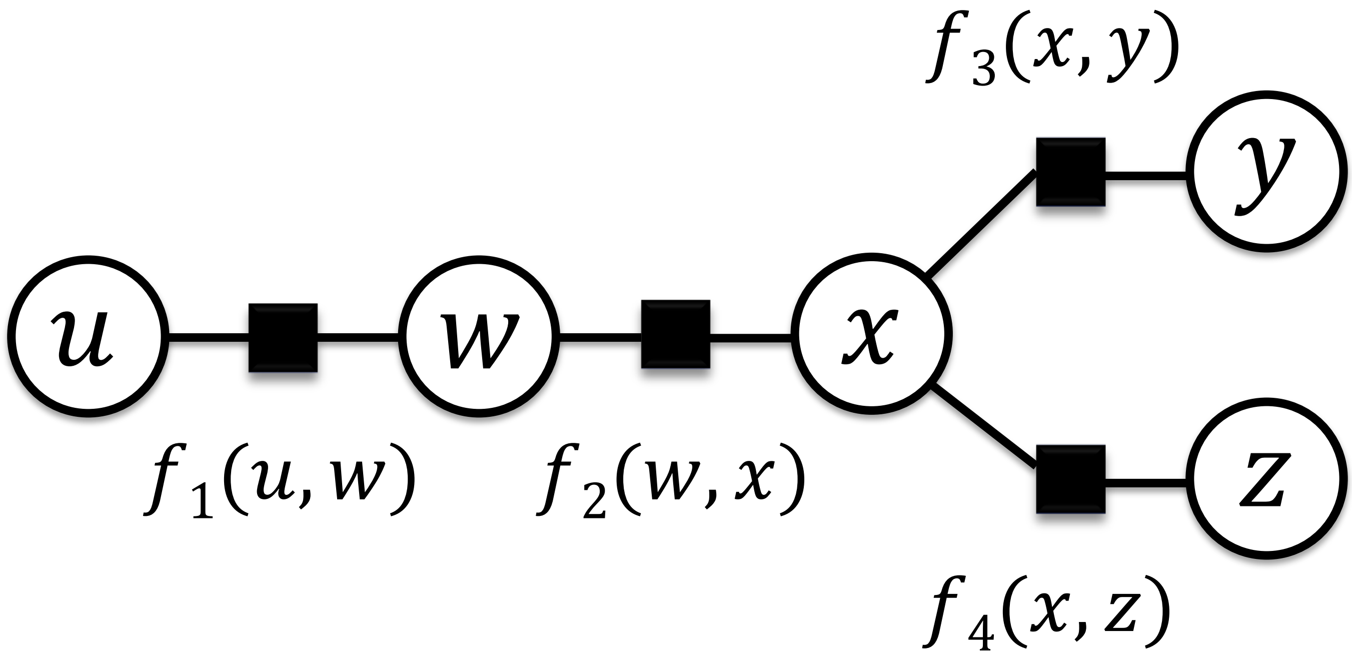
Key idea: Distributive Law

$$\sum_x \sum_y xy = x_1 y_1 + x_2 y_1 + x_1 y_2 + x_2 y_2$$

$$= (x_1 + x_2)(y_1 + y_2)$$

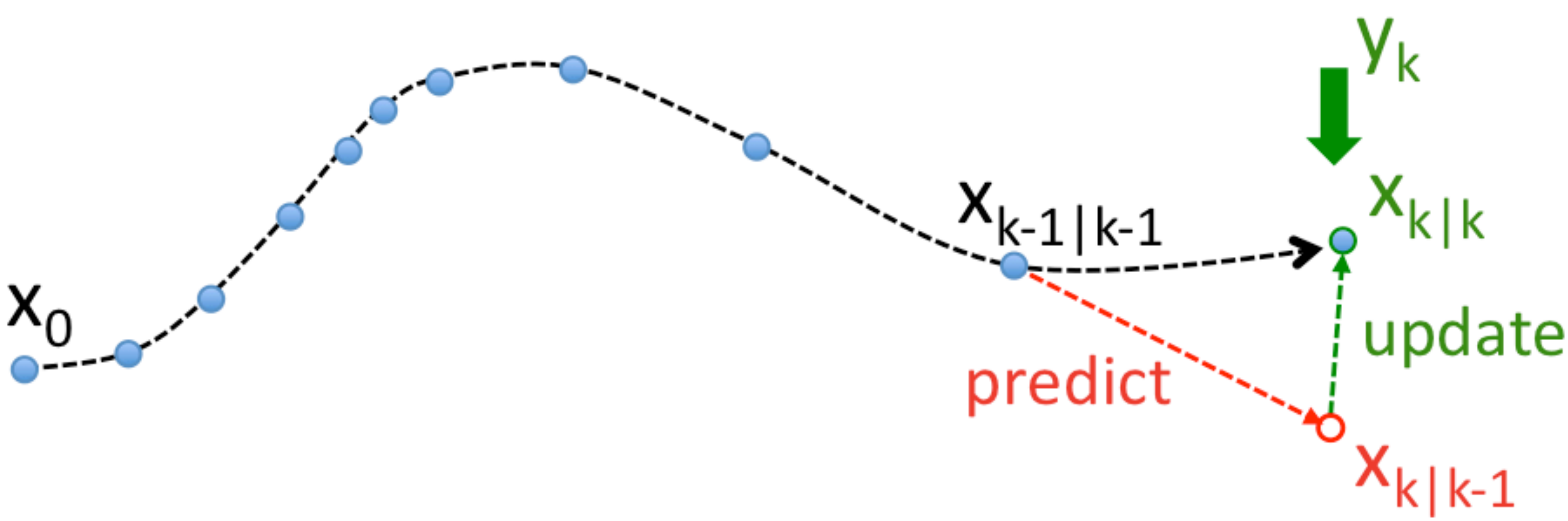
7 versus 3 operations

The Sum-Product Algorithm



How do we solve it and what does the solution look like?

KF/PFs offer solutions to dynamical systems, nonlinear in general, using prediction and update as data becomes available. Tracking in time or space offers an ideal framework for studying KF/PF.



Kalman Framework

$\mathbf{x}_k = f_{k-1}(\mathbf{x}_{k-1}, \mathbf{v}_k)$ state equation

$\mathbf{y}_k = h_k(\mathbf{x}_k, \mathbf{w}_k)$ measurement equation

↓

$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{v}_k$ state equation

$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{w}_k$ measurement equation

$\mathbf{x}_k$, $\mathbf{y}_k$, $\mathbf{v}_k$, $\mathbf{w}_k$: Gaussian

$\mathbf{F}_k$, $\mathbf{H}_k$: Linear

