Systems Dynamics

Course ID: 267MI – Fall 2023

Thomas Parisini
Gianfranco Fenu

University of Trieste
Department of Engineering and Architecture

267MI –Fall 2023

Lecture 14
State Estimation from Observed Data

Lecture 14: Table of Contents

14. State Estimation from Observed Data

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

$$\hat{\vartheta} = \mathrm{E}\left(\vartheta\right) = \int \vartheta p\left(\vartheta\right) \, d\vartheta$$

  and the estimate uncertainty $\mathrm{var}(\hat{\vartheta})$ 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, $\mathrm{E}(\vartheta)$ and $\mathrm{var}(\vartheta)$ change as well. In particular, we expect $\mathrm{var}(\vartheta)$ 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

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

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

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

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

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

- We obtain:

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

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

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

$$\mathrm{var}\,(\vartheta - \hat{\vartheta}) = \mathrm{E}\left[(\vartheta - \hat{\vartheta})^2\right] = \mathrm{E}\left[\left(\vartheta - \frac{\lambda_{\vartheta d}}{\lambda_{dd}}\, d\right)^2\right]$$

$$= \mathrm{E}\left(\vartheta^2\right) - 2\,\frac{\lambda_{\vartheta d}}{\lambda_{dd}}\,\mathrm{E}(\vartheta d) + \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}^2}\,\mathrm{E}\left(d^2\right)$$

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

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 $\mathrm{E}(d) = 0$, $\mathrm{E}(\vartheta) = 0$
- As before, let us use the notations $\mathrm{E}(d^2) = \lambda_{dd}$, $\mathrm{E}(\vartheta^2) = \lambda_{\vartheta\vartheta}$, $\mathrm{E}(\vartheta d) = \lambda_{\vartheta d}$
- 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:

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

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

Optimal linear estimator

Optimal Linear Estimator (cont.)

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

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

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 $\mathrm{E}(d) = d_m\,,\ \mathrm{E}(\vartheta) = \vartheta_m$

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

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

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

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

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

$$\mathrm{var}\,(\vartheta - \hat{\vartheta}) = \Lambda_{\vartheta\vartheta} - \Lambda_{\vartheta d}\,\Lambda_{dd}^{-1}\Lambda_{d\vartheta}$$

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:

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

$$\hat{\vartheta} = \vartheta_m \qquad \mathrm{var}\,(\vartheta - \hat{\vartheta}) = \lambda_{\vartheta\vartheta} = \mathrm{var}\,(\vartheta)$$

- 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

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

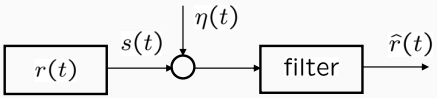
and this is correct because it expresses the fact that the estimate cannot but improve whenever the observed data convey useful information

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).



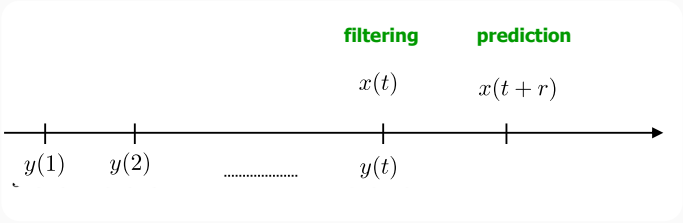
Example:
signal filtering

Kalman estimation (cont.)

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

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

- $v_1 \sim \mathsf{WGN}(0, V_1)$, $v_2 \sim \mathsf{WGN}(0, V_2)$
- $v_1(\cdot)$, $v_2(\cdot)$ 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. $\Longrightarrow$ both the data $y(t)$, $y(t-1)$, ... and the unknown $x(t)$ are r.v. $\Longrightarrow$ 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:

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

State estimation and Bayes estimation (cont.)

- But:

$$\mathrm{E}[v_1(t)] = 0,\ \mathrm{E}[v_2(t)] = 0 \quad \Longrightarrow \quad \mathrm{E}[x(t)] = 0,\ \mathrm{E}[y(t)] = 0$$

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

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
  - $\mathrm{E}(\vartheta) = 0$, $\mathrm{E}[d(1)] = 0$, $\mathrm{E}[d(2)] = 0$
- 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 $\lambda_{\vartheta\vartheta} = \mathrm{E}(\vartheta^2)$, $\lambda_{\vartheta 1} = \mathrm{E}[\vartheta d(1)]$, $\ldots$

Recursive form of Bayes estimation (cont.)

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

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

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

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

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

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

Recursive form of Bayes estimation (cont.)

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

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

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

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

Recursive form of Bayes estimation (cont.)

recursion

$$\mathrm{E}[\vartheta \mid d(1), d(2)] = \frac{1}{\lambda^2}\left(\lambda_{\vartheta 2} - \lambda_{\vartheta 1}\frac{\lambda_{12}}{\lambda_{11}}\right) d(2) + \frac{1}{\lambda^2}\left(\lambda_{\vartheta 1}\frac{\lambda_{22}}{\lambda_{11}} - \lambda_{\vartheta 2}\frac{\lambda_{12}}{\lambda_{11}} - \lambda_{\vartheta 1}\frac{\lambda^2}{\lambda_{11}}\right) d(1) + \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1)$$

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

$$\mathrm{E}[\vartheta \mid d(1), d(2)] = \frac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1) + \frac{1}{\lambda^2}\left(\lambda_{\vartheta 2} - \lambda_{\vartheta 1}\frac{\lambda_{12}}{\lambda_{11}}\right)\left[d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1)\right]$$

