Systems Dynamics

Course ID: 267MI – Fall 2018

Thomas Parisini
Gianfranco Fenu

University of Trieste
Department of Engineering and Architecture

267MI –Fall 2018

Lecture 13
State estimation from observed data

Kalman estimation

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

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

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

$$\boxed{h^{\circ}(x) = E\left(\vartheta \,|\, d = x\right)}$$

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

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

- We obtain:

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

- expected value $\dfrac{\lambda_{\vartheta d}}{\lambda_{dd}}\, d$
- 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

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

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

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

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

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

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:

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

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

Optimal linear estimator

Optimal Linear Estimator (cont.)

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

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

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

$$\boxed{\begin{aligned} \hat{\vartheta} &= \vartheta_m + \frac{\lambda_{\vartheta d}}{\lambda_{dd}}\,(d - d_m) \\ \mathrm{var}\,(\vartheta - \hat{\vartheta}) &= \lambda_{\vartheta\vartheta} - \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}} \end{aligned}}$$

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

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

$$\boxed{\begin{aligned} \hat{\vartheta} &= \vartheta_m + \Lambda_{\vartheta d}\,{\Lambda_{dd}}^{-1}\,(d - d_m) \\ \mathrm{var}\,(\vartheta - \hat{\vartheta}) &= \Lambda_{\vartheta\vartheta} - \Lambda_{\vartheta d}\,{\Lambda_{dd}}^{-1}\Lambda_{d\vartheta} \end{aligned}}$$

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:

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

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:

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

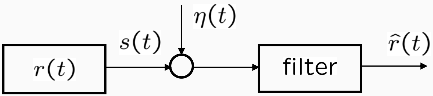
  and thus $\mathrm{var}\,(\vartheta - \hat{\vartheta}) \leq \mathrm{var}\,(\vartheta)$ and

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

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

Kalman estimation

- 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).



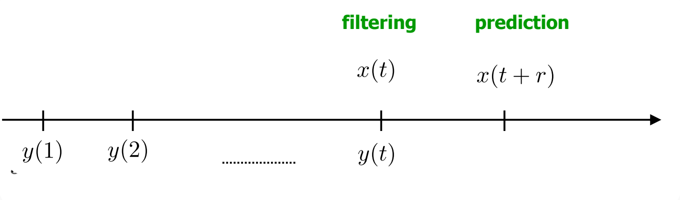
Example: signal filtering

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 $y(t),\ y(t-1),\ \ldots$ 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:

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

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

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

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 $\mathrm{E}[\vartheta \mid d(1)] = \dfrac{\lambda_{\vartheta 1}}{\lambda_{11}} d(1)$

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

Recursive form of Bayes estimation (cont.)

recursion

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

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

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

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

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

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 $\vartheta,\ d(1),\ e$ 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

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

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

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

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

Recursive form of Bayes estimation (cont.)

- Observe that

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

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:

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

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 $\vartheta,\ d(1),\ d(2)$ 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:

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

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):

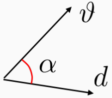
- Suppose that d and ϑ are marginally and jointly Gaussian random variables:

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

  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:

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

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



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

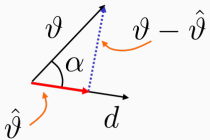
- Uncorrelated random variables correspond to orthogonal vectors

Geometric interpretation of Bayes recursive estimation (cont.)

- Now:

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

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:

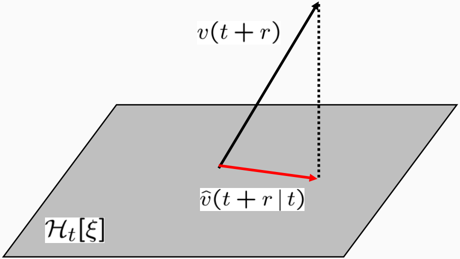
$$\|\vartheta - \hat{\vartheta}\|^2 = \|\vartheta\|^2 - \|\hat{\vartheta}\|^2 = \|\vartheta\|^2 - \|\vartheta\|^2\, (\cos\alpha)^2$$
$$= \lambda_{\vartheta\vartheta} - \lambda_{\vartheta\vartheta}\, \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}\lambda_{\vartheta\vartheta}} = \lambda_{\vartheta\vartheta} - \frac{\lambda_{\vartheta d}^2}{\lambda_{dd}}$$

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)

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

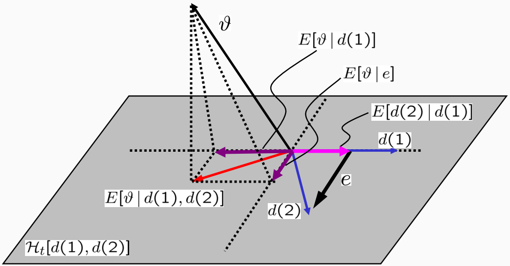


Geometric interpretation of Bayes recursive estimation (cont.)