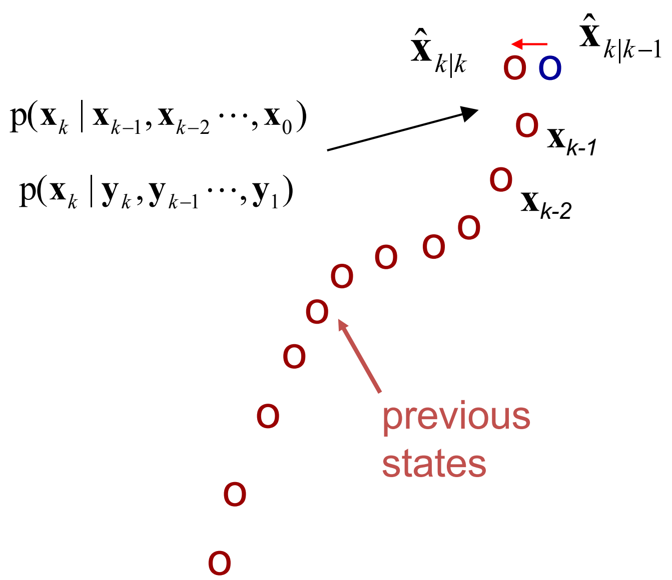
Optimal Filter = Kalman Filter

1963

A Single Kalman Iteration

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{v}_k$$
$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{w}_k$$

$$\mathbf{x}_{k|k} \sim \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$$



PREDICT

1. Predict the mean $\hat{\mathbf{x}}_{k|k-1}$ using previous history.

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$$

$$\hat{\mathbf{x}}_{k|k-1} = \mathrm{E}\{\mathbf{x}_k \mid \mathbf{x}_{k-1}\} = \int \mathbf{x}_k \, p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) d\mathbf{x}_k$$

2. Predict the covariance $\mathbf{P}_{k|k-1}$ using previous history.

UPDATE

3. Correct/update the mean using new data $\mathbf{y}_k$

$$p(\mathbf{x}_k \mid \mathbf{Y}_k)$$

$$\hat{\mathbf{x}}_{k|k} = \mathrm{E}\{\mathbf{x}_k \mid \mathbf{Y}_k\} = \int \mathbf{x}_k \, p(\mathbf{x}_k \mid \mathbf{Y}_k) d\mathbf{x}_k$$