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

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

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 $\Longrightarrow$ e is Gaussian. Moreover ϑ, $d(1)$, e are jointly Gaussian
- $\mathrm{E}(e) = 0$
- $\lambda_{ee} = \mathrm{E}\left[\left(d(2) - \dfrac{\lambda_{12}}{\lambda_{11}} d(1)\right)^2\right] = \lambda_{22} + \dfrac{\lambda_{12}^2}{\lambda_{11}^2}\lambda_{11} - 2\dfrac{\lambda_{12}^2}{\lambda_{11}} = \lambda^2$
- $\lambda_{\vartheta e} = \mathrm{E}\left[\vartheta\left(d(2) - \dfrac{\lambda_{12}}{\lambda_{11}} d(1)\right)\right] = \lambda_{\vartheta 2} - \lambda_{\vartheta 1}\dfrac{\lambda_{12}}{\lambda_{11}}$
- $\lambda_{1e} = \mathrm{E}\left[d(1)\left(d(2) - \dfrac{\lambda_{12}}{\lambda_{11}} d(1)\right)\right] = \lambda_{12} - \lambda_{11}\dfrac{\lambda_{12}}{\lambda_{11}} = 0$

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

Recursive form of Bayes estimation (cont.)

- Hence

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

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

$$\boxed{\mathrm{E}[\vartheta \mid d(1), d(2)] = \mathrm{E}[\vartheta \mid d(1)] + \mathrm{E}[\vartheta \mid e]}$$

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

Recursive form of Bayes estimation (cont.)

- Observe that

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

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:

$$e = d(2) - \mathrm{E}[d(2) \mid d(1)] \quad \Longrightarrow \quad d(2) = \mathrm{E}[d(2) \mid d(1)] + e$$

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:

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

  and hence:

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

Generalization to the non-zero mean case

- Now, if $\vartheta,\ d(1),\ d(2)$ 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:

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

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

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

Geometric interpretation of Bayes recursive estimation

Recall (Bayes estimation):

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

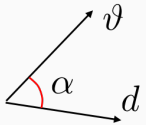
$$\begin{bmatrix} d \\ \vartheta \end{bmatrix} \sim G\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_{dd} & \lambda_{d\vartheta} \\ \lambda_{\vartheta d} & \lambda_{\vartheta\vartheta} \end{bmatrix} \right)$$

  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:

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

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

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



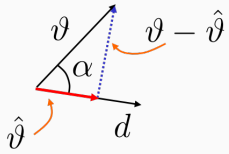
- Uncorrelated random variables correspond to orthogonal vectors

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:

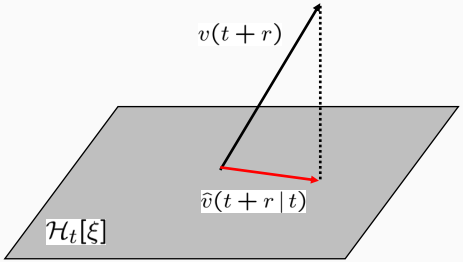
$$\begin{aligned}\|\vartheta - \hat{\vartheta}\|^2 &= \|\vartheta\|^2 - \|\hat{\vartheta}\|^2 = \|\vartheta\|^2 - \|\vartheta\|^2\, (\cos\alpha)^2 \\ &= \lambda_{\vartheta\vartheta} - \lambda_{\vartheta\vartheta}\, \frac{\lambda^2_{\vartheta d}}{\lambda_{dd}\lambda_{\vartheta\vartheta}} = \lambda_{\vartheta\vartheta}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

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)

$$\mathcal{H}_t[\xi]\ (=\mathcal{H}_t[v])$$

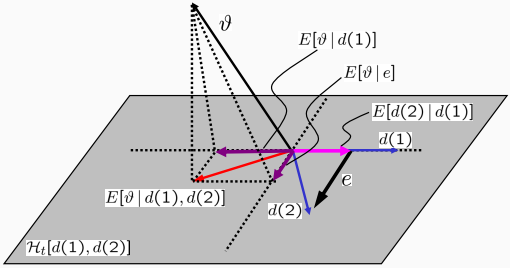


Geometric interpretation of Bayes recursive estimation (cont.)

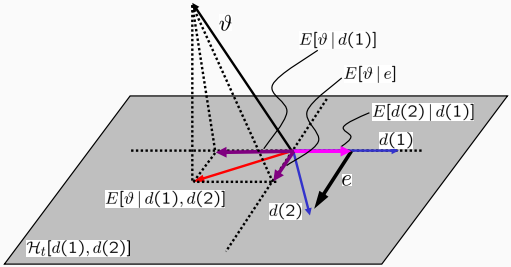
- If

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

we are able to consider $\vartheta,\ d(1),\ d(2)$ 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 $d(1)$
  - $\mathrm{E}[\vartheta \mid d(1)]$ is orthogonal to $\mathrm{E}[\vartheta|e]$
  - $\mathrm{E}[\vartheta \mid d(1), d(2)] = \mathrm{E}[\vartheta \mid d(1)] + \mathrm{E}[\vartheta \mid e]$ ← **not true in general**
  - $\mathrm{E}[\vartheta \mid d(1), d(2)] \neq \mathrm{E}[\vartheta \mid d(1)] + \mathrm{E}[\vartheta \mid d(2)]$

Kalman Estimation

Kalman Predictor

One-step ahead Kalman predictor

- Consider the dynamic system

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

  - $v_1 \sim WGN(0, V_1)$, $v_2 \sim WGN(0, V_2)$
  - $v_1(\cdot)$, $v_2(\cdot)$ 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:

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

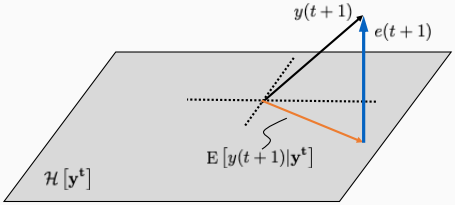
One-step ahead Kalman predictor (cont.)