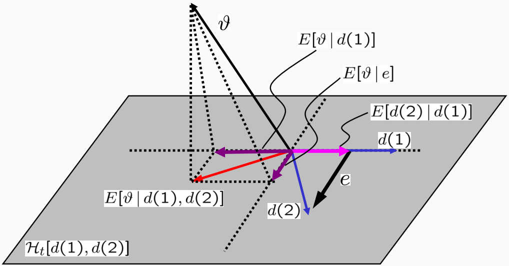
- If

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

  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

  - $\mathrm{E}[\vartheta \mid d(1), d(2)] \neq \mathrm{E}[\vartheta \mid d(1)] + \mathrm{E}[\vartheta \mid d(2)]$

One-step ahead Kalman predictor

- Consider the dynamic system

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

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

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

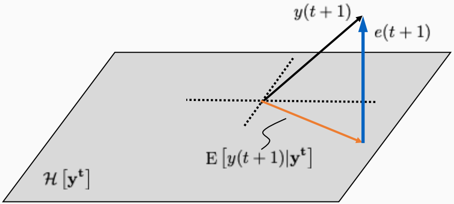
One-step ahead Kalman predictor (cont.)

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:

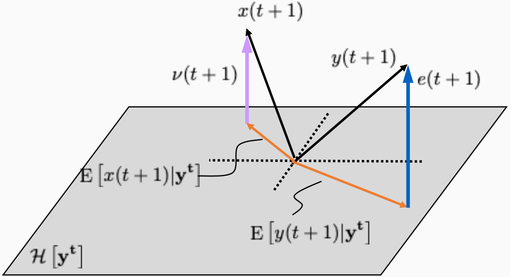


One-step ahead Kalman predictor (cont.)

- The state prediction error is:

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

and thus the situation now is:



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

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

$$x(t) = f\left[\mathbf{v}_1^{\mathbf{t}-1}, x(1)\right] = f[v_1(t-1), v_1(t-2), \ldots, v_1(1), x(1)]$$

$$y(t) = \bar{f}\left[\mathbf{v}_1^{\mathbf{t}-1}, x(1),\ v_2(t)\right] \quad \Longrightarrow \quad \mathbf{y^t} = \bar{f}\left[\mathbf{v}_1^{\mathbf{t}-1}, x(1),\ \mathbf{v}_2^{\mathbf{t}}\right]$$

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

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

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

Recursive one-step ahead prediction

- We have

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

- From the recursive Bayes formula:

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

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

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

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

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

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

Recursive one-step ahead prediction (cont.)

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

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

  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:

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

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.)

- $(\star\star)$ $\quad \mathrm{E}\left\{v_1(t)\left[H\left(x(t)-\hat{x}(t\,|\,t1)\right)+v_2(t)\right]^\top\right\}$

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

- but $v_1(\cdot)$ white $\Longrightarrow v_1(t)$ independent with $\mathbf{v}_2^{\mathbf{t}-1}$
- $v_1(\cdot)$ 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 $\hat{x}(t \,|\, t-1)$ 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.

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

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

- 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:

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

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

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

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

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

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

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

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

  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

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

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

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 $v_2(t)$:

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

Riccati equation

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

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)

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

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:

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

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:

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

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

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

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

$$\hat{x}(2\,|\,1) = F P_1 H^\top \left(H P_1 H^\top + V_2\right)^{-1} y(1) \qquad (\star)$$

Interpretation

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

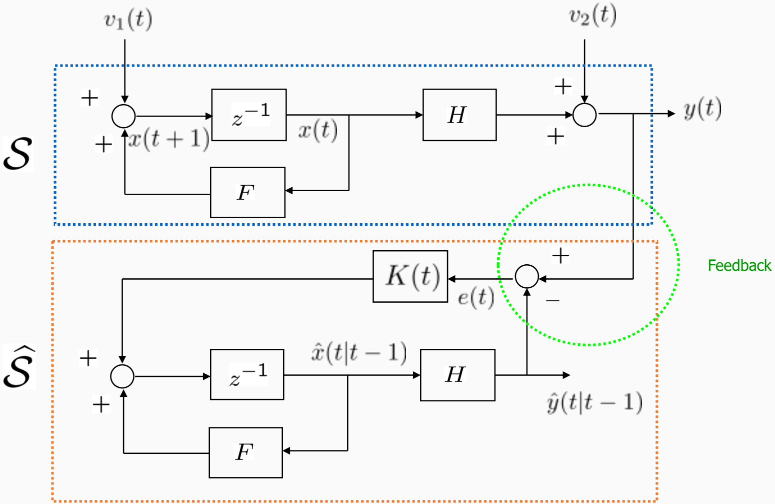
then relation $(\star)$ is "compatible" with the recursive one and the interpretation is obvious: *a priori*, without available data, the more reasonable estimate is the *a priori* expected value.

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 $V_2 > 0$.

267MI –Fall 2018

Lecture 13

State estimation from observed data

END