

Systems Dynamics

Course ID: 267MI – Fall 2021

Thomas Parisini
Gianfranco Fenu

University of Trieste
Department of Engineering and Architecture



267MI –Fall 2021

Lecture 13

**State Estimation from Observed
Data**

13. State Estimation from Observed Data

13.1 Kalman Estimation

13.1.1 Bayes Estimation in the Gaussian Case

13.1.2 State Estimation in the Bayes Estimation Framework

13.1.3 Kalman Predictor

13.2 Kalman Estimation: Generalizations and Steady-state Estimator

13.2.1 r -steps Ahead Kalman Prediction

13.2.2 Kalman Filter

13.2.3 Kalman Predictor in the Presence of Exogenous Inputs

13.2.4 Steady-state Kalman Estimator

13.2.5 Observability

13.2.6 Reachability

13.2.7 Example of Riccati Equation Convergence Analysis

Kalman Estimation

Kalman Estimation

Bayes Estimation in the Gaussian Case

Recall the basic facts about Bayes estimation

- We look for an estimation method allowing to **embed the possible *a-priori* knowledge on the unknown quantity** to be estimated
- In the framework of Bayes estimation also **the unknown vector ϑ is interpreted as a random vector**
- The probability density function $p(\vartheta)$ **in absence of observed data** is the *a-priori* probability density function embedding the available information on ϑ before collecting the data. Hence, in absence of data, **the *a-priori* estimator** could be

$$\hat{\vartheta} = \mathbb{E}(\vartheta) = \int \vartheta p(\vartheta) d\vartheta$$

and the estimate uncertainty $\text{var}(\hat{\vartheta})$ would be the ***a-priori* uncertainty**.

Recall the basic facts about Bayes estimation (cont.)

- Clearly, as soon as new data are collected, the probability density function $p(\vartheta)$ changes. As a consequence, $E(\vartheta)$ and $\text{var}(\vartheta)$ change as well. In particular, we expect $\text{var}(\vartheta)$ to decrease.
- **Summing up**, the basic idea is to consider a **joint random experiment** with respect to ϑ and to d and this is the conceptual peculiarity of the Bayes estimation approach.

Recall the basic facts about Bayes estimation (cont.)

- Consider the generic estimator as function of the data

$$\hat{\vartheta} = h(d)$$

and define the cost **functional**

$$J[h(\cdot)] = \mathbb{E} \left[\|\vartheta - h(d)\|^2 \right]$$

- The goal is to determine an estimator $h^\circ(\cdot)$ such that $J[h(\cdot)]$ is minimized, that is we have to determine

$$h^\circ(\cdot) : \mathbb{E} \left[\|\vartheta - h^\circ(d)\|^2 \right] \leq \mathbb{E} \left[\|\vartheta - h(d)\|^2 \right], \quad \forall h(\cdot)$$

where **the expected values are computed with reference to the joint random experiment.**

Recall the basic facts about Bayes estimation (cont.)

Assuming for the moment that ϑ and d are scalar

$$h^\circ(x) = E(\vartheta | d = x)$$

The optimal Bayes estimator is the expected value conditioned to the actual observed data

and thus $\hat{\vartheta} = h^\circ(\delta)$, where δ is the specific value taken on by d in the random experiment.

Remark. The generalization to the vector case is trivial.

Bayes Estimation in the Gaussian Case

- Assume that d and ϑ are **marginally and jointly Gaussian** random variables:

$$\begin{bmatrix} d \\ \vartheta \end{bmatrix} \sim G \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_{dd} & \lambda_{d\vartheta} \\ \lambda_{\vartheta d} & \lambda_{\vartheta\vartheta} \end{bmatrix} \right)$$

and

$$p(d, \vartheta) = C \exp \left(-\frac{1}{2} [d \ \vartheta] \begin{bmatrix} \lambda_{dd} & \lambda_{d\vartheta} \\ \lambda_{\vartheta d} & \lambda_{\vartheta\vartheta} \end{bmatrix}^{-1} \begin{bmatrix} d \\ \vartheta \end{bmatrix} \right)$$

- We obtain:

$p(\vartheta | d)$ is Gaussian with:

- expected value $\frac{\lambda_{\vartheta d}}{\lambda_{dd}} d$
- variance $\lambda^2 = \lambda_{\vartheta\vartheta} - \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}}$


Bayes Estimation in the Gaussian Case (cont.)

- Then, the optimal Bayes estimator is given by

$$\hat{\vartheta} = h^{\circ}(x) = E(\vartheta | d = x) = \frac{\lambda_{\vartheta d}}{\lambda_{dd}} d$$

Recalling that $E(d) = 0$, $E(\vartheta) = 0$ by assumption, we obtain that $E(\hat{\vartheta}) = 0$ and hence the variance of the optimal estimator is

$$\begin{aligned}\text{var}(\vartheta - \hat{\vartheta}) &= E\left[(\vartheta - \hat{\vartheta})^2\right] = E\left[\left(\vartheta - \frac{\lambda_{\vartheta d}}{\lambda_{dd}} d\right)^2\right] \\ &= E(\vartheta^2) - 2 \frac{\lambda_{\vartheta d}}{\lambda_{dd}} E(\vartheta d) + \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}^2} E(d^2)\end{aligned}$$


$$\text{var}(\vartheta - \hat{\vartheta}) = \lambda_{\vartheta\vartheta} - \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}} = \lambda^2$$

Optimal Linear Estimator


- Let us **remove the assumption** for which d and ϑ are marginally and jointly Gaussian random variables, and let us just assume that $E(d) = 0$, $E(\vartheta) = 0$
- As before, let us use the notations $E(d^2) = \lambda_{dd}$, $E(\vartheta^2) = \lambda_{\vartheta\vartheta}$, $E(\vartheta d) = \lambda_{\vartheta d}$
- **Impose** that the estimator takes on a **linear structure**:

$$\hat{\vartheta} = \alpha d + \beta$$

where α and β are suitable parameters to be determined.

- Introduce the cost function:

$$J = E \left[\left(\vartheta - \hat{\vartheta} \right)^2 \right] = E \left[\left(\vartheta - \alpha d - \beta \right)^2 \right]$$


$$\hat{\vartheta} = \frac{\lambda_{\vartheta d}}{\lambda_{dd}} d$$

Optimal linear estimator

Optimal Linear Estimator (cont.)

- The variance of the optimal linear estimator is given by:

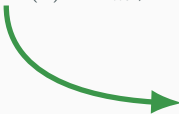
$$\begin{aligned}\text{var}(\vartheta - \hat{\vartheta}) &= \text{E} \left[(\vartheta - \hat{\vartheta})^2 \right] = \lambda_{\vartheta\vartheta} + \alpha^2 \lambda_{dd} + \beta^2 - 2\alpha \lambda_{\vartheta d} \\ &= \lambda_{\vartheta\vartheta} + \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}} \lambda_{dd} - 2 \frac{\lambda_{\vartheta d}}{\lambda_{dd}} \lambda_{\vartheta d} = \lambda_{\vartheta\vartheta} - \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}} = \lambda^2\end{aligned}$$

Therefore:

- the optimal linear estimator is **formally** equal to the Bayes one.
- If the Gaussian assumption on the random variables holds, then the optimal linear estimator actually is the best possible in the minimum variance sense
- If the Gaussian assumption on the random variables does not hold, then the linear estimator is sub-optimal, but still it is the best estimator constrained to take on a linear structure in the case in which no further assumptions are introduced on the probabilistic characteristics of the random variables


Generalizations

- If $E(d) = d_m$, $E(\vartheta) = \vartheta_m$


$$\hat{\vartheta} = \vartheta_m + \frac{\lambda_{\vartheta d}}{\lambda_{dd}} (d - d_m)$$
$$\text{var}(\vartheta - \hat{\vartheta}) = \lambda_{\vartheta\vartheta} - \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}}$$

- If d and ϑ are vectors with $E(d) = d_m$, $E(\vartheta) = \vartheta_m$ and

$$\text{var} \left(\begin{bmatrix} d \\ \vartheta \end{bmatrix} \right) = \begin{bmatrix} \Lambda_{dd} & \Lambda_{d\vartheta} \\ \Lambda_{\vartheta d} & \Lambda_{\vartheta\vartheta} \end{bmatrix} \quad \Lambda_{d\vartheta} = \Lambda_{\vartheta d}^\top$$


$$\hat{\vartheta} = \vartheta_m + \Lambda_{\vartheta d} \Lambda_{dd}^{-1} (d - d_m)$$
$$\text{var}(\vartheta - \hat{\vartheta}) = \Lambda_{\vartheta\vartheta} - \Lambda_{\vartheta d} \Lambda_{dd}^{-1} \Lambda_{d\vartheta}$$

Interpretations and remarks

- Consider for simplicity the Bayes estimator in the simple case:

$$\hat{\vartheta} = \vartheta_m + \frac{\lambda_{\vartheta d}}{\lambda_{dd}} (d - d_m)$$

Then:

- $\vartheta_m = \mathbb{E}(\vartheta)$ is the **a priori estimate**: in case of no observations availability, it is the more reasonable estimate. In this case, we have:

$$\hat{\vartheta} = \vartheta_m \quad \text{var}(\vartheta - \hat{\vartheta}) = \lambda_{\vartheta\vartheta} = \text{var}(\vartheta)$$

- Instead, when observations are available, we have:

$$\hat{\vartheta} = \vartheta_m + \frac{\lambda_{\vartheta d}}{\lambda_{dd}} (d - d_m)$$

**A-priori part
of the estimate**

**Correction term
exploiting
observed data**

Interpretations and remarks (cont.)