Let us enhance the role played by the innovation:

- the prediction of $x(t+1)$ is based on the data $y(t),\ y(t-1),\ \ldots,\ y(1)$
- $\mathbf{y^t} = \text{col}[y(t),\ y(t-1),\ \ldots,\ y(1)]$ 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

$$e(t+1) = y(t+1) - \mathrm{E}\left[y\,(t+1)\,|\mathbf{y^t}\right]$$

and hence the situation is:

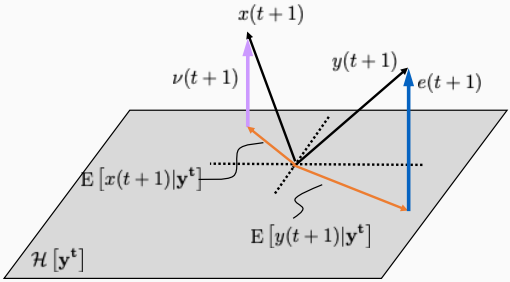


One-step ahead Kalman predictor (cont.)

- The state prediction error is:

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

and thus the situation now is:



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

Optimal one-step ahead output prediction

- We have:

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

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

$$x(t) = f\left[\mathbf{v_1^{t-1}}, x(1)\right] = f[v_1(t-1), v_1(t-2), \ldots, v_1(1), x(1)]$$
$$y(t) = \bar{f}\left[\mathbf{v_1^{t-1}}, x(1),\, v_2(t)\right] \quad \Longrightarrow \quad \mathbf{y^t} = \bar{f}\left[\mathbf{v_1^{t-1}}, x(1),\, \mathbf{v_2^t}\right]$$

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

$$\mathrm{E}\left[v_2(t+1)\,|\,\mathbf{y^t}\right] = \mathrm{E}\left[v_2(t+1)\right] = 0$$

  - $v_2(t+1)$ independent with $\mathbf{y^t}$

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

Recursive one-step ahead prediction

- We have

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

- From the recursive Bayes formula:

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

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

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

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

$$\mathrm{E}\left[v_1(t)\,|\,\mathbf{y}^{\mathbf{t-1}}\right] = \mathrm{E}\left[v_1(t)\right] = 0$$

$$\boxed{\mathrm{E}\left[x(t+1)\,|\,\mathbf{y}^{\mathbf{t-1}}\right] = F\,\hat{x}(t\,|\,t-1)}$$

Recursive one-step ahead prediction (cont.)

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

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

  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 $\Lambda_{x(t+1)e(t)} = \mathrm{E}\left[x(t+1)\,e(t)^\top\right]$

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

  Hence:

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

Recursive one-step ahead prediction (cont.)

- Now, let us analyze separately the terms $F\ \mathrm{E}\left\{x(t)v_2(t)^\top\right\}$ and $\mathrm{E}\left\{v_1(t)\left[H\left(x(t)-\hat{x}(t\,|\,t-1)\right)+v_2(t)\right]^\top\right\}$

- $(\star)\quad F\ \mathrm{E}\left[x(t)v_2(t)^\top\right]$
  - $v_1(\cdot)$, $v_2(\cdot)$ independent, mutually and with $x(1)$ [Hp.]
  - $v_2(t)$ independent with $x(t)$

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

Recursive one-step ahead prediction (cont.)

- $(\star\star)$ $\quad \mathrm{E}\left\{ v_1(t) \left[H \left(x(t) - \hat{x}(t \,|\, t-1) \right) + v_2(t) \right]^\top \right\}$

$$
\begin{aligned}
&= \mathrm{E}\left[v_1(t) x(t)^\top \right] H^\top \\
&- \mathrm{E}\left[v_1(t) \hat{x}(t \,|\, t-1)^\top \right] H^\top \\
&+ \mathrm{E}\left[v_1(t) v_2(t)^\top \right]
\end{aligned}
$$

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

$$
\mathrm{E}\left[v_1(t) x(t)^\top \right] = \mathrm{E}\left[v_1(t) \right] \mathrm{E}\left[x(t)^\top \right] = 0
$$

Recursive one-step ahead prediction (cont.)

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

$$\mathrm{E}\left[v_1(t)\hat{x}(t \mid t-1)^\top\right] = 0$$
$$\Lambda_{x(t+1)e(t)} = F \cdot \mathrm{E}\left\{x(t)\left[x(t) - \hat{x}(t \mid t-1)\right]^\top\right\} \cdot H^\top$$

- 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:

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

Recursive one-step ahead prediction (cont.)

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

$$P(t) = \mathrm{E}\left[\nu(t)\nu(t)^\top\right]$$

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

$$\hat{x}(t \mid t-1) \in \mathcal{H}[\mathbf{y^t}]$$
$$\mathrm{E}\left[\hat{x}(t \mid t-1)\nu(t)^\top\right] = \mathrm{E}\left[\hat{x}(t \mid t-1)\right] \mathrm{E}\left[\nu(t)^\top\right] = 0$$

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

Recursive one-step ahead prediction (cont.)

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

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

  Hence

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

  and

$$\nu(t) = \breve{f}\left[y^{t-1}, v_2(t)\right] \quad \Longrightarrow \quad H \cdot \mathrm{E}\left[\nu(t)v_2(t)^\top\right] = 0$$

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

Recursive one-step ahead prediction (cont.)

- Summing up

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

where

$$\begin{aligned}\mathrm{E}\left[x(t+1)\,|\,\mathbf{y^{t-1}}\right] &= F\,\hat{x}(t\,|\,t-1)\\ \mathrm{E}\left[x(t+1)\,|\,e(t)\right] &= \Lambda_{x(t+1)e(t)}\,\Lambda^{-1}_{e(t)e(t)}\,e(t)\\ &= F\cdot P(t)\cdot H^{\top}\left[H\cdot P(t)\cdot H^{\top}+V_2\right]^{-1}e(t)\end{aligned}$$

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

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

Riccati equation

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

- Then, we need to express in recursive way

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

But:

$$\begin{aligned} x(t+1) &= Fx(t) + v_1(t) \\ \hat{x}(t+1\,|\,t) &= F\,\hat{x}(t\,|\,t-1) + K(t)\cdot e(t) \end{aligned}$$

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

Riccati equation (cont.)

- On the other hand

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

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

Riccati equation (cont.)

Hence

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

Riccati equation (cont.)

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

$$
\begin{aligned}
E\left[v_1(t)\nu(t)^\top\right] &= E\left[\nu(t)v_1(t)^\top\right] = 0 \\
E\left[v_2(t)\nu(t)^\top\right] &= E\left[\nu(t)v_2(t)^\top\right] = 0 \\
E\left[v_1(t)v_2(t)^\top\right] &= E\left[v_2(t)v_1(t)^\top\right] = 0
\end{aligned}
$$

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 manupulations)

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

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 $P(2)$:

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

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

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

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

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

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

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

Initialization of the Riccati recursive equation (cont.)

- Then:

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

$(\star)$ 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

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

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

Initialization of the estimate

- Let us address the initialization of

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

We have:

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

Initialization of the estimate (cont.)

- We have

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

Interpretation

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

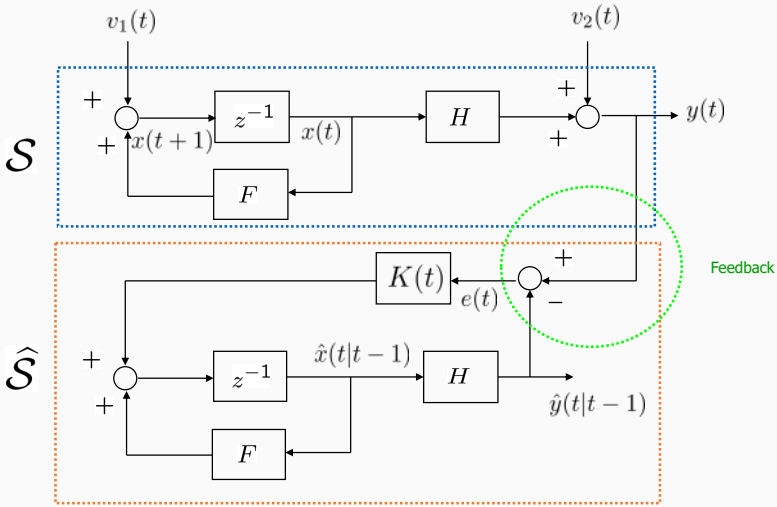
then relation $(\star)$ is "compatible" with the recursive one and the interpretation is obvious: *a priori*, without available data, the more reasonable estimate is the *a priori* expected value.

Remark

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

Kalman predictor

The Kalman predictor architecture can be drawn as follows:



Remarks

- 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 $P(1) = P_1 \geq 0$
- $\left(HP_1H^\top + V_2\right) > 0$ as we assumed $V_2 > 0$.

Kalman Estimation: Generalizations and Steady-state Estimator

Introduction

- 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 $t \to \infty$.

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:

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

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

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

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

$$\hat{x}\,(t+r|t) = \mathrm{E}\left[x\,(t+r)\,|y^t\right]$$

- 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

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

- 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

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

Summing up:

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

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

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

Kalman Estimation: Generalizations and Steady-state Estimator

Kalman Filter

Kalman Filter

- The "0-steps ahead prediction" is called filtering:

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

Already known

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

Kalman Filter (cont.)

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

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

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

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

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

Kalman Filter (cont.)

- Summing up, we have:

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

Filter gain matrix

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

- 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

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

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

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

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 x n requires $o\left(n^3\right)$ 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:

$$\bar{K} := \lim_{t\to\infty} K(t) \qquad \bar{P} := \lim_{t\to\infty} P(t)$$

- 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

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

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

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

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

- 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:

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

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

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

$$\begin{cases} x(t+1) = Fx(t) + G_v\xi(t) \\ y(t) = Hx(t) + v_2(t) \end{cases} \qquad \xi(t) \sim \mathsf{WGN}(0,1)$$

$$G_v: \; G_v\, G_v^T = V_1$$

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

Observability

- Given the linear time-invariant dynamic system

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

- 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

$$\mathcal{O}_n \triangleq \left[H^T \mid F^T H^T \mid {F^T}^2 H^T \mid \ldots \mid {F^T}^{n-1} H^T\right]$$

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

Kalman Estimation: Generalizations and Steady-state Estimator

Reachability

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

$$\mathcal{R}_n \stackrel{\triangle}{=} \left[G_v \mid FG_v \mid F^2G_v \mid \ \ldots \ \mid F^{n-1}G_v\right]$$

- 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:

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

- 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.

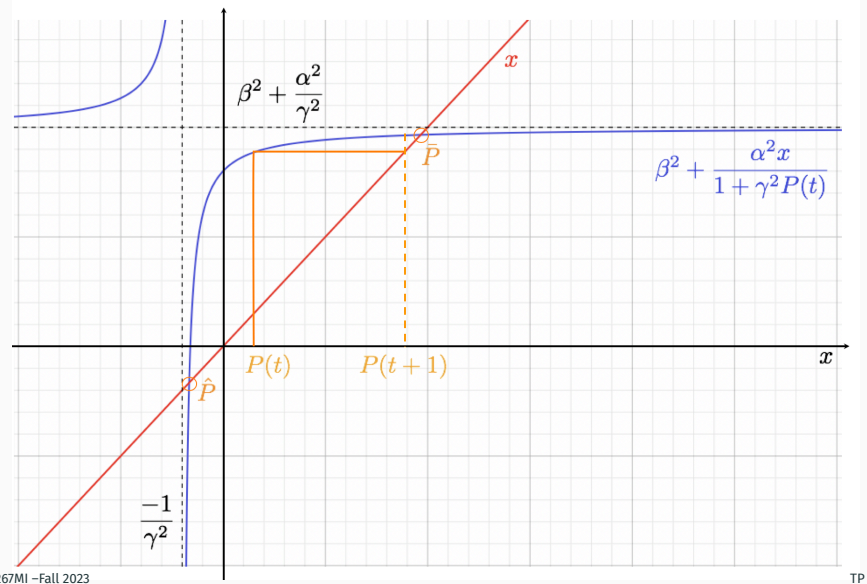
Example – Case 1

- 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

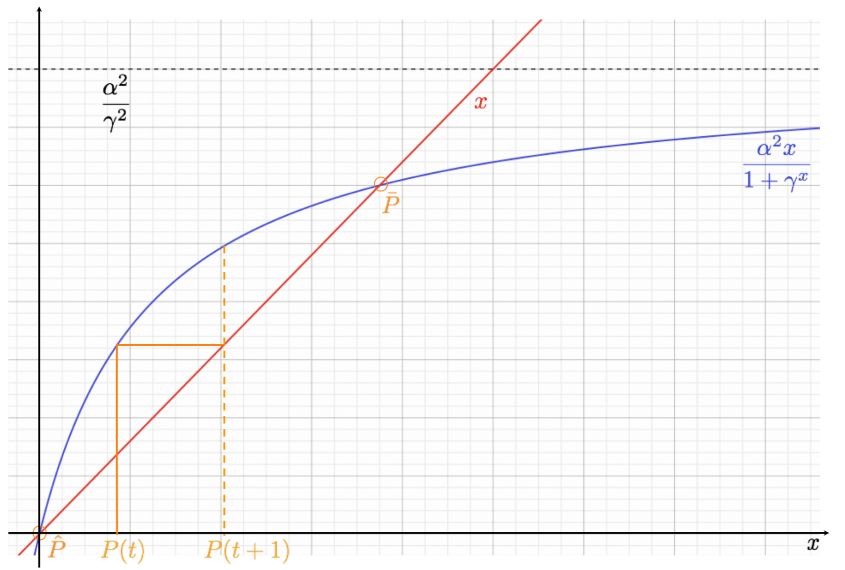
- Consider $|\alpha| > 1$, $\beta = 0$ and $\gamma \neq 0$
- The recursive Riccati equation becomes

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

- 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.)



Kalman Estimation: the Extended Kalman Filter

Kalman Estimation: the Extended Kalman Filter

The Estimation of State Variables of Nonlinear Systems

State Estimation for Nonlinear Systems

So far we have seen that the optimal linear filtering theorems and algorithms are clean and powerful.

What about nonlinear system state variables estimation?

Can we adapt/extend the Kalman filtering ideas to the nonlinear systems?

State Estimation for Nonlinear Systems (cont.)

We refer to system's descriptions through state equations:

$$\begin{cases} x(t+1) = f\big(x(t)\big) + v_1(t) & x,\, v_1 \in \mathbb{R}^n \\ y(t) = h\big(x(t)\big) + v_2(t) & y,\, v_2 \in \mathbb{R}^p \end{cases}$$

where

- $v_1 \sim \mathsf{WGN}(0, V_1)$, $v_2 \sim \mathsf{WGN}(0, V_2)$
- $v_1(\cdot)$, $v_2(\cdot)$ independent, mutually and with $x(1) \sim \mathsf{G}(\bar{x}, P_1)$
- $f(\cdot)$, $h(\cdot)$, V_1, V_2, P_1 known
- even if $v_1(\cdot)$, $v_2(\cdot)$ are Gaussian r.v., in general the state x and the measurements y are not Gaussian, due to the nonlinear transformations $f(\cdot)$ and $h(\cdot)$.

Still, we would solve the state prediction and filtering problems:

$$\hat{x}(t|t-1) = \mathrm{E}\left[x(t)|y(0), \ldots, y(t-1)\right] \quad \hat{x}(t|t) = \mathrm{E}\left[x(t)|y(0), \ldots, y(t)\right]$$

State Estimation for Nonlinear Systems: Linearisation

Given the nonlinear system

$$\begin{cases} x(t+1) = f\big(x(t)\big) + v_1(t) & x,\, v_1 \in \mathbb{R}^n \\ y(t) = h\big(x(t)\big) + v_2(t) & y,\, v_2 \in \mathbb{R}^p \end{cases}$$

Let us determine the **nominal state movement** as the solution of

$$\begin{aligned} \tilde{x}(t+1) &= f\big(\tilde{x}(t)\big) \\ \tilde{x}(1) &= \mathrm{E}\left[x(1)\right] \end{aligned}$$

and the corresponding **nominal output movement** $\tilde{y}(t) = h\big(\tilde{x}(t)\big)$.

Remark

The nominal state and output movements $\tilde{x}(t)$ and $\tilde{y}(t)$ can be computed *a priori*, by solving a deterministic difference equation, without any observed data.

State Estimation for Nonlinear Systems: Linearisation (cont.)

Linearisation via the Taylor Series Expansion

The nonlinear state and output movements, close to the nominal ones, may be approximated using the first order Taylor series expansion evaluated at the nominal movements $\tilde{x}(t)$ and $\tilde{y}(t)$.

Defining the variations

$$\Delta x(t) = x(t) - \tilde{x}(t) \qquad \Delta y(t) = y(t) - \tilde{y}(t)$$

the linearisation leads to

$$\begin{cases} \Delta x(t+1) = \tilde{F}(t)\,\Delta x(t) + v_1(t) \\ \Delta y(t) = \tilde{H}(t)\,\Delta x(t) + v_2(t) \end{cases}$$

where

$$\tilde{F}(t) = \frac{\partial\, f(x)}{\partial x}\Big|_{x=\tilde{x}(t)} \qquad \tilde{H}(t) = \frac{\partial\, h(x)}{\partial x}\Big|_{x=\tilde{x}(t)}$$

State Estimation for Nonlinear Systems: Linearisation (cont.)

Kalman theory can be applied to the linearised system, obtaining the predicted estimate for the state variation as follows

$$\widehat{\Delta x}\,(t+1|t) = \tilde{F}(t)\widehat{\Delta x}\,(t|t-1) + \tilde{K}(t)e(t)$$
$$e(t) = \Delta y(t) - \tilde{H}(t)\,\widehat{\Delta x}\,(t|t-1)$$

where

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

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

State Estimation for Nonlinear Systems: Linearisation (cont.)

Taking into account that the state estimate variation is

$$\widehat{\Delta x}\left(t|t-1\right) = \hat{x}\left(t|t-1\right) - \tilde{x}(t)$$

and recalling the expression of the output variation $\Delta y(t)$, we may rewrite the state prediction estimate as

$$\begin{aligned}
\hat{x}\left(t+1|t\right) &= \tilde{x}(t+1) + \widehat{\Delta x}\left(t+1|t\right) \\
&= f\left(\tilde{x}(t)\right) + \tilde{F}(t)\widehat{\Delta x}\left(t|t-1\right) + \tilde{K}(t)e(t) \\
&= f\left(\tilde{x}(t)\right) + \tilde{F}(t)\widehat{\Delta x}\left(t|t-1\right) \\
&\quad + \tilde{K}(t)\left\{y(t) - \left[\tilde{y}(t) + \tilde{H}(t)\,\widehat{\Delta x}\left(t|t-1\right)\right]\right\}
\end{aligned}$$

$$\approx f\left(\hat{x}\left(t|t-1\right)\right)$$

$$\approx h\left(\hat{x}\left(t|t-1\right)\right)$$

Linearised Kalman Predictor and Filter

Summarising, given the nonlinear system

$$\begin{cases} x(t+1) = f\big(x(t)\big) + v_1(t) & x,\, v_1 \in \mathbb{R}^n \\ y(t) = h\big(x(t)\big) + v_2(t) & y,\, v_2 \in \mathbb{R}^p \end{cases}$$

the approximate predicted state estimate can be expressed as

Linearised Kalman Predictor

$$\hat{x}\left(t+1|t\right) = f\left(\hat{x}\left(t|t-1\right)\right) + \tilde{K}(t)\left[y(t) - h\left(\hat{x}\left(t|t-1\right)\right)\right]$$

The gain $\tilde{K}(t)$ is evaluated using the usual expression of **the Kalman predictor gain**, employing the solution of the **Riccati equation**, computed with matrices

$$\tilde{F}(t) = \frac{\partial\, f(x)}{\partial x}\Big|_{x=\tilde{x}(t)} \qquad \tilde{H}(t) = \frac{\partial\, h(x)}{\partial x}\Big|_{x=\tilde{x}(t)}$$

where $\tilde{x}(\cdot)$ is the nominal state movement, solution of
$\tilde{x}(t+1) = f\big(\tilde{x}(t)\big)$ with $\tilde{x}(1) = \mathrm{E}\left[x(1)\right]$

Linearised Kalman Predictor and Filter (cont.)

In a similar way, the approximated filtered state estimate can be written as

Linearised Kalman Filter

$$\hat{x}\left(t|t\right) = f\left(\hat{x}\left(t-1|t-1\right)\right) + \tilde{K}_0(t)\left[y(t) - h\left(\hat{x}\left(t-1|t-1\right)\right)\right]$$

The gain $\tilde{K}_0(t)$ is computed using the usual expression of **the Kalman filter gain**, employing the solution of the **Riccati equation**, computed with matrices

$$\tilde{F}(t) = \left.\frac{\partial f(x)}{\partial x}\right|_{x=\tilde{x}(t)} \qquad \tilde{H}(t) = \left.\frac{\partial h(x)}{\partial x}\right|_{x=\tilde{x}(t)}$$

where $\tilde{x}(\cdot)$ is the nominal state movement, solution of
$\tilde{x}(t+1) = f\big(\tilde{x}(t)\big)$ with $\tilde{x}(1) = \mathrm{E}\left[x(1)\right]$

Remarks

Important Considerations about the Linearisation

- The *linearised Kalman predictor* and the *linearised Kalman filter* have been obtained by approximating the nonlinear functions $f(\cdot)$ and $h(\cdot)$ around the nominal state movement $\tilde{x}(t)$.
- The nominal state movement has been computed by solving the difference equation

$$\begin{aligned} \tilde{x}(t+1) &= f\big(\tilde{x}(t)\big) \\ \tilde{x}(1) &= \mathrm{E}\left[x(1)\right] \end{aligned}$$

  without taking into consideration any measured data.
- The true state movement may diverge, step by step, from the nominal one. The performance of the predictor/filter may deteriorate in the long run.
- How to cope with this issue?

Kalman Estimation: the Extended Kalman Filter

The Extended Kalman Filter

The Extended Kalman Filter

Instead to consider an off-line precalculated state movement as nominal state for the linearisation, a better approximation can be achieved by linearising at each step around the last estimate of the state:

$$\hat{F}\left(t|t-1\right) = \frac{\partial\, f(x)}{\partial x}\Big|_{x=\hat{x}(t|t-1)} \qquad \hat{H}\left(t|t-1\right) = \frac{\partial\, h(x)}{\partial x}\Big|_{x=\hat{x}(t|t-1)}$$

This means that we can no longer solve the Riccati equation off-line, as we could in the case of the linearized Kalman predictor: the solution of the DRE must be performed just after the estimate $\hat{x}\left(t|t-1\right)$ is available.

The Extended Kalman Filter (cont.)

The Extended Kalman Filter (EKF) as State Predictor

Given the nonlinear system

$$\begin{cases} x(t+1) = f\big(x(t)\big) + v_1(t) & x,\, v_1 \in \mathbb{R}^n\,,\ v_1 \sim \mathsf{WGN}(0, V_1) \\ y(t) = h\big(x(t)\big) + v_2(t) & y,\, v_2 \in \mathbb{R}^p\,,\ v_2 \sim \mathsf{WGN}(0, V_2) \end{cases}$$

the **approximate predicted state estimate** can be expressed as

$$\hat{x}\,(t+1|t) = f\left(\hat{x}\,(t|t-1)\right) + \tilde{K}(t)\left[y(t) - h\left(\hat{x}\,(t|t-1)\right)\right]$$

The gain $\tilde{K}(t)$ is evaluated using the usual expression of **the Kalman predictor gain**, employing the solution of the **Riccati equation**, computed with matrices

$$\hat{F}\,(t|t-1) = \frac{\partial\, f(x)}{\partial x}\Big|_{x=\hat{x}(t|t-1)} \qquad \hat{H}\,(t|t-1) = \frac{\partial\, h(x)}{\partial x}\Big|_{x=\hat{x}(t|t-1)}$$

and with $\hat{x}\,(1|0) = \mathrm{E}\,[x(1)] \qquad \hat{P}(1) = \hat{P}_1 = \mathrm{var}\,[x(1)]$

Kalman Estimation: the Extended Kalman Filter

Generalisation

The EKF Predictor in the Presence of Exogenous Inputs

Consider the system model

$$x(t+1) = f\left(x(t), u(t)\right) + v_1(t) \qquad v_1(\cdot) \sim \text{WGN}\,(0,\, V_1)$$
$$y(t) = h\left(x(t)\right) + v_2(t) \qquad v_2(\cdot) \sim \text{WGN}\,(0,\, V_2)$$

and linearise the state and output equations around the state estimate $\hat{x}\,(t|t-1)$, as usual.

This results in the following equations for the EKF predictor

$$\hat{x}\,(t+1|t) = f\left(\hat{x}\,(t|t-1)\,,\ u(t)\right) + K(t)\left[y(t) - h\left(\hat{x}\,(t|t-1)\right)\right]$$
$$K(t) = F_t\,P(t)\,H_t^\top\left[H_t P(t) H_t^\top + V_2\right]^{-1}$$
$$P(t+1) = F_t\left\{P(t) - P(t)H_t^\top\left[V_2 + H_t P(t) H_t^\top\right]^{-1} H_t P(t)\right\} F_t^\top + V_1$$

where

$$F_t = \left.\frac{\partial\, f(x,\,u)}{\partial x}\right|_{\substack{x\,=\,\hat{x}\,(t|t-1)\\ u\,=\,u(t)}} \qquad H_t = \left.\frac{\partial\, h(x)}{\partial x}\right|_{x=\hat{x}(t|t-1)}$$

Kalman Estimation: the Extended Kalman Filter

Concluding Remarks

EKF Predictor and Filter Properties

- The Extended Kalman Filter (EKF) is nothing else than a standard and exact Kalman filter for the linearised system. When applied to the nonlinear system, **the EKF is no more linear or optimal**.
- The notations $\hat{x}(t|t-1)$ and $P(t|t-1)$ are now denoting approximate conditional mean values and covariances.

EKF Predictor and Filter Properties (cont.)

- The equations for calculating the filter gain $K(t)$ and the covariance $P(t)$ are coupled to the filter equations, since F_t and H_t are function of the previous filter estimate $\hat{x}\,(t|t-1)$. So in general the evaluation of $K(t)$ and $P(t)$ cannot be carried out off-line.
- The smaller will be $\|x(t) - \hat{x}\,(t|t-1)\,\|^2$ or $\|x(t) - \hat{x}\,(t|t)\,\|^2$ the better the linearised system will approximate the original nonlinear system. Therefore, **in high signal-to-noise ratio situations, there would be fewer difficulties in using an Extended Kalman Filter**. However, in general, **convergence is not guaranteed**.

Kalman Estimation: the Extended Kalman Filter

Application Example: a Two-Phase PMSM Drive

PMSM Motor State Estimation

- Consider a two-phase permanent magnet synchronous drive: we would like to estimate the states of the motor, for control purposes or for some other different reasons.
- We can measure the motor winding currents, but we can't acquire the rotor position and angular speed, so we want to use an EKF to estimate both the rotor position and the rotational speed.