4. Correct/update the covariance $\mathbf{P}_{k|k}$ using $\mathbf{y}_k$

$$\cdots \Rightarrow p(\mathbf{x}_{k-1} \mid \mathbf{Y}_{k-1}) \Rightarrow p(\mathbf{x}_k \mid \mathbf{Y}_{k-1}) \Rightarrow p(\mathbf{x}_k \mid \mathbf{Y}_k) \Rightarrow \cdots$$

PREDICTOR-CORRECTOR

DENSITY PROPAGATOR

The Model

Consider the discrete, linear system,

$$\mathbf{x}_{k+1} = \mathbf{M}_k \mathbf{x}_k + \mathbf{w}_k, \quad k = 0, 1, 2, \ldots, \qquad (1)$$

where

- $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector at time t_k
- $\mathbf{M}_k \in \mathbb{R}^{n \times n}$ is the state transition matrix (mapping from time t_k to t_{k+1}) or model
- $\{\mathbf{w}_k \in \mathbb{R}^n; k = 0, 1, 2, \ldots\}$ is a white, Gaussian sequence, with $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$, often referred to as model error
- $\mathbf{Q}_k \in \mathbb{R}^{n \times n}$ is a symmetric positive definite covariance matrix (known as the model error covariance matrix).

Some of the following slides are from: Sarah Dance, University of Reading

The Observations

We also have discrete, linear observations that satisfy

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \quad k = 1, 2, 3, \dots, \tag{2}$$

where

- $\mathbf{y}_k \in \mathbb{R}^p$ is the vector of actual measurements or observations at time t_k
- $\mathbf{H}_k \in \mathbb{R}^{n \times p}$ is the observation operator. Note that this is not in general a square matrix.
- $\{\mathbf{v}_k \in \mathbb{R}^p; k = 1, 2, \dots\}$ is a white, Gaussian sequence, with $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$, often referred to as observation error.
- $\mathbf{R}_k \in \mathbb{R}^{p \times p}$ is a symmetric positive definite covariance matrix (known as the observation error covariance matrix).

We assume that the initial state, $\mathbf{x}_0$ and the noise vectors at each step, $\{\mathbf{w}_k\}$, $\{\mathbf{v}_k\}$, are assumed mutually independent.

The Prediction and Filtering Problems

We suppose that there is some uncertainty in the initial state, i.e.,

$$\mathbf{x}_0 \sim N(0, \mathbf{P}_0) \tag{3}$$

with $\mathbf{P}_0 \in \mathbb{R}^{n \times n}$ a symmetric positive definite covariance matrix.

The problem is now to compute an improved estimate of the stochastic variable $\mathbf{x}_k$, provided $\mathbf{y}_1, \ldots \mathbf{y}_j$ have been measured:

$$\widehat{\mathbf{x}}_{k|j} = \widehat{\mathbf{x}}_{k|y_1,\ldots,y_j}. \tag{4}$$

- When $j = k$ this is called the filtered estimate.
- When $j = k - 1$ this is the one-step predicted, or (here) the predicted estimate.

- The Kalman filter (Kalman, 1960) provides estimates for the linear discrete prediction and filtering problem.
- We will take a minimum variance approach to deriving the filter.
- We assume that all the relevant probability densities are Gaussian so that we can simply consider the mean and covariance.
- Rigorous justifcation and other approaches to deriving the filter are discussed by Jazwinski (1970), Chapter 7.

Prediction step

We first derive the equation for one-step prediction of the mean using the state propagation model (1).

$$\begin{aligned} \widehat{\mathbf{x}}_{k+1|k} &= \mathbb{E}\left[\mathbf{x}_{k+1}|\mathbf{y}_1, \ldots \mathbf{y}_k\right], \\ &= \mathbb{E}\left[\mathbf{M}_k \mathbf{x}_k + \mathbf{w}_k\right], \\ &= \mathbf{M}_k \widehat{\mathbf{x}}_{k|k} \end{aligned} \tag{5}$$