Clearly:

- If $\lambda_{\vartheta d} = 0$ then $\hat{\vartheta} = \vartheta_m$ and this is correct: it means that the data observation d is uncorrelated with ϑ and hence it does not convey useful information for the estimate: *the a-posteriori estimate coincides with the a-priori one.*
- If $\lambda_{\vartheta d} \neq 0$ then *the estimate is corrected on the basis of the observed data:*
 - If $\lambda_{\vartheta d} > 0$ then $\hat{\vartheta} - \vartheta_m$ and $d - d_m$ *in the average* keep the same sign and the correction is more likely to keep the same sign as well
 - If $\lambda_{\vartheta d} < 0$ then $\hat{\vartheta} - \vartheta_m$ and $d - d_m$ *in the average* have a different sign and the correction is more likely to change the same sign as well

Interpretations and remarks (cont.)

- It also very important to enhance the role played by the variance λ_{dd} that “quantifies” the degree of **uncertainty of the observed data**:

$$\hat{\vartheta} = \vartheta_m + \frac{\lambda_{\vartheta d}}{\lambda_{dd}} (d - d_m)$$

the larger λ_{dd} , the smaller the applied correction, that is, **the update is “more cautious”**

- Moreover:

$$\text{var}(\vartheta - \hat{\vartheta}) = \lambda_{\vartheta\vartheta} - \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}} = \lambda_{\vartheta\vartheta} \left(1 - \frac{\lambda_{\vartheta d}^2}{\lambda_{\vartheta\vartheta}\lambda_{dd}}\right)$$

and thus $\text{var}(\vartheta - \hat{\vartheta}) \leq \text{var}(\vartheta)$ and

$$\text{var}(\vartheta - \hat{\vartheta}) < \text{var}(\vartheta) \text{ if } \lambda_{\vartheta d} \neq 0$$

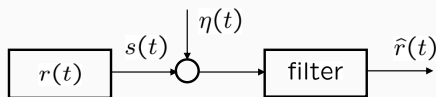
and this is correct because it expresses the fact that the estimate cannot but improve whenever the observed data convey useful information

Kalman Estimation

State Estimation in the Bayes Estimation Framework

Kalman estimation

- In Kalman estimation we address the problem of estimating variables that are **not directly available** and **without making any assumption on the stationarity of the stochastic processes** (unlike what has been done since now).



Example:
signal filtering

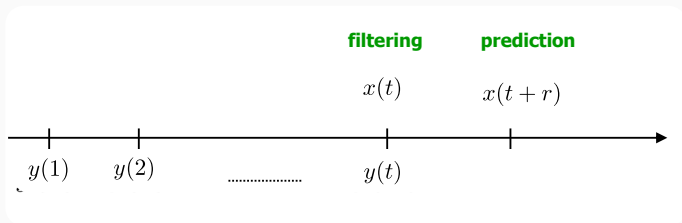
Kalman estimation (cont.)

- We refer to system's descriptions through **state equations**:

$$x(t+1) = Fx(t) + v_1(t) \quad x, v_1 \in \mathbb{R}^n$$

$$y(t) = Hx(t) + v_2(t) \quad y, v_2 \in \mathbb{R}^p$$

- $v_1 \sim \text{WGN}(0, V_1)$, $v_2 \sim \text{WGN}(0, V_2)$
- $v_1(\cdot)$, $v_2(\cdot)$ independent, mutually and with $x(1)$
- F , H , V_1 , V_2 known



State estimation and Bayes estimation

- Since $v_1(t)$ and $v_2(t)$ are random variables, also $x(t)$ and $y(t)$ are r.v. \implies both the data $y(t), y(t-1), \dots$ and the unknown $x(t)$ are r.v. \implies it is natural to resort to the Bayes framework
- From the Gaussian assumption on the exogenous variables and the linearity of the dynamic system it follows that **the probability density functions of the state, the output and the state/output joint probability density functions are Gaussian as well.**

$$\hat{x}(t+r|t) = x(t+r)_m + \Lambda_{x(t+r)d} \Lambda_{dd}^{-1} (d - d_m)$$

where:

- $x(t+r)_m := \mathbb{E}[x(t+r)]$
- $d := y^t := \text{col}[y(t), y(t-1), \dots, y(1)]$
- $d_m := \mathbb{E}[d]$

State estimation and Bayes estimation (cont.)

- But:

$$E[v_1(t)] = 0, E[v_2(t)] = 0 \implies E[x(t)] = 0, E[y(t)] = 0$$



$$\hat{x}(t+r|t) = \Lambda_{x(t+r)d} \Lambda_{dd}^{-1} d \quad (\star)$$

Remark: formula (\star) makes sense also if the Gaussian assumptions do not hold. In such a case $\Lambda_{x(t+r)d} \Lambda_{dd}^{-1} d$ is the best linear estimator

- (\star) solves the problem but **it is NOT recursive**. Instead, we want to obtain a recursive estimator of the form:

$$\hat{x}(t+r|t) = f [\hat{x}(t+r-1|t-1)]$$

Recursive form of Bayes estimation

- For now, denote by ϑ the unknown to be estimated and by d the observed data.
- Suppose (just for simplicity and without loss of generality) that
 - ϑ scalar
 - $d(1), d(2)$ two scalar data
 - $E(\vartheta) = 0, E[d(1)] = 0, E[d(2)] = 0$
- Then

$$\begin{bmatrix} \vartheta \\ d(1) \\ d(2) \end{bmatrix} \sim G \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_{\vartheta\vartheta} & \lambda_{\vartheta 1} & \lambda_{\vartheta 2} \\ \lambda_{1\vartheta} & \lambda_{11} & \lambda_{12} \\ \lambda_{2\vartheta} & \lambda_{21} & \lambda_{22} \end{bmatrix} \right)$$

where $\lambda_{\vartheta\vartheta} = E(\vartheta^2), \lambda_{\vartheta 1} = E[\vartheta d(1)], \dots$

Recursive form of Bayes estimation (cont.)

- The estimate of ϑ based on the **single data point** $d(1)$ is given by

$$E[\vartheta | d(1)] = \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1)$$

- Instead, the estimate of ϑ based on **two data points** $d(1)$, $d(2)$ is

$$E[\vartheta | d(1), d(2)] = [\lambda_{\vartheta 1} \quad \lambda_{\vartheta 2}] \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}^{-1} \begin{bmatrix} d(1) \\ d(2) \end{bmatrix}$$

where $\lambda_{12} = \lambda_{21}$ But

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}^{-1} = \frac{1}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} \begin{bmatrix} \lambda_{22} & -\lambda_{12} \\ -\lambda_{12} & \lambda_{11} \end{bmatrix}$$

and hence

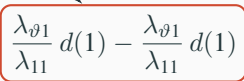
$$E[\vartheta | d(1), d(2)] = \frac{1}{\lambda_{11}\lambda_{22} - \lambda_{12}^2} [(\lambda_{\vartheta 1}\lambda_{22} - \lambda_{\vartheta 2}\lambda_{12}) d(1) + (-\lambda_{\vartheta 1}\lambda_{12} + \lambda_{\vartheta 2}\lambda_{11}) d(2)]$$

Recursive form of Bayes estimation (cont.)