PMSM Motor State Estimation (cont.)

- The system equations are

$$
\begin{aligned}
\dot{I}_a &= -\frac{R}{L} I_a + \frac{\omega\lambda}{L} \sin\vartheta + \frac{u_a + \eta_a}{L} \\
\dot{I}_b &= -\frac{R}{L} I_b - \frac{\omega\lambda}{L} \cos\vartheta + \frac{u_b + \eta_b}{L} \\
\dot{\omega} &= -\frac{3\lambda}{2\,J} I_a \, \sin\vartheta + \frac{3\lambda}{2\,J} I_b \, \cos\vartheta - \frac{F\,\omega}{J} + \eta_\alpha \\
\dot{\vartheta} &= \omega \\
y_1 &= I_a + \epsilon_1 \\
y_2 &= I_b + \epsilon_2
\end{aligned}
$$

The variables are defined as follows

- I_a and I_b are the current intensities in the two motor windings.
- R and L are the motor winding's resistance and inductance.
- ϑ and ω are respectively the angular position and the rotational speed of the rotor.

PMSM Motor State Estimation (cont.)

The remaining variables are defined as follows

- λ is the flux constant of the motor.
- F and J are respectively the the coefficient of viscous friction that acts on the motor shaft and the moment of inertia of the motor shaft and its load.
- u_a and u_b are the voltages that are applied across the two motor windings. The terms η_a and η_b are noise terms, taking into account possible errors in the applied voltages.
- η_α is a noise term due to uncertainty in the load torque.
- y_1 and y_2 are the measurements of the winding currents, distorted by measurement noises ϵ_1 and ϵ_2.

We assume known and constant the values of R, L, F, J, λ and the noise variances. Moreover, we assume that the noises are uncorrelated.

PMSM Motor State Estimation (cont.)

The state vector is $x = [I_a\ ,\ I_b\ ,\ \omega\ ,\ \vartheta]^T$. By discretising the system equation with sampling time period Δ, we obtain

$$
\begin{aligned}
x(t+1) &= x(t) + \begin{bmatrix} -\dfrac{Rx_1(t)}{L} + \dfrac{x_3(t)\lambda \sin x_4(t)}{L} + \dfrac{u_1(t)}{L} \\ -\dfrac{Rx_2(t)}{L} + \dfrac{x_3(t)\lambda \cos x_4(t)}{L} + \dfrac{u_2(t)}{L} \\ -\dfrac{3\lambda x_1(t) \sin x_4(t)}{2\,J} + \dfrac{3\lambda x_2(t) \cos x_4(t)}{2\,J} - \dfrac{F x_3(t)}{J} \\ x_3(t) \end{bmatrix} \Delta + \begin{bmatrix} \dfrac{\eta_a(t)}{L} \\ \dfrac{\eta_b(t)}{L} \\ \eta_\alpha(t) \\ 0 \end{bmatrix} \Delta \\
&= f\left(x(t),\, u(t)\right) + v_1(t)
\end{aligned}
$$

PMSM Motor State Estimation (cont.)

The discretised output equation appears simply as follows

$$
\begin{aligned}
y(t) &= \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} \epsilon_1(t) \\ \epsilon_2(t) \end{bmatrix} \\
&= h\left(x(t)\right) + v_2(t)
\end{aligned}
$$

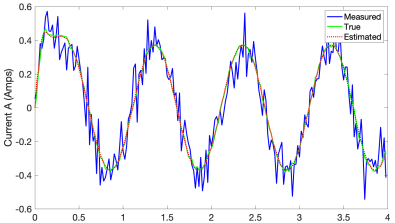
Let's assume

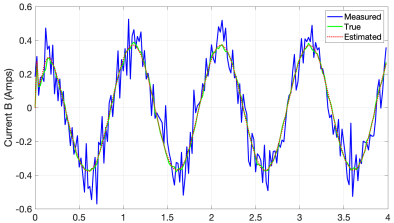
$$
\begin{aligned}
\epsilon_1(\cdot) &\sim \mathsf{WGN}\left(0, 10^{-2}\right) & \epsilon_2(\cdot) &\sim \mathsf{WGN}\left(0, 10^{-2}\right) \\
\eta_a(\cdot) &\sim \mathsf{WGN}\left(0, 10^{-6}\right) & \eta_b(\cdot) &\sim \mathsf{WGN}\left(0, 10^{-6}\right) \\
\eta_\alpha(\cdot) &\sim \mathsf{WGN}\left(0, 25 \cdot 10^{-4}\right)
\end{aligned}
$$

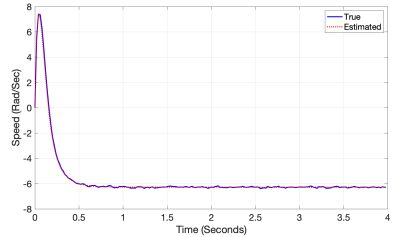
Moreover, let's assume that the sampled control inputs are (sampling period Δ)

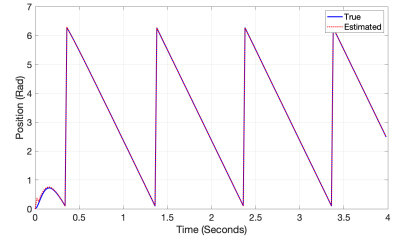
$$
u_1(t) = \sin 2\pi\Delta t \qquad u_2(t) = \cos 2\pi\Delta t \quad t \in \mathbb{Z}
$$

PMSM Motor State Estimation: the Results









267MI –Fall 2023

Lecture 14
State Estimation from Observed Data

END