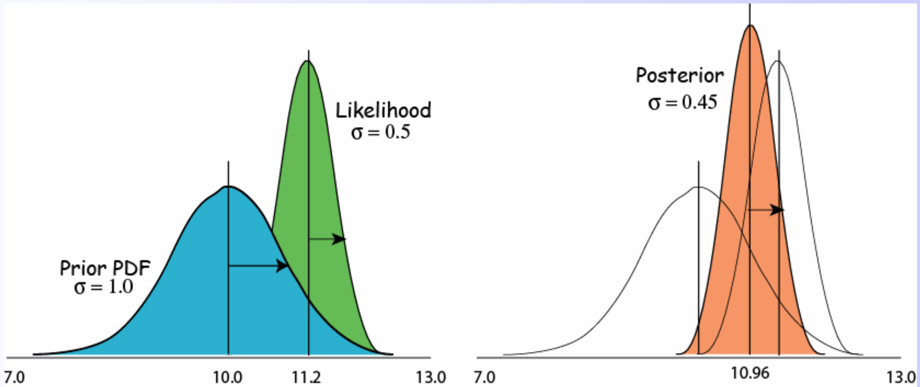
The one step prediction of the covariance is defined by,

$$\mathbf{P}_{k+1|k} = \mathbb{E}\left[(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1|k})^T|\mathbf{y}_1, \ldots \mathbf{y}_k\right]. \quad (6)$$

Exercise: Using the state propagation model, (1), and one-step prediction of the mean, (5), show that

$$\mathbf{P}_{k+1|k} = \mathbf{M}_k\mathbf{P}_{k|k}\mathbf{M}_k^T + \mathbf{Q}_k. \quad (7)$$

Product of Gaussians=Gaussian:



One data point problem

For the general linear inverse problem we would have

Prior: $$p(\boldsymbol{m}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{m}-\boldsymbol{m}_o)^T C_m^{-1}(\boldsymbol{m}-\boldsymbol{m}_o)\right\}$$

Likelihood: $$p(\boldsymbol{d}|\boldsymbol{m}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{d}-G\boldsymbol{m})^T C_d^{-1}(\boldsymbol{d}-G\boldsymbol{m})\right\}$$

Posterior PDF

$$\propto \exp\left\{-\frac{1}{2}[(\boldsymbol{d}-G\boldsymbol{m})^T C_d^{-1}(\boldsymbol{d}-G\boldsymbol{m}) + (\boldsymbol{m}-\boldsymbol{m}_o)^T C_m^{-1}(\boldsymbol{m}-\boldsymbol{m}_o)]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\mathbf{m}-\hat{\mathbf{m}}\right]^T \mathbf{S}^{-1}\left[\mathbf{m}-\hat{\mathbf{m}}\right]\right\}$$

$$\mathbf{S}^{-1} = \mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}$$

$$\hat{\mathbf{m}} = \left(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}\right)^{-1}\left(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{d} + \mathbf{C}_m^{-1}\mathbf{m}_0\right)$$

$$= \mathbf{m}_0 + \left(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}\right)^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}\left(\mathbf{d} - \mathbf{G}\mathbf{m}_0\right)$$

Filtering Step

At the time of an observation, we assume that the update to the mean may be written as a linear combination of the observation and the previous estimate:

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}), \tag{8}$$

where $\mathbf{K}_k \in \mathbb{R}^{n \times p}$ is known as the Kalman gain and will be derived shortly.

But first we consider the covariance associated with this estimate:

$$\mathbf{P}_{k|k} = \mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})^T | \mathbf{y}_1, \ldots \mathbf{y}_k\right]. \tag{9}$$

Using the observation update for the mean (8) we have,

$$\begin{aligned}
\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k} &= \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}) \\
&= \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}), \\
&\quad \text{replacing the observations with their model equivalent,} \\
&= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1}) - \mathbf{K}_k\mathbf{v}_k. \qquad (10)
\end{aligned}$$

Thus, since the error in the prior estimate, $\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1}$ is uncorrelated with the measurement noise we find

$$\begin{aligned}
\mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1})^T\right](\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T \\
&\quad + \mathbf{K}_k\mathbb{E}\left[\mathbf{v}_k\mathbf{v}_k^T\right]\mathbf{K}_k^T. \qquad (11)
\end{aligned}$$

Simplification of the a posteriori error covariance formula

Using this value of the Kalman gain we are in a position to simplify the Joseph form as

$$\mathbf{P}_{k|k} = (\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T = (\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}. \tag{15}$$

Exercise: Show this.

Note that the covariance update equation is independent of the actual measurements: so $\mathbf{P}^{k|k}$ could be computed in advance.

Summary of the Kalman filter

Prediction step

Mean update: $\widehat{\mathbf{x}}_{k+1|k} = \mathbf{M}_k \widehat{\mathbf{x}}_{k|k}$

Covariance update: $\mathbf{P}_{k+1|k} = \mathbf{M}_k \mathbf{P}_{k|k} \mathbf{M}_k^T + \mathbf{Q}_k.$

Observation update step

Mean update: $\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k \widehat{\mathbf{x}}_{k|k-1})$

Kalman gain: $\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}^T + \mathbf{R}_k)^{-1}$

Covariance update: $\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}.$