- letting $\lambda^2 = \lambda_{22} - \frac{\lambda_{12}^2}{\lambda_{11}}$ we have

$$\begin{aligned} E[\vartheta \mid d(1), d(2)] &= \frac{1}{\lambda_{11}\lambda^2} (-\lambda_{\vartheta 1}\lambda_{12} + \lambda_{\vartheta 2}\lambda_{11}) d(2) \\ &\quad + \frac{1}{\lambda_{11}\lambda^2} (\lambda_{\vartheta 1}\lambda_{22} - \lambda_{\vartheta 2}\lambda_{12}) d(1) \end{aligned}$$

- Adding and subtracting the term $E[\vartheta \mid d(1)] = \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1)$

$$\begin{aligned} E[\vartheta \mid d(1), d(2)] &= \frac{1}{\lambda_{11}\lambda^2} (-\lambda_{\vartheta 1}\lambda_{12} + \lambda_{\vartheta 2}\lambda_{11}) d(2) \\ &\quad + \frac{1}{\lambda_{11}\lambda^2} (\lambda_{\vartheta 1}\lambda_{22} - \lambda_{\vartheta 2}\lambda_{12}) d(1) + \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1) - \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1) \end{aligned}$$


Recursive form of Bayes estimation (cont.)

recursion

$$\begin{aligned} E[\vartheta | d(1), d(2)] &= \frac{1}{\lambda^2} \left(\lambda_{\vartheta 2} - \lambda_{\vartheta 1} \frac{\lambda_{12}}{\lambda_{11}} \right) d(2) \\ &+ \frac{1}{\lambda^2} \left(\lambda_{\vartheta 1} \frac{\lambda_{22}}{\lambda_{11}} - \lambda_{\vartheta 2} \frac{\lambda_{12}}{\lambda_{11}} - \lambda_{\vartheta 1} \frac{\lambda^2}{\lambda_{11}} \right) d(1) + \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1) \end{aligned}$$

- substituting $\lambda^2 = \lambda_{22} - \frac{\lambda_{12}^2}{\lambda_{11}}$ we have

$$\begin{aligned} E[\vartheta | d(1), d(2)] &= \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1) \\ &+ \frac{1}{\lambda^2} \left(\lambda_{\vartheta 2} - \lambda_{\vartheta 1} \frac{\lambda_{12}}{\lambda_{11}} \right) \left[d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1) \right] \end{aligned}$$

- **Definition.** Given two random variables $d(1)$ and $d(2)$ we call **innovation** of $d(2)$ with respect to $d(1)$ the quantity:

$$e = d(2) - E[d(2) | d(1)] = d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1)$$

Recursive form of Bayes estimation (cont.)

Let us analyze the random variable e :

- e is a linear combination of $d(1)$ and of $d(2)$ that are Gaussian $\implies e$ is Gaussian. Moreover ϑ , $d(1)$, e are jointly Gaussian
- $E(e) = 0$
- $\lambda_{ee} = E \left[\left(d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1) \right)^2 \right] = \lambda_{22} + \frac{\lambda_{12}^2}{\lambda_{11}^2} \lambda_{11} - 2 \frac{\lambda_{12}^2}{\lambda_{11}} = \lambda^2$
- $\lambda_{\vartheta e} = E \left[\vartheta \left(d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1) \right) \right] = \lambda_{\vartheta 2} - \lambda_{\vartheta 1} \frac{\lambda_{12}}{\lambda_{11}}$
- $\lambda_{1e} = E \left[d(1) \left(d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1) \right) \right] = \lambda_{12} - \lambda_{11} \frac{\lambda_{12}}{\lambda_{11}} = 0$

The innovation e is uncorrelated with $d(1)$

Recursive form of Bayes estimation (cont.)

- Hence

$$\begin{aligned} E[\vartheta | d(1), d(2)] &= \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1) \\ &\quad + \frac{1}{\lambda^2} \left(\lambda_{\vartheta 2} - \lambda_{\vartheta 1} \frac{\lambda_{12}}{\lambda_{11}} \right) \left[d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1) \right] \\ &= \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1) + \frac{\lambda_{\vartheta e}}{\lambda_{ee}} e \end{aligned}$$

and, since ϑ , $d(1)$, e are jointly Gaussian, we have

$$E[\vartheta | d(1), d(2)] = E[\vartheta | d(1)] + E[\vartheta | e]$$

Thus: the optimal estimate can be expressed also as a function of the innovation.

Recursive form of Bayes estimation (cont.)

- Observe that

$$\mathbb{E}[\vartheta | d(1), e] = \mathbb{E}[\vartheta | d(1)] + \mathbb{E}[\vartheta | e]$$

because e is uncorrelated with $d(1)$; thus, the optimal estimate given $d(1)$, $d(2)$ coincides with the optimal estimate given $d(1)$, e

$d(2)$ and e have the **same information content**

In particular:

$$e = d(2) - \mathbb{E}[d(2) | d(1)] \implies d(2) = \mathbb{E}[d(2) | d(1)] + e$$

and hence the innovation represents the “part” of $d(2)$ which is not predictable on the basis of $d(1)$.

The innovation represents the actual information content of $d(2)$ with respect to $d(1)$

Generalization to the vector case

- Now, if ϑ , $d(1)$, $d(2)$ are **zero-mean vectors** we have:

$$\begin{bmatrix} \vartheta \\ d(1) \\ d(2) \end{bmatrix} \sim G \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Lambda_{\vartheta\vartheta} & \Lambda_{\vartheta 1} & \Lambda_{\vartheta 2} \\ \Lambda_{1\vartheta} & \Lambda_{11} & \Lambda_{12} \\ \Lambda_{2\vartheta} & \Lambda_{21} & \Lambda_{22} \end{bmatrix} \right)$$

where $\Lambda_{\vartheta 1} = \Lambda_{1\vartheta}^\top$, $\Lambda_{\vartheta 2} = \Lambda_{2\vartheta}^\top$, $\Lambda_{21} = \Lambda_{12}^\top$

- We obtain:

$$e = d(2) - \mathbb{E}[d(2) | d(1)] = d(2) - \Lambda_{21}\Lambda_{11}^{-1} d(1)$$

and hence:

$$\begin{aligned} \mathbb{E}[\vartheta | d(1), d(2)] &= \mathbb{E}[\vartheta | d(1)] + \mathbb{E}[\vartheta | e] \\ &= \Lambda_{\vartheta 1}\Lambda_{11}^{-1} d(1) + \Lambda_{\vartheta e}\Lambda_{ee}^{-1} e \end{aligned}$$

Generalization to the non-zero mean case

- Now, if ϑ , $d(1)$, $d(2)$ are **non-zero mean vectors** we have:

$$\begin{bmatrix} \vartheta \\ d(1) \\ d(2) \end{bmatrix} \sim G \left(\begin{bmatrix} \vartheta_m \\ d(1)_m \\ d(2)_m \end{bmatrix}, \begin{bmatrix} \Lambda_{\vartheta\vartheta} & \Lambda_{\vartheta 1} & \Lambda_{\vartheta 2} \\ \Lambda_{1\vartheta} & \Lambda_{11} & \Lambda_{12} \\ \Lambda_{2\vartheta} & \Lambda_{21} & \Lambda_{22} \end{bmatrix} \right)$$

- We obtain:

$$\begin{aligned} E[\vartheta \mid d(1), d(2)] &= E[\vartheta \mid d(1)] + E[\vartheta \mid e] - \vartheta_m \\ &= \vartheta_m + \Lambda_{\vartheta 1} \Lambda_{11}^{-1} [d(1) - d(1)_m] + \Lambda_{\vartheta e} \Lambda e e^{-1} e \end{aligned}$$

where, in analogy with the zero-mean scalar case we have:

- $E(e) = 0$
- $\Lambda_{1e} = E \left\{ [d(1) - d(1)_m]^\top e \right\} = 0$
- $\Lambda_{\vartheta e} = \Lambda_{\vartheta 2} - \Lambda_{\vartheta 1} \Lambda_{11}^{-1} \Lambda_{12}$

Geometric interpretation of Bayes recursive estimation

Recall (Bayes estimation):

- Suppose that d and ϑ are marginally and jointly Gaussian random variables:

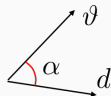
$$\begin{bmatrix} d \\ \vartheta \end{bmatrix} \sim G \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_{dd} & \lambda_{d\vartheta} \\ \lambda_{\vartheta d} & \lambda_{\vartheta\vartheta} \end{bmatrix} \right)$$

Hence d and ϑ can be interpreted **geometric vectors**

- Define the scalar product $(\vartheta, d) = E(\vartheta \cdot d)$
- The usual properties of vector spaces equipped with scalar product hold true. In particular:

$$\|\vartheta\| = \sqrt{(\vartheta, \vartheta)}$$

$$\|d\| = \sqrt{(d, d)}$$



$$(\vartheta, d) = \|\vartheta\| \|d\| \cos \alpha$$

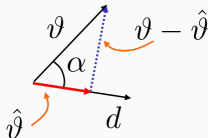
- **Uncorrelated** random variables correspond to orthogonal vectors

Geometric interpretation of Bayes recursive estimation (cont.)

- Now:

$$\begin{aligned}\hat{\vartheta} &= \frac{\lambda_{\vartheta d}}{\lambda_{dd}} d = \frac{E(\vartheta \cdot d)}{E(d \cdot d)} d = \frac{(\vartheta, d)}{\|d\|^2} d = \frac{(\vartheta, d)}{\|d\|^2} \frac{\|\vartheta\|}{\|\vartheta\|} d \\ &= \frac{(\vartheta, d)}{\|\vartheta\| \|d\|} \|\vartheta\| \frac{d}{\|d\|} = \|\vartheta\| \cos \alpha \frac{d}{\|d\|}\end{aligned}$$

The optimal estimate $\hat{\vartheta}$ is the projection of ϑ on the data vector d



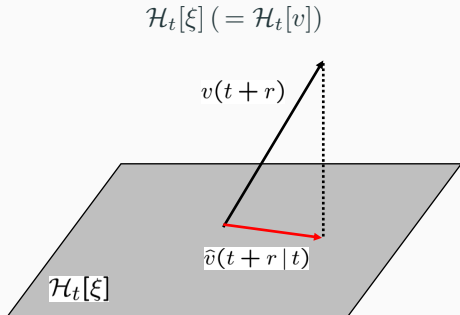
- Then consider the vector $\vartheta - \hat{\vartheta}$. It follows that:

$$\begin{aligned}\|\vartheta - \hat{\vartheta}\|^2 &= \|\vartheta\|^2 - \|\hat{\vartheta}\|^2 = \|\vartheta\|^2 - \|\vartheta\|^2 (\cos \alpha)^2 \\ &= \lambda_{\vartheta\vartheta} - \lambda_{\vartheta\vartheta} \frac{\lambda_{\vartheta d}^2}{\lambda_{dd} \lambda_{\vartheta\vartheta}} = \lambda_{\vartheta\vartheta} - \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}}\end{aligned}$$

The error variance is the square of the length of vector $\vartheta - \hat{\vartheta}$.

Geometric interpretation of Bayes recursive estimation (cont.)

- In the prediction problem, $\hat{v}(t+r|t)$ is the projection of $v(t+r)$ (interpreted as a geometric vector) on the subspace (hyperplane)

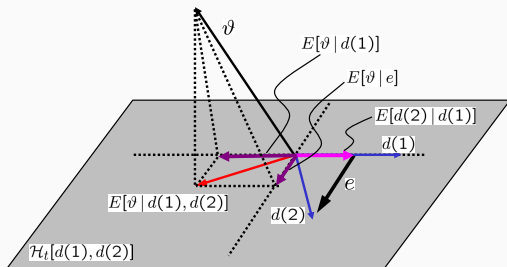


Geometric interpretation of Bayes recursive estimation (cont.)

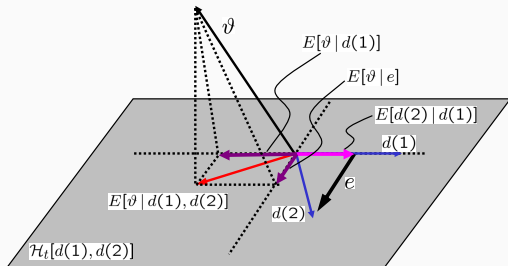
- If

$$\begin{bmatrix} \vartheta \\ d(1) \\ d(2) \end{bmatrix} \sim G \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Lambda_{\vartheta\vartheta} & \Lambda_{\vartheta 1} & \Lambda_{\vartheta 2} \\ \Lambda_{1\vartheta} & \Lambda_{11} & \Lambda_{12} \\ \Lambda_{2\vartheta} & \Lambda_{21} & \Lambda_{22} \end{bmatrix} \right)$$

we are able to consider ϑ , $d(1)$, $d(2)$ as geometric vectors, and hence



Geometric interpretation of Bayes recursive estimation (cont.)



- Note that:

- e lies on the plane $\mathcal{H}_t[d(1), d(2)]$ and is orthogonal to $d(1)$
- $E[\vartheta | d(1)]$ is orthogonal to $E[\vartheta | e]$
- $E[\vartheta | d(1), d(2)] = E[\vartheta | d(1)] + E[\vartheta | e]$
- $E[\vartheta | d(1), d(2)] \neq E[\vartheta | d(1)] + E[\vartheta | d(2)]$

not true in general

Kalman Estimation

Kalman Predictor

One-step ahead Kalman predictor

- Consider the dynamic system

$$\begin{cases} x(t+1) = Fx(t) + v_1(t) \\ y(t) = Hx(t) + v_2(t) \end{cases} \quad x, v_1 \in \mathbb{R}^n, y, v_2 \in \mathbb{R}^p$$

- $v_1 \sim WGN(0, V_1)$, $v_2 \sim WGN(0, V_2)$
- $v_1(\cdot)$, $v_2(\cdot)$ independent, mutually and with $x(1)$
- F, H, V_1, V_2 known, $V_2 > 0$
- We want to design a one step ahead state predictor in recursive form:

$$\hat{x}(t+1|t) \text{ function of } \hat{x}(t|t-1)$$

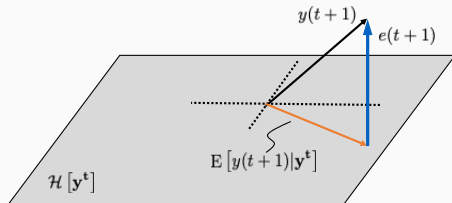
One-step ahead Kalman predictor (cont.)

Let us enhance the role played by the **innovation**:

- the prediction of $x(t+1)$ is based on the data $y(t), y(t-1), \dots, y(1)$
- $\mathbf{y}^t = \text{col}[y(t), y(t-1), \dots, y(1)]$ generates the subspace of the past $\mathcal{H}[\mathbf{y}^t]$
- The innovation provided by the $(t+1)$ -th data-point with respect to \mathbf{y}^t is given by

$$e(t+1) = y(t+1) - \mathbb{E}[y(t+1) | \mathbf{y}^t]$$

and hence the situation is:

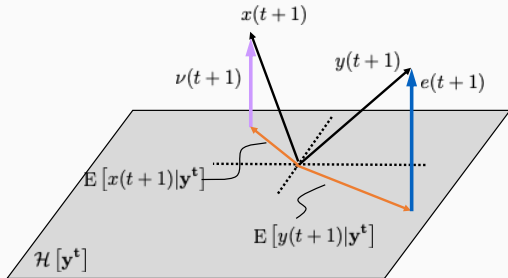


One-step ahead Kalman predictor (cont.)

- The **state prediction error** is:

$$\nu(t+1) = x(t+1) - \hat{x}(t+1|t) = x(t+1) - \mathbb{E}[x(t+1)|\mathbf{y}^t]$$

and thus the situation now is:



The state prediction error $\nu(t+1)$ is orthogonal to the past $\mathcal{H}[\mathbf{y}^t]$

Optimal one-step ahead output prediction

- We have:

$$\begin{aligned}\hat{y}(t+1|t) &= \mathbb{E}[y(t+1) | \mathbf{y}^t] \\ &= \mathbb{E}[Hx(t+1) + v_2(t+1) | \mathbf{y}^t] \\ &= H \mathbb{E}[x(t+1) | \mathbf{y}^t] + \mathbb{E}[v_2(t+1) | \mathbf{y}^t] \\ &= H \hat{x}(t+1|t) + \mathbb{E}[v_2(t+1) | \mathbf{y}^t]\end{aligned}$$

- Let us analyze the term $\mathbb{E}[v_2(t+1) | \mathbf{y}^t]$:

$$\begin{aligned}x(t) &= f[\mathbf{v}_1^{t-1}, x(1)] = f[v_1(t-1), v_1(t-2), \dots, v_1(1), x(1)] \\ y(t) &= \bar{f}[\mathbf{v}_1^{t-1}, x(1), v_2(t)] \implies \mathbf{y}^t = \bar{f}[\mathbf{v}_1^{t-1}, x(1), \mathbf{v}_2^t]\end{aligned}$$

- $v_2(\cdot)$ white $\implies v_2(t+1)$ independent from \mathbf{v}_2^t
- $v_1(\cdot), v_2(\cdot)$ independent, mutually and with $x(1)$ [Hp.]
- $v_2(t+1)$ independent with \mathbf{y}^t

$$\mathbb{E}[v_2(t+1) | \mathbf{y}^t] = \mathbb{E}[v_2(t+1)] = 0$$

$$\hat{y}(t+1|t) = H \hat{x}(t+1|t)$$

Recursive one-step ahead prediction

- We have

$$\begin{aligned}\hat{x}(t+1|t) &= \mathbb{E}[x(t+1) | \mathbf{y}^t] \\ &= \mathbb{E}[x(t+1) | \mathbf{y}^{t-1}, y(t)]\end{aligned}$$

- From the recursive Bayes formula:

$$\hat{x}(t+1|t) = \mathbb{E}[x(t+1) | \mathbf{y}^{t-1}] + \mathbb{E}[x(t+1) | e(t)]$$

- Let us first compute the term $\mathbb{E}[x(t+1) | \mathbf{y}^{t-1}]$:

$$\begin{aligned}\mathbb{E}[x(t+1) | \mathbf{y}^{t-1}] &= \mathbb{E}[Fx(t) + v_1(t) | \mathbf{y}^{t-1}] \\ &= F \mathbb{E}[x(t) | \mathbf{y}^{t-1}] + \mathbb{E}[v_1(t) | \mathbf{y}^{t-1}]\end{aligned}$$

But $v_1(t)$ independent with \mathbf{y}^{t-1}

$$\mathbb{E}[v_1(t) | \mathbf{y}^{t-1}] = \mathbb{E}[v_1(t)] = 0$$

$$\mathbb{E}[x(t+1) | \mathbf{y}^{t-1}] = F \hat{x}(t|t-1)$$

Recursive one-step ahead prediction (cont.)

- Now compute the term $E [x(t + 1) | e(t)]$. From Bayes formula

$$E [x(t + 1) | e(t)] = \Lambda_{x(t+1)e(t)} \Lambda_{e(t)e(t)}^{-1} e(t)$$

And hence the problem has been reduced to the one of determining the matrices $\Lambda_{x(t+1)e(t)}$, $\Lambda_{e(t)e(t)}$

- Expression of $\Lambda_{x(t+1)e(t)} = E [x(t + 1) e(t)^T]$

$$\begin{aligned} e(t) &= y(t) - E [y(t) | \mathbf{y}^{t-1}] = y(t) - \hat{y}(t | t - 1) \\ &= H x(t) + v_2(t) - H \hat{x}(t | t - 1) \\ &= H [x(t) - \hat{x}(t | t - 1)] + v_2(t) \end{aligned}$$

Hence:

$$\begin{aligned} \Lambda_{x(t+1)e(t)} &= E \left\{ [F x(t) + v_1(t)] \cdot [H [x(t) - \hat{x}(t | t - 1)] + v_2(t)]^T \right\} \\ &= F E \left\{ x(t) [x(t) - \hat{x}(t | t - 1)]^T \right\} \cdot H^T \\ &\quad + F E [x(t)v_2(t)^T] \\ &\quad + E \left\{ v_1(t) [H (x(t) - \hat{x}(t | t - 1)) + v_2(t)]^T \right\} \end{aligned}$$

Recursive one-step ahead prediction (cont.)

- Now, let us analyze separately the terms $F \mathbb{E} \{x(t)v_2(t)^\top\}$ and $\mathbb{E} \left\{ v_1(t) [H(x(t) - \hat{x}(t|t-1)) + v_2(t)]^\top \right\}$
- (★) $F \mathbb{E} [x(t)v_2(t)^\top]$
 - $v_1(\cdot), v_2(\cdot)$ independent, mutually and with $x(1)$ [Hp.]
 - $v_2(t)$ independent with $x(t)$

$$\mathbb{E} [x(t)v_2(t)^\top] = \mathbb{E} [x(t)] \mathbb{E} [v_2(t)^\top] = 0$$

Recursive one-step ahead prediction (cont.)

- (★★)
$$\mathbb{E} \left\{ v_1(t) [H(x(t) - \hat{x}(t|t-1)) + v_2(t)]^\top \right\}$$

$$= \mathbb{E} [v_1(t)x(t)^\top] H^\top$$

$$- \mathbb{E} [v_1(t)\hat{x}(t|t-1)^\top] H^\top$$

$$+ \mathbb{E} [v_1(t)v_2(t)^\top]$$

- but $v_1(\cdot)$ white $\implies v_1(t)$ independent with \mathbf{v}_1^{t-1}
- $v_1(\cdot)$ independent with $x(1)$ [Hp.] $\implies v_1(t)$ independent with $x(t)$

$$\mathbb{E} [v_1(t)x(t)^\top] = \mathbb{E} [v_1(t)] \mathbb{E} [x(t)^\top] = \mathbf{0}$$

Recursive one-step ahead prediction (cont.)

- Moreover $\hat{x}(t|t-1)$ depends on \mathbf{y}^{t-1} which, in turn, depends on $v_1(t-2), v_1(t-3), \dots, x(1)$ and on $v_2(t-1)$ etc.

$$\mathbb{E} [v_1(t)\hat{x}(t|t-1)^\top] = 0$$

$$\Lambda_{x(t+1)e(t)} = F \cdot \mathbb{E} \left\{ x(t) [x(t) - \hat{x}(t|t-1)]^\top \right\} \cdot H^\top$$

- Now, introduce the term $\hat{x}(t|t-1)$ in order to make the state prediction error $\nu(t) = x(t) - \hat{x}(t|t-1)$ to show up in the overall formula:

$$\begin{aligned} \Lambda_{x(t+1)e(t)} &= F \cdot \mathbb{E} \left\{ [x(t) - \hat{x}(t|t-1)] [x(t) - \hat{x}(t|t-1)]^\top \right\} \cdot H^\top \\ &\quad + F \cdot \mathbb{E} \left\{ \hat{x}(t|t-1) [x(t) - \hat{x}(t|t-1)]^\top \right\} \cdot H^\top \end{aligned}$$

$$\Lambda_{x(t+1)e(t)} = F \cdot \mathbb{E} [\nu(t)\nu(t)^\top] \cdot H^\top + F \cdot \mathbb{E} [\hat{x}(t|t-1)\nu(t)^\top] \cdot H^\top$$

Recursive one-step ahead prediction (cont.)

- It is now worth introducing the **state prediction error covariance matrix**:

$$P(t) = \mathbb{E} [\nu(t)\nu(t)^\top]$$

- Finally, notice that $\nu(t)$ is orthogonal to $\mathcal{H}[\mathbf{y}^t]$, whereas

$$\hat{x}(t | t-1) \in \mathcal{H}[\mathbf{y}^t]$$

$$\mathbb{E} [\hat{x}(t | t-1)\nu(t)^\top] = \mathbb{E} [\hat{x}(t | t-1)] \mathbb{E} [\nu(t)^\top] = \mathbf{0}$$

$$\Lambda_{x(t+1)e(t)} = F \cdot P(t) \cdot H^\top$$

Recursive one-step ahead prediction (cont.)

- Expression of $\Lambda_{e(t)e(t)}$

Recall that

$$\begin{aligned}e(t) &= H [x(t) - \hat{x}(t | t - 1)] + v_2(t) \\ &= H \nu(t) + v_2(t)\end{aligned}$$

Hence

$$\begin{aligned}\Lambda_{e(t)e(t)} &= E [e(t)e(t)^\top] \\ &= H \cdot E [\nu(t)\nu(t)^\top] \cdot H^\top + E [v_2(t)v_2(t)^\top] \\ &\quad + H \cdot E [\nu(t)v_2(t)^\top] + E [v_2(t)\nu(t)^\top] \cdot H^\top\end{aligned}$$

and

$$\nu(t) = \check{f} [y^{t-1}, v_2(t)] \implies H \cdot E [\nu(t)v_2(t)^\top] = 0$$

$$\Lambda_{e(t)e(t)} = H \cdot P(t) \cdot H^\top + V_2$$

Recursive one-step ahead prediction (cont.)

- Summing up

$$\hat{x}(t+1|t) = \mathbb{E}[x(t+1)|\mathbf{y}^{t-1}] + \mathbb{E}[x(t+1)|e(t)]$$

where

$$\mathbb{E}[x(t+1)|\mathbf{y}^{t-1}] = F \hat{x}(t|t-1)$$

$$\begin{aligned}\mathbb{E}[x(t+1)|e(t)] &= \Lambda_{x(t+1)e(t)} \Lambda_{e(t)e(t)}^{-1} e(t) \\ &= F \cdot P(t) \cdot H^T [H \cdot P(t) \cdot H^T + V_2]^{-1} e(t)\end{aligned}$$

and hence

$$\hat{x}(t+1|t) = F \hat{x}(t|t-1) + K(t) \cdot e(t)$$

where the **gain matrix** “weighting” the innovation is

$$K(t) = F \cdot P(t) \cdot H^T [H \cdot P(t) \cdot H^T + V_2]^{-1}$$

We want to determine a **recursive formula** also for state prediction error covariance matrix $P(t) = E [\nu(t)\nu(t)^\top]$

- Then, we need to express in recursive way

$$\nu(t+1) = x(t+1) - \hat{x}(t+1|t)$$

But:

$$x(t+1) = Fx(t) + v_1(t)$$

$$\hat{x}(t+1|t) = F\hat{x}(t|t-1) + K(t) \cdot e(t)$$

$$\begin{aligned}\nu(t+1) &= F[x(t) - \hat{x}(t|t-1)] + v_1(t) - K(t)e(t) \\ &= F\nu(t) + v_1(t) - K(t)e(t)\end{aligned}$$

- On the other hand

$$\begin{aligned}e(t) &= y(t) - \hat{y}(t | t - 1) \\ &= Hx(t) + v_2(t) - H\hat{x}(t | t - 1) \\ &= H\nu(t) + v_2(t)\end{aligned}$$

$$\nu(t + 1) = [F - K(t)H] \nu(t) + v_1(t) - K(t)v_2(t)$$

Hence

$$\begin{aligned}P(t+1) &= E [\nu(t+1)\nu(t+1)^\top] \\&= E \left\{ [F - K(t)H] \nu(t)\nu(t)^\top [F - K(t)H]^\top \right\} + E [v_1(t)v_1(t)^\top] \\&+ E [K(t)v_2(t)v_2(t)^\top K(t)^\top] + E \left\{ [F - K(t)H] \nu(t)v_1(t)^\top \right\} \\&- E \left\{ [F - K(t)H] \nu(t)v_2(t)^\top K(t)^\top \right\} \\&+ E \left\{ v_1(t)\nu(t)^\top [F - K(t)H]^\top \right\} - E [v_1(t)v_2(t)^\top K(t)^\top] \\&- E \left\{ K(t)v_2(t)\nu(N)^\top [F - K(t)H]^\top \right\}\end{aligned}$$

However $\nu(t)$ is independent with $v_1(t)$ and with $v_2(t)$:

$$E [v_1(t)\nu(t)^\top] = E [\nu(t)v_1(t)^\top] = 0$$

$$E [v_2(t)\nu(t)^\top] = E [\nu(t)v_2(t)^\top] = 0$$

$$E [v_1(t)v_2(t)^\top] = E [v_2(t)v_1(t)^\top] = 0$$

Difference Riccati Equation (DRE)

$$P(t+1) = [F - K(t)H] P(t) [F - K(t)H]^\top + V_1 + K(t)V_2K(t)^\top$$

Riccati equation (cont.)

- Therefore, the Riccati equation is a recursive matrix equation which, once initialized, allows to compute the matrix $P(t)$
- There are several equivalent forms of Riccati equation. The following one is very useful because it does not explicitly involve the gain matrix $K(t)$ (this form can be derived by very simple algebraic manipulations)

$$P(t+1) = F \left\{ P(t) - P(t)H^T [V_2 + HP(t)H^T]^{-1} HP(t) \right\} F^T + V_1$$

Initialization of the Riccati recursive equation

- Notice that $\nu(1) = x(1) - \hat{x}(1|0)$ but $y(0)$ is not available and thus we are not able to compute $\nu(1)$ and hence $P(1)$
- Then, let us “formally” start the recursion from $P(2)$:

$$P(2) = E \left\{ [x(2) - \hat{x}(2|1)] [x(2) - \hat{x}(2|1)]^\top \right\}$$

and since $\hat{x}(2|1)$ is the Bayes estimate of $x(2)$ we can write:

$$P(2) = \Lambda_{x(2)x(2)} - \Lambda_{x(2)y(1)} \Lambda_{y(1)y(1)}^{-1} \Lambda_{y(1)x(2)}$$

but

$$\Lambda_{x(2)x(2)} = E \left\{ [Fx(1) + v_1(1)] [Fx(1) + v_1(1)]^\top \right\} = FP_1F^\top + V_1$$

where we set $P_1 = \text{var} [x(1)]$. Moreover:

$$\Lambda_{x(2)y(1)} = E \left\{ [Fx(1) + v_1(1)] [Hx(1) + v_2(1)]^\top \right\} = FP_1H^\top$$

$$\Lambda_{y(1)x(2)} = \Lambda_{x(2)y(1)}^\top$$

$$\Lambda_{y(1)y(1)} = E \left\{ [Hx(1) + v_2(1)] [Hx(1) + v_2(1)]^\top \right\} = HP_1H^\top + V_2$$

Initialization of the Riccati recursive equation (cont.)

- Then:

$$P(2) = FP_1F^\top + V_1 - FP_1H^\top (HP_1H^\top + V_2)^{-1} HP_1F^\top \quad (\star)$$



(\star) formally coincides with the Riccati equation with the position $P_1 = P(1)$

Interpretation

At instant 1, in which no past observed data are available, we assume that $\hat{x}(1|0) = E[x(1)] = 0$. Thus

$$P(1) = E \left\{ [x(1) - \hat{x}(1|0)] [x(1) - \hat{x}(1|0)]^\top \right\} = P_1$$

The Riccati is initialized with $P_1 = P(1) = \text{var} [x(1)]$ at instant 1 and not at instant 2.

- Let us address the initialization of

$$\hat{x}(t+1|t) = F \hat{x}(t|t-1) + K(t) \cdot e(t)$$

We have:

$$\begin{aligned}\hat{x}(2|1) &= \mathbb{E}[x(2) | x(1)] = \Lambda_{x(2)y(1)} \Lambda_{y(1)y(1)}^{-1} y(1) \\ &= \mathbb{E} \left\{ [Fx(1) + v_1(1)] [Hx(1) + v_2(1)]^\top \right\} \\ &\quad \times \left(\mathbb{E} \left\{ [Hx(1) + v_2(1)] [Hx(1) + v_2(1)]^\top \right\} \right)^{-1} y(1) \\ &= FP_1H^\top (HP_1H^\top + V_2)^{-1} y(1) \quad (\star)\end{aligned}$$

- We have

$$\hat{x}(2|1) = FP_1H^\top (HP_1H^\top + V_2)^{-1} y(1) \quad (\star)$$

Interpretation

Letting $\hat{x}(1|0) = 0 \implies e(1) = y(1) - H\hat{x}(1|0) = y(1)$

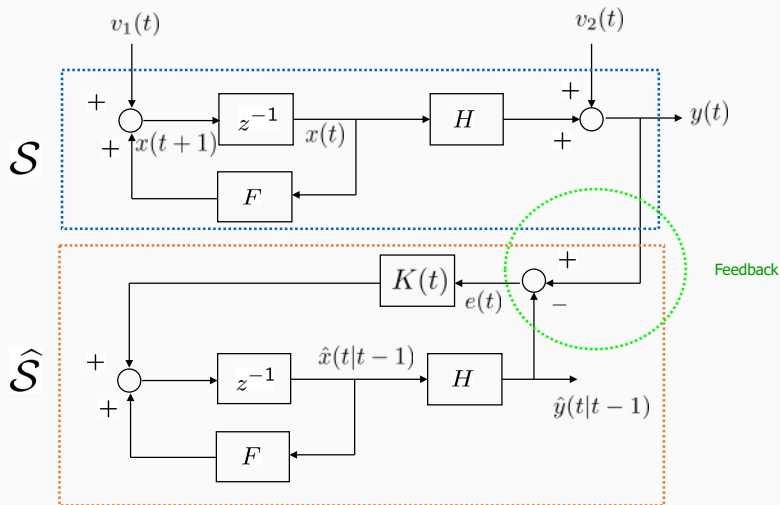
then relation (\star) is “compatible” with the recursive one and the interpretation is obvious: *a priori*, without available data, the more reasonable estimate is the *a priori* expected value.

Remark

If $E[x(1)] = \bar{x}_1 \neq 0$ we just initialize by $\hat{x}(1|0) = \bar{x}_1$.

Kalman predictor

The Kalman predictor architecture can be drawn as follows:



- The gain matrix $K(t)$ plays a fundamental role: the term $K(t)e(t)$ corrects the prediction based on a known state-space model of the system through the observed data collected on line.
- The Riccati equation can be solved off line, that is, the matrices $P(t)$ can be determined *a priori* and hence also the gain matrix $K(t)$.
- $P(t) \geq 0, \forall t > 1$ if $P(1) = P_1 \geq 0$
- $(HP_1H^T + V_2) > 0$ as we assumed $V_2 > 0$.

Kalman Estimation: Generalizations and Steady-state Estimator

- In the previous section, a recursive optimal state one-step ahead predictor has been devised in the context of Kalman estimation
- In the present last part, the following generalizations will be addressed:
 - The **Optimal r -steps ahead predictor** to compute $\hat{x}(t+r|t)$
 - The **Optimal Filter** to compute $\hat{x}(t|t)$
 - The optimal predictor with **exogenous inputs**
- Finally, the **steady-state behaviour** of the optimal predictor will be analyzed, that is, the characteristics and properties of the estimator when $t \rightarrow \infty$.

Kalman Estimation: Generalizations and Steady-state Estimator

r-steps Ahead Kalman Prediction

Recall the structure of the one-step ahead Kalman predictor

- We assumed that a state-space model of the system is available:

$$\begin{cases} x(t+1) = Fx(t) + v_1(t) \\ y(t) = Hx(t) + v_2(t) \end{cases} \quad x, v_1 \in \mathbb{R}^n, y, v_2 \in \mathbb{R}^p$$

- The one-step ahead Kalman predictor equations take on the form (see previous section):

$$\hat{x}(t+1|t) = F \hat{x}(t|t-1) + K(t) \cdot e(t)$$

$$K(t) = F \cdot P(t) \cdot H^T [H \cdot P(t) \cdot H^T + V_2]^{-1}$$

$$P(t+1) = F \left\{ P(t) - P(t)H^T [V_2 + HP(t)H^T]^{-1} HP(t) \right\} F^T + V_1$$

r-steps Ahead Kalman Prediction

- We want to estimate the variable $x(t+r)$ by observed data collected till instant t , with $r > 1$
- The optimal predictor is given by

$$\hat{x}(t+r|t) = \mathbb{E}[x(t+r)|y^t]$$

- Since $x(t+r) = Fx(t+r-1) + v_1(t+r-1)$ and owing to the fact that, for $\forall r > 1$, the noise sample $v_1(t+r-1)$ is uncorrelated with the observed data till instant t

➔ $\hat{x}(t+r|t) = F\hat{x}(t+r-1|t)$

- By iterating, we obtain

$$\hat{x}(t+r|t) = F^{r-1}\hat{x}(t+1|t)$$

- Moreover, concerning the output, as $v_2(t+r)$ is uncorrelated with the observed data till instant t

➔ $\hat{y}(t+r|t) = H\hat{x}(t+r|t)$

Summing up:

- The r -steps ahead Kalman predictor can be easily obtained from the one-step ahead predictor derived in the previous section:

$$\hat{x}(t+1|t) = F \hat{x}(t|t-1) + K(t) \cdot e(t)$$

$$K(t) = F \cdot P(t) \cdot H^T [H \cdot P(t) \cdot H^T + V_2]^{-1}$$

$$P(t+1) = F \left\{ P(t) - P(t)H^T [V_2 + HP(t)H^T]^{-1} HP(t) \right\} F^T + V_1$$



$$\hat{x}(t+r|t) = F^{r-1} \hat{x}(t+1|t)$$

$$\hat{y}(t+r|t) = H \hat{x}(t+r|t)$$

Kalman Estimation: Generalizations and Steady-state Estimator

Kalman Filter

- The “0-steps ahead prediction” is called **filtering**:

$$\begin{aligned}\hat{x}(t|t) &= \mathbb{E}[x(t)|y^t] = \mathbb{E}[x(t)|y^{t-1}, y(t)] \\ &= \mathbb{E}[x(t)|y^{t-1}] + \mathbb{E}[x(t)|e(t)] \\ &= \hat{x}(t|t-1) + \Lambda_{x(t)e(t)} \cdot \Lambda_{e(t)e(t)}^{-1} \cdot e(t)\end{aligned}$$

Already known 

$$\begin{aligned}\Lambda_{x(t)e(t)} &= \mathbb{E}\left\{x(t) [H(x(t) - \hat{x}(t|t-1)) + v_2(t)]^T\right\} \\ &= \mathbb{E}\left\{[x(t) + \hat{x}(t|t-1) - \hat{x}(t|t-1)] \cdot\right. \\ &\quad \left.[H(x(t) - \hat{x}(t|t-1)) + v_2(t)]^T\right\} \\ &= \dots\end{aligned}$$


Kalman Filter (cont.)

$$\Lambda_{x(t) e(t)} = \mathbb{E} \left\{ [x(t) - \hat{x}(t|t-1)] [x(t) - \hat{x}(t|t-1)]^T \right\} H^T$$

$$+ \mathbb{E} \left\{ [x(t) - \hat{x}(t|t-1)] v_2(t)^T \right\} \rightarrow 0$$

$$+ \mathbb{E} \left\{ \hat{x}(t|t-1) [x(t) - \hat{x}(t|t-1)]^T \right\} \rightarrow 0$$

$$+ \mathbb{E} \left\{ \hat{x}(t|t-1) v_2(t)^T \right\} \rightarrow 0$$


$$\Lambda_{x(t) e(t)} = P(t) H^T$$

- Summing up, we have:

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K_0(t) \cdot e(t)$$

Filter gain matrix

$$K_0(t) = P(t) \cdot H^T [H \cdot P(t) \cdot H^T + V_2]^{-1}$$

- Notice that

$$K(t) = FK_0(t)$$

Kalman Estimation: Generalizations and Steady-state Estimator

**Kalman Predictor in the Presence of
Exogenous Inputs**

Kalman Predictor in the Presence of Exogenous Inputs

- In this case, the state-space model is written as

$$\begin{cases} x(t+1) = Fx(t) + Gu(t) + v_1(t) \\ y(t) = Hx(t) + v_2(t) \end{cases} \quad x, v_1 \in \mathbb{R}^n, y, v_2 \in \mathbb{R}^p, u \in \mathbb{R}^m$$

- The equations for the **one-step ahead Kalman predictor** become

$$\begin{aligned} \hat{x}(t+1|t) &= F \hat{x}(t|t-1) + Gu(t) + K(t) \cdot e(t) \\ K(t) &= F \cdot P(t) \cdot H^\top [H \cdot P(t) \cdot H^\top + V_2]^{-1} \\ P(t+1) &= F \left\{ P(t) - P(t)H^\top [V_2 + HP(t)H^\top]^{-1} HP(t) \right\} F^\top + V_1 \end{aligned}$$

Kalman Estimation: Generalizations and Steady-state Estimator

Steady-state Kalman Estimator

Steady-State Kalman Estimator

- The equations to update $K(t)$ and $P(t + 1)$, at each time-step, require **the inversion of a matrix**.
- The inversion of a matrix of dimension $n \times n$ requires $o(n^3)$ operations (in general), and hence the **computational complexity** of each iteration increases with the cubic power of the dimension of the state vector.
- It would practically very appealing to be able to replace the time-varying matrices $K(t)$ and $P(t + 1)$ with **constant** matrices \bar{K} and \bar{P} computed off-line before hand.

Steady-State Kalman Estimator (cont.)

- Clearly, this would give rise to a **sub-optimal predictor** but would allow to handle in practice high-dimensional problems.
- In practice, we **replace** $K(t)$ and $P(t+1)$ with their **steady-state values**:

$$\bar{K} := \lim_{t \rightarrow \infty} K(t) \quad \bar{P} := \lim_{t \rightarrow \infty} P(t)$$

- The corresponding predictor is called **steady-state Kalman predictor**



$$\hat{x}(t+1|t) = F\hat{x}(t|t-1) + \bar{K}e(t)$$

Computation of the Steady-State Gain Matrix

- The steady-state gain matrix is simply computed as

$$\bar{K} = F\bar{P}H^T (H\bar{P}H^T + V_2)^{-1}$$

- The matrix \bar{P} is a solution of the **Algebraic Riccati Equation (ARE)**:

$$P(t+1) = P(t) = \bar{P}$$

→
$$\bar{P} = F \left[\bar{P} - \bar{P}H^T (V_2 + H\bar{P}H^T)^{-1} H\bar{P} \right] F^T + V_1$$
 ARE

- In case of multiple solutions of the ARE, it is necessary to choose the **positive-semidefinite** one.

Stability of the Steady-State Kalman Predictor

- Let us recall the equations of the steady-state Kalman predictor:

$$\begin{cases} \hat{x}(t+1|t) = F\hat{x}(t|t-1) + \bar{K}e(t) \\ \hat{y}(t+1|t) = H\hat{x}(t+1|t) \\ e(t) = y(t) - \hat{y}(t|t-1) \end{cases}$$

$$\hat{x}(t+1|t) = [F - \bar{K}H] \hat{x}(t|t-1) + \bar{K}y(t)$$

The predictor's stability depends on the eigenvalues of this matrix.

