KALMAN FILTER THEORY

Michele TARAGNA

Dipartimento di Elettronica e Telecomunicazioni

Politecnico di Torino

michele.taragna@polito.it

Master Course in Mechatronic Engineering

Master Course in Computer Engineering

01RKYQW / 01RKYOV "Estimation, Filtering and System Identification"

Academic Year 2019/2020

Background: state estimation in a deterministic noise-free framework

Setup: let us consider a discrete-time, linear time-invariant (LTI), dynamical system $\mathcal{S}$ described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = Ax(t) + Bu(t) & \text{(state equation)} \\ y(t) = Cx(t) + Du(t) & \text{(output equation)} \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^p$, $y(t) \in \mathbb{R}^q$ are deterministic, and assume that:

- $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{q \times n}$ are known constant matrices
- $D = \mathbf{0}_{q \times p}$, i.e., $\mathcal{S}$ is a physically realizable system
- the input $u(t)$ is known $\forall t \geq 1$
- the output $y(t)$ is measured $\forall t \geq 1$, <u>without</u> any noise or disturbance
- the initial state $x(t=1)$ is unknown

Goal: find a state estimate $\hat{x}(t)$ that suitably approximates the state $x(t)$, $\forall t \geq 1$

Solution #1: use a copy $\mathcal{S}'$ of the system $\mathcal{S}$, fed by the same input $u(t)$ but starting from a different initial state $\hat{x}(t=1) \neq x(t=1)$:

$$\mathcal{S}' : \begin{cases} \hat{x}(t+1) = A\hat{x}(t) + Bu(t) \\ \hat{y}(t) = C\hat{x}(t) \end{cases} \qquad t = 1, 2, \ldots$$

To evaluate the performances, let us define the state estimation error:

$$\tilde{x}(t) = x(t) - \hat{x}(t)$$

that solves the following difference equation:

$$\begin{aligned} \tilde{x}(t+1) &= x(t+1) - \hat{x}(t+1) = Ax(t) + Bu(t) - [A\hat{x}(t) + Bu(t)] = \\ &= Ax(t) - A\hat{x}(t) = A\left[x(t) - \hat{x}(t)\right] = A\tilde{x}(t) \end{aligned}$$