- If \bar{K} stabilizes $F - \bar{K}H$ the solution of the ARE is **stabilizing**.

Stability of the Steady-State Kalman Predictor (cont.)

- However, this does not say anything about the existence of a positive (definite or semi-definite) solution of the ARE.
- Then, it is worth asking under what conditions the recursive Riccati equation converges, that is, the ARE has (at least) one positive semi-definite solution.
- Recall that \bar{P} is the state prediction error covariance matrix.

First DRE Convergence Theorem

Assume that the dynamic system generating the observed data is **asymptotically stable**, that is, all eigenvalues of matrix F are strictly inside the unit disc.



- For every $P_1 \geq 0$ the recursive Riccati equation converges to the same matrix $\bar{P} \geq 0$.
- The steady-state predictor is asymptotically stable

Second DRE Convergence Theorem

- Consider

$$\begin{cases} x(t+1) = Fx(t) + G_v \xi(t) \\ y(t) = Hx(t) + v_2(t) \end{cases} \quad \begin{cases} \xi(t) \sim \text{WGN}(0, 1) \\ G_v : G_v G_v^T = V_1 \end{cases}$$

This decomposition is not unique

- Assume that the pair (F, H) is **observable** and (F, G_v) is **reachable**
 - For every positive semi-definite initial condition, the recursive Riccati equation asymptotically **converges** to the same matrix \bar{P}
 - The limit matrix \bar{P} is **positive-definite**
 - The steady-state Kalman predictor is **asymptotically stable**

Kalman Estimation: Generalizations and Steady-state Estimator

Observability

- Given the linear time-invariant dynamic system

$$\begin{cases} x(t+1) = Fx(t) \\ y(t) = Hx(t) \end{cases}$$

- The pair (F, H) is **observable** if two different initial states yielding output modes of behaviour that coincide for all future time-instants do not exist.

Theorem: Observability Necessary and Sufficient Condition

- Define the **observability matrix** for a system with n state variables

$$\mathcal{O}_n \triangleq \begin{bmatrix} H^T & | & F^T H^T & | & F^{T^2} H^T & | & \dots & | & F^{T^{n-1}} H^T \end{bmatrix}$$

- The pair (F, H) is **observable if and only if** \mathcal{O}_n has **rank n** .

Kalman Estimation: Generalizations and Steady-state Estimator

Reachability

- Given the linear time-invariant dynamic system (zero initial state)

$$x(t+1) = Fx(t) + G_v \xi(t)$$

- The pair (F, G_v) is **reachable** if, for a given final state \bar{x} , there exists a input sequence $\xi(t)$ and a time-instant t_N such that $x(t_N) = \bar{x}$

Theorem: Reachability Necessary and Sufficient Condition

- Define the **reachability matrix** for a system with n state variables

$$\mathcal{R}_n \triangleq [G_v \mid FG_v \mid F^2G_v \mid \dots \mid F^{n-1}G_v]$$

- The pair (F, G_v) is **reachable if and only if** \mathcal{R}_n has **rank n** .

Kalman Estimation: Generalizations and Steady-state Estimator

**Example of Riccati Equation
Convergence Analysis**

Example of Riccati Equation Convergence Analysis

Consider the dynamic system with one state variable:

$$\begin{cases} x(t+1) = \alpha x(t) + v_1(t) & v_1(t) \sim \text{WGN}(0, \beta^2) \\ y(t) = \gamma x(t) + v_2(t) & v_2(t) \sim \text{WGN}(0, 1) \end{cases}$$

- α , β and γ are system's parameters; we show that depending on their specific values, the Riccati equation may converge or not
- In all cases, we consider anyway the system to be **unstable**, that is $|\alpha| > 1$
- Even in this case of unstable system generating the data, the steady-state Kalman predictor may be able to track the state with a bounded prediction error.

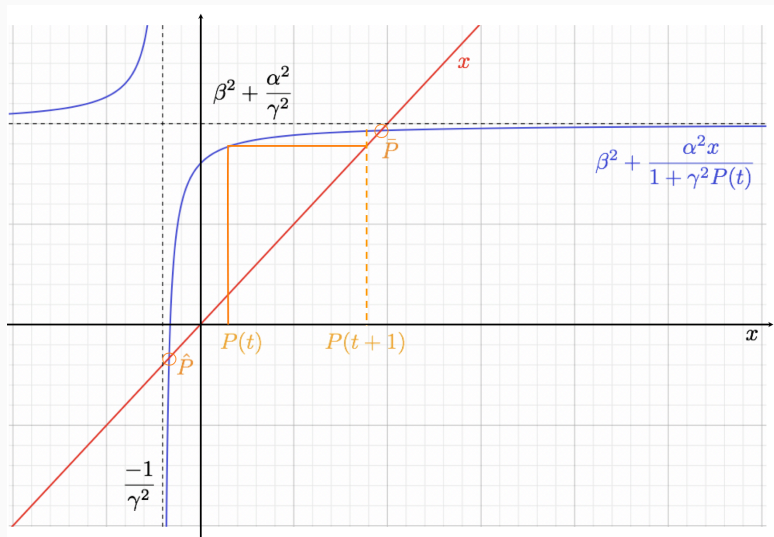
Example – Case 1

- Consider $|\alpha| > 1$, $\beta \neq 0$ and $\gamma \neq 0$
- The recursive Riccati equation becomes

$$P(t+1) = \beta^2 + \frac{\alpha^2 P(t)}{1 + \gamma^2 P(t)}$$

- It is easy to see that the corresponding ARE has two solutions, one positive and one negative
- Selecting the positive solution, we can see that $|\alpha - \bar{K}\gamma| < 1$, hence the solution is stabilizing

Example – Case 1 (cont.)



Example – Case 2

- Consider $|\alpha| > 1$, $\beta \neq 0$ and $\gamma = 0$
- The recursive Riccati equation becomes

$$P(t+1) = \beta^2 + \alpha^2 P(t)$$

- The solution always **diverges**
 - As $\gamma = 0$ the gain \bar{K} is zero (no use of the measurements), thus the predictor behaves in **open-loop**
 - The predictor tries to “mimic” the dynamic system (this is the best it can do) and actually the system is unstable, thus the divergence:

$$\hat{x}(t+1|t) = \alpha \hat{x}(t|t-1)$$

Example – Case 3

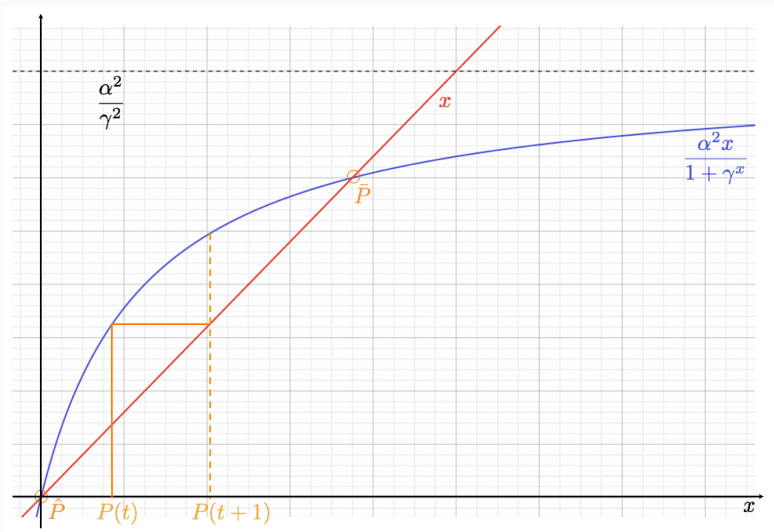
- Consider $|\alpha| > 1$, $\beta = 0$ and $\gamma \neq 0$
- The recursive Riccati equation becomes

$$P(t+1) = \frac{\alpha^2 P(t)}{1 + \gamma^2 P(t)}$$

- The ARE has two solutions, one zero and one positive
- In case of non-zero initial conditions, the Riccati equation converges to the positive solution
- If the initial condition P_1 is zero, the solution of the recursive Riccati equation is constantly equal to zero.
 - This is equivalent to state that there is no uncertainty on the state $x(0)$ at the initial time, nor at all subsequent time-instants
 - The predictor is unstable (as the dynamic system), operates in open-loop and yields an error-free prediction

$$\hat{x}(t+1|t) = \alpha \hat{x}(t|t-1)$$

Example – Case 3 (cont.)



267MI –Fall 2021

Lecture 13

**State Estimation from Observed
Data**

END