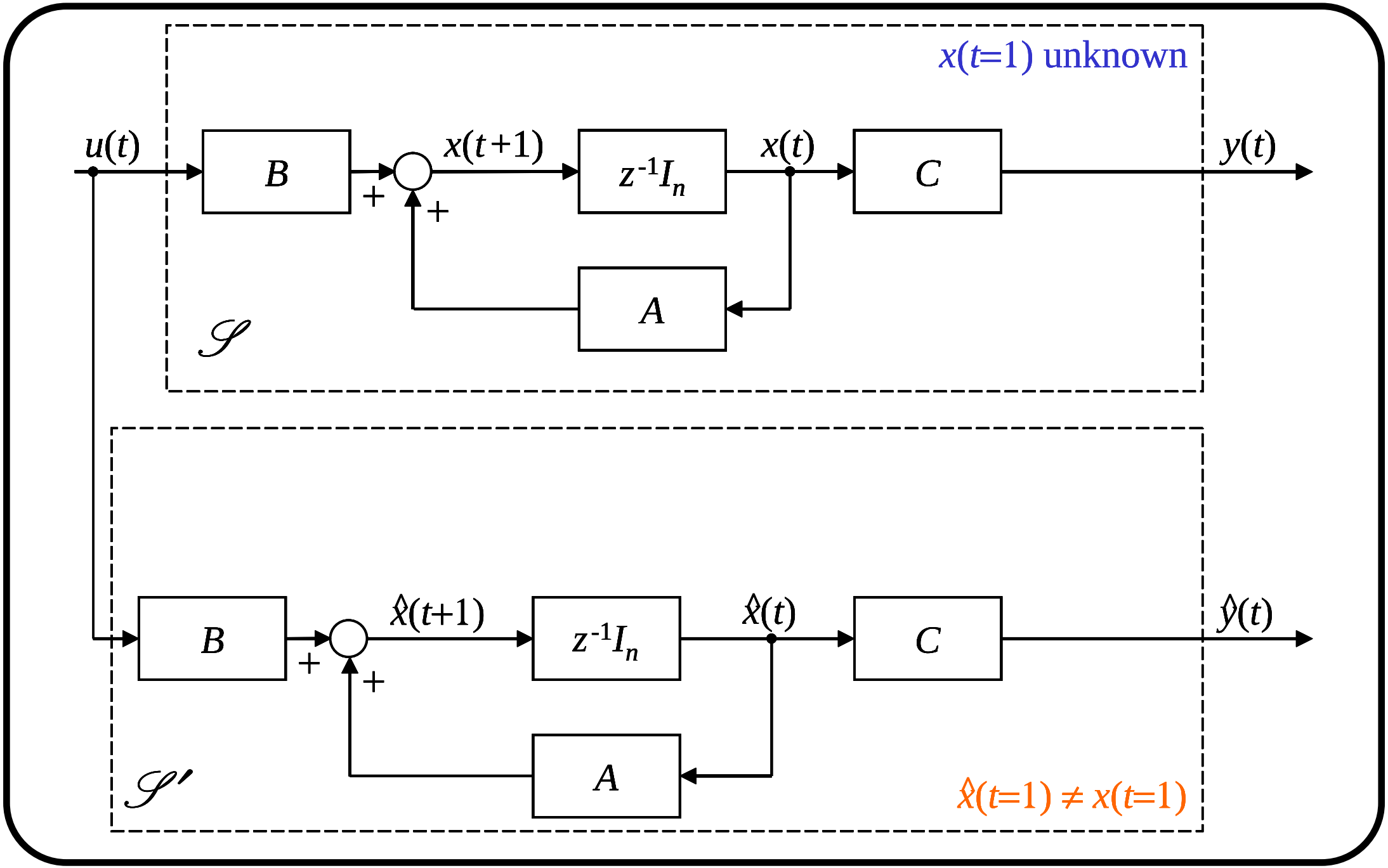
$$\Downarrow$$

$$\tilde{x}(t) = A^{t-1}\tilde{x}(t=1) = A^{t-1}\left[x(t=1) - \hat{x}(t=1)\right], \quad \forall t \geq 1$$

Drawback: only if A is asymptotically stable (i.e., such that $|\lambda_i(A)| < 1, \forall i$) $\Rightarrow$

$$\lim_{t\to\infty} \tilde{x}(t) = \mathbf{0}, \forall \hat{x}(t=1)$$

i.e., $\hat{x}(t)$ asymptotically converges to $x(t)$ independently of $\hat{x}(t=1)$


x(t=1) unknown
u(t)
B
+
+
x(t+1)
z-1In
x(t)
C
y(t)
A
𝒮
B
+
+
x̂(t+1)
z-1In
x̂(t)
C
ŷ(t)
A
𝒮′
x̂(t=1) ≠ x(t=1)

Solution #2 (Asymptotic State Observer): modify the Solution #1 in order to exploit the information on $y(t)$, using a copy $\mathcal{S}''$ of the system $\mathcal{S}$, fed by the same input $u(t)$ and with initial state $\hat{x}(t=1) \neq x(t=1)$, adding a term that linearly depends on $y(t)$:

$$\mathcal{S}'' : \begin{cases} \hat{x}(t+1) = A\hat{x}(t) + Bu(t) + L[y(t) - \hat{y}(t)] \\ \hat{y}(t) = C\hat{x}(t) \end{cases} \qquad t = 1, 2, \ldots$$

The state estimation error $\tilde{x}(t) = x(t) - \hat{x}(t)$ now solves this difference equation:

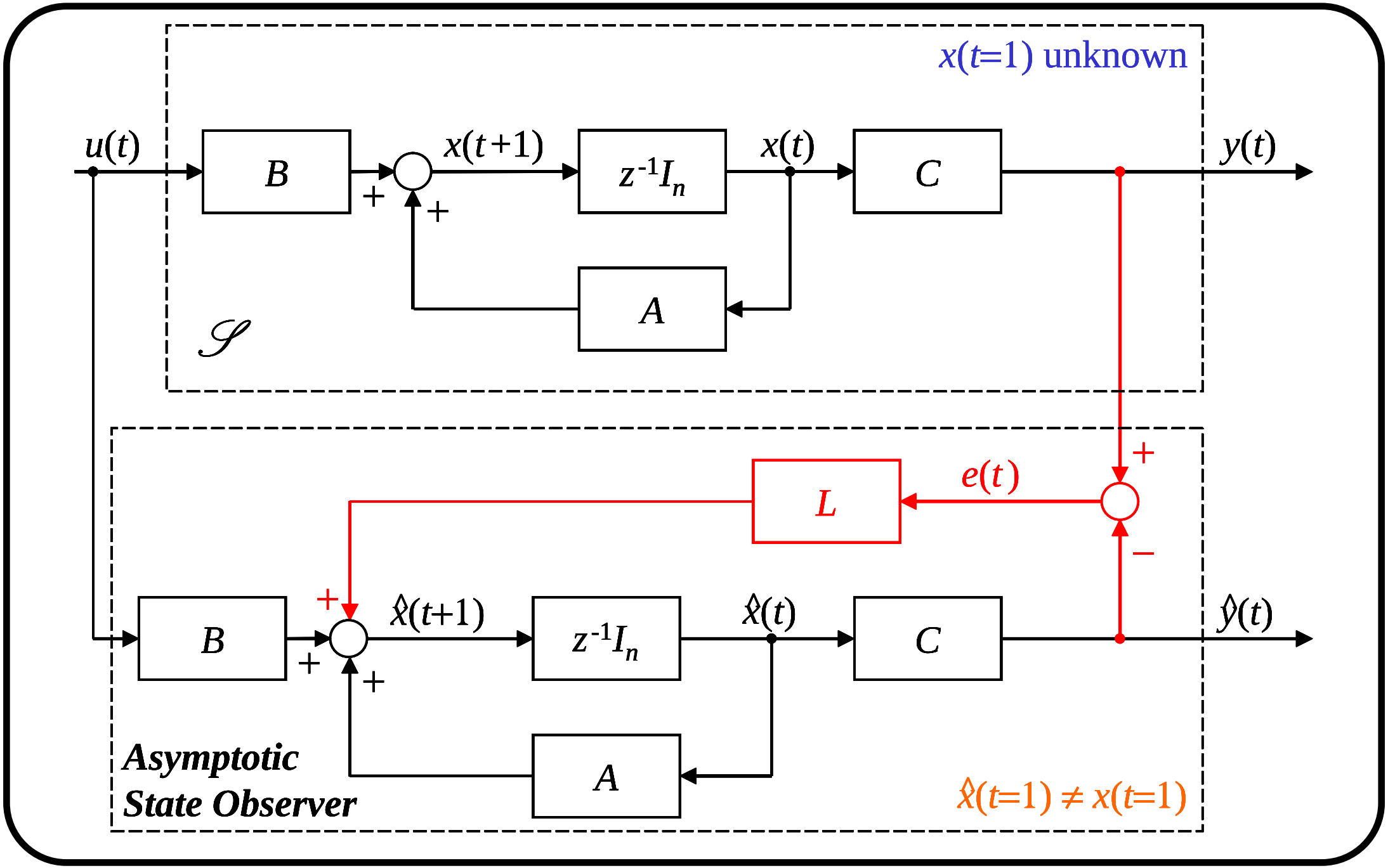
$$\begin{aligned} \tilde{x}(t+1) &= x(t+1) - \hat{x}(t+1) = Ax(t) + Bu(t) - \{A\hat{x}(t) + Bu(t) + L[y(t) - \hat{y}(t)]\} \\ &= Ax(t) - A\hat{x}(t) - L[Cx(t) - C\hat{x}(t)] = A\,[x(t) - \hat{x}(t)] - LC[x(t) - \hat{x}(t)] \\ &= (A - LC)\,[x(t) - \hat{x}(t)] = (A - LC)\tilde{x}(t) \end{aligned}$$

$$\Downarrow$$

$$\tilde{x}(t) = (A - LC)^{t-1}\,\tilde{x}(t=1) = (A - LC)^{t-1}\,[x(t=1) - \hat{x}(t=1)]\,, \quad \forall t \geq 1$$

If $A - LC$ is such that $|\lambda_i(A - LC)| < 1, \forall i \;\Rightarrow\; \lim_{t\to\infty} \tilde{x}(t) = \mathbf{0}, \forall \hat{x}(t=1)$,

i.e., $\hat{x}(t)$ asymptotically converges to $x(t)$ independently of $\hat{x}(t=1)$


x(t=1) unknown
u(t)
B
+
+
x(t+1)
z-1In
x(t)
C
y(t)
A
𝒮
L
e(t)
+
−
+
B
+
+
x̂(t+1)
z-1In
x̂(t)
C
ŷ(t)
A
Asymptotic
State Observer
x̂(t=1) ≠ x(t=1)

Main result: if the system $\mathcal{S}$ is fully observable, i.e., if the observability matrix

$$M_O = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \in \mathbb{R}^{nq \times n}$$

is full rank (i.e., $\rho\left(M_O\right) = n$), then the matrix $L \in \mathbb{R}^{n \times q}$ can be designed such that

$$|\lambda_i(A - LC)| < 1, \ \forall i$$

$$\Downarrow$$

$\lim_{t \to \infty} \tilde{x}(t) = \mathbf{0}$, $\forall \hat{x}(t=1)$, i.e., $\hat{x}(t)$ asymptotically converges to $x(t)$ $\forall \hat{x}(t=1)$

Kalman filtering problem

Let us consider a discrete-time, linear time-invariant (LTI), dynamical system $\mathcal{S}$ described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = Ax(t) + v_1(t) \\ y(t) = Cx(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

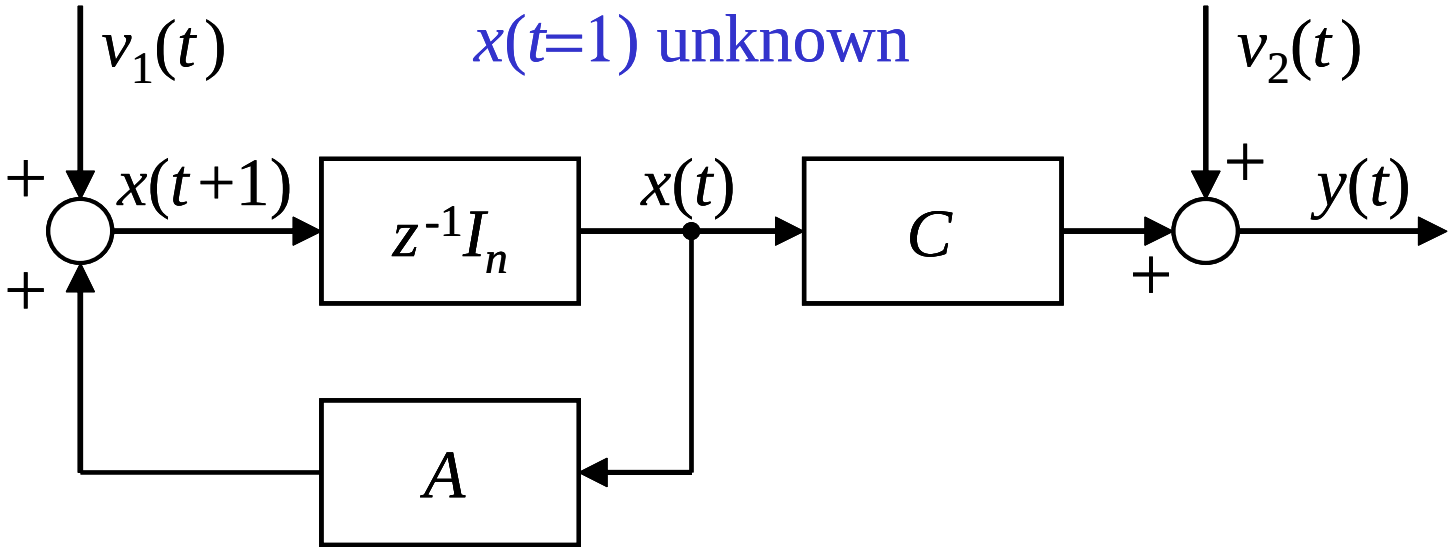
where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- the process noise $v_1(t)$ and the measurement noise $v_2(t)$ are white noises with zero mean value and known variance which are uncorrelated with each other, i.e., $v_1(t) \sim WN(\mathbf{0}, V_1)$ and $v_2(t) \sim WN(\mathbf{0}, V_2)$ are such that:

$$E[v_1(t_1)v_1(t_2)^T] = V_1\delta(t_2 - t_1) \qquad \text{(whiteness of } v_1(t)\text{)}$$
$$E[v_2(t_1)v_2(t_2)^T] = V_2\delta(t_2 - t_1) \qquad \text{(whiteness of } v_2(t)\text{)}$$
$$E[v_1(t_1)v_2(t_2)^T] = \mathbf{0}_{n\times q}, \quad \forall t_1, t_2 \qquad \text{(uncorrelation of } v_1(t) \text{ and } v_2(t)\text{)}$$

- $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{q\times n}$, $V_1 \in \mathbb{R}^{n\times n}$, $V_2 \in \mathbb{R}^{q\times q}$ are known matrices

- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\mathbf{0}, P_1)$, with known $P_1 \in \mathbb{R}^{n \times n}$, which is uncorrelated with noises $v_1(t)$ and $v_2(t)$
- the output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$



- **Goal:** estimate the state $x(N+r)$:
  - if $r = 1 \quad \Rightarrow \quad$ one-step prediction problem
  - if $r > 1 \quad \Rightarrow \quad$ multi-step prediction problem
  - if $r = 0 \quad \Rightarrow \quad$ filtering problem
  - if $r < 0 \quad \Rightarrow \quad$ smoothing (or regularization) problem

- $v_1(t)$ and $v_2(t)$ are random variables $\Rightarrow x(t)$ and $y(t)$ are random variables too

$$\Downarrow$$

the Bayesian estimate of the state $x(N+r)$ given the N measurements $y(N), y(N-1), \ldots, y(1)$, is equal to:

$$\hat{x}(N+r|N) = E\left[x(N+r)|\, d\right] = \bar{x}(N+r) + \Sigma_{x(N+r)d}\Sigma_{dd}^{-1}\left(d-\bar{d}\right)$$

where
$$\begin{aligned}
\bar{x}(N+r) &= E[x(N+r)] \in \mathbb{R}^n \\
d = y^N &= \left[y(N)^T\ y(N-1)^T\ \cdots\ y(1)^T\right]^T \in \mathbb{R}^{Nq} \\
\bar{d} &= E[d] = E[y^N] \in \mathbb{R}^{Nq} \\
\Sigma_{dd} &= E\left[(d-\bar{d})(d-\bar{d})^T\right] \in \mathbb{R}^{Nq\times Nq} \\
\Sigma_{x(N+r)d} &= E\left[\left(x(N+r)-\bar{x}(N+r)\right)(d-\bar{d})^T\right] \in \mathbb{R}^{n\times Nq}
\end{aligned}$$

- $v_1(t)$, $v_2(t)$ have zero mean value, $\forall t \Rightarrow x(t)$, $y(t)$ have zero mean value, $\forall t$

$$\Downarrow$$

$$\hat{x}(N+r|N) = E\left[x(N+r)|\, y^N\right] = \Sigma_{x(N+r)y^N}\Sigma_{y^Ny^N}^{-1}y^N$$

- The main drawback of the proposed form of the Bayesian estimate
$$\hat{x}(N+r|N) = E\left[x(N+r)|\, y^N\right] = \Sigma_{x(N+r)y^N}\Sigma^{-1}_{y^N y^N} y^N$$
is that it requires the *batch processing* of the measurements, since all information is incorporated in one single step into the estimate
$$\Downarrow$$
this leads to the necessity for inverting $\Sigma_{y^N y^N} \in \mathbb{R}^{Nq\times Nq}$, which may be a very difficult task when N is large

- To avoid this drawback, *recursive* or *sequential estimation* schemes are looked for, in which the current estimate depends on the previous estimate and the current measurement
$$\Downarrow$$
such schemes rely on *sequential processing* of data, where the measurements are processed in stages

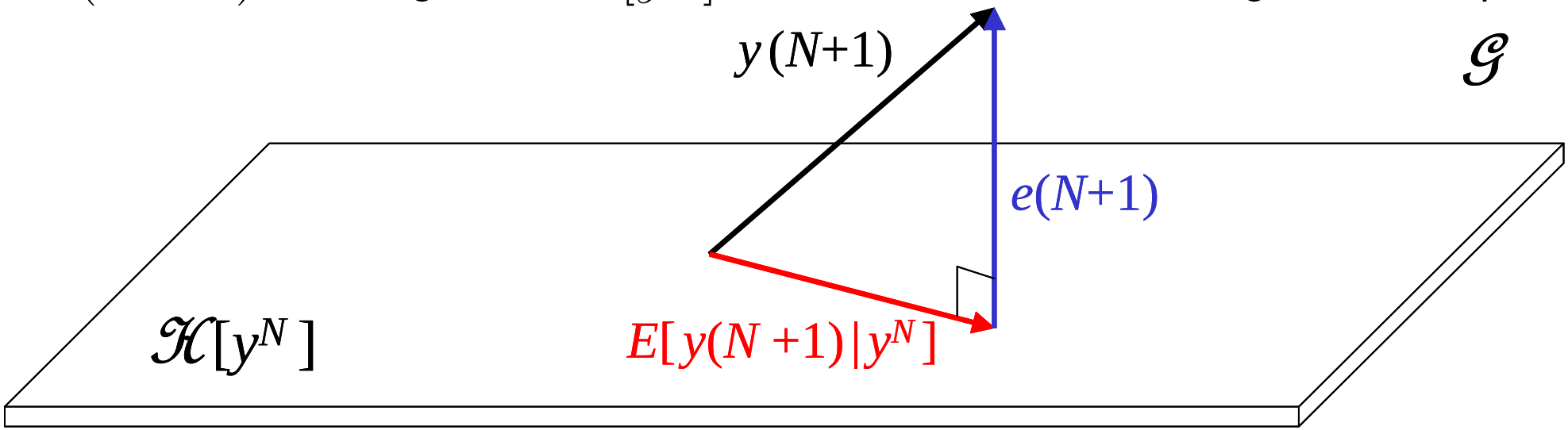
- Using the recursive Bayesian estimation method:
  - first, the one-step prediction problem will be solved
  - then, the multi-step prediction and the filtering problems will be afforded

One-step Kalman predictor

- **Goal:** given the data vector $y^N = \left[y(N)^T \; y(N-1)^T \; \cdots \; y(1)^T\right]^T$, find a recursive Bayesian estimate $\hat{x}(N+1|N)$ of the state $x(N+1)$ starting from the estimate $\hat{x}(N|N-1)$ of the state $x(N)$ obtained at the previous stage
- The **innovation of** $y(N+1)$ **given** y^N is defined by:

$$e(N+1) = y(N+1) - E\left[y(N+1)|\, y^N\right] \in \mathbb{R}^q$$

where $E\left[y(N+1)|\, y^N\right]$ is the projection of $y(N+1)$ over the subspace $\mathcal{H}[y^N]$ generated by all the components of y^N, called *past subspace* $\Rightarrow$ $e(N+1)$ is orthogonal to $\mathcal{H}[y^N]$, i.e., the innovation is orthogonal to the past

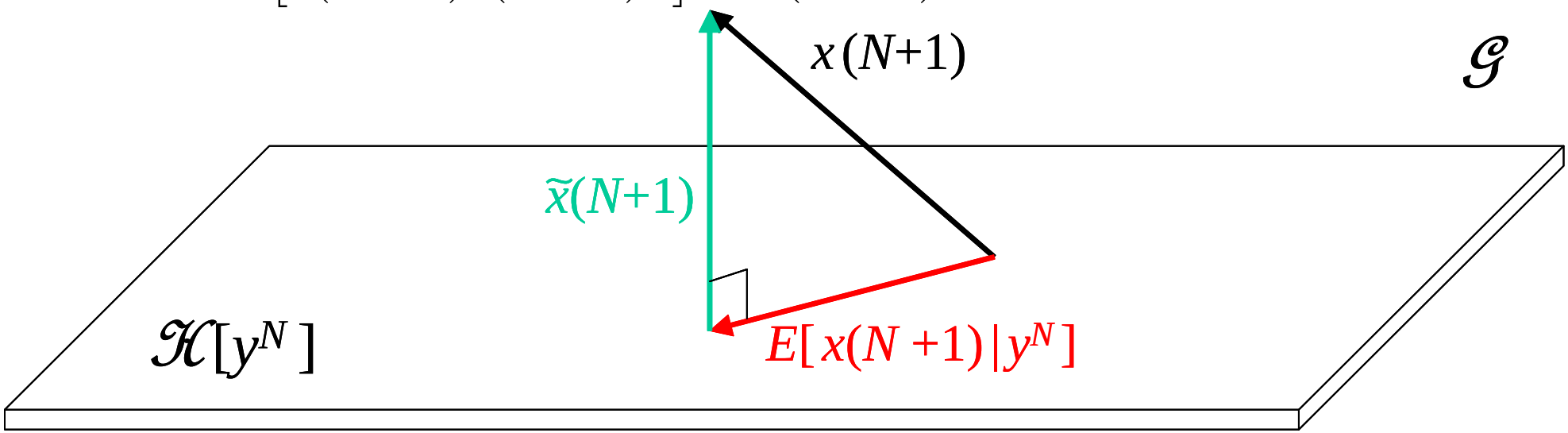


- The **prediction error of the state** $x(N+1)$ is defined by:

$$\tilde{x}(N+1) = x(N+1) - E\left[x(N+1)|\, y^N\right] \in \mathbb{R}^n$$

where $E\left[x(N+1)|\, y^N\right]$ is the projection of $x(N+1)$ over $\mathcal{H}[y^N]$ $\Rightarrow$ $\tilde{x}(N+1)$ is orthogonal to $\mathcal{H}[y^N]$, i.e., it is orthogonal to the past

- $E[\tilde{x}(N+1)] = E\left[x(N+1) - \Sigma_{x(N+1)y^N}\Sigma^{-1}_{y^N y^N} y^N\right] =$
  $= E[x(N+1)] - \Sigma_{x(N+1)y^N}\Sigma^{-1}_{y^N y^N} E\left[y^N\right] = \mathbf{0}$
- $Var[\tilde{x}(N+1)] = E\left[(\tilde{x}(N+1) - E[\tilde{x}(N+1)])\,(\tilde{x}(N+1) - E[\tilde{x}(N+1)])^T\right] =$
  $= E\left[\tilde{x}(N+1)\tilde{x}(N+1)^T\right] = P(N+1)$



- $e(N+1)$ and $\tilde{x}(N+1)$ are orthogonal to $\mathcal{H}[y^N] \Rightarrow$
  $e(N+1)$ and $\tilde{x}(N+1)$ are parallel, i.e., they are linearly dependent:

$$
\begin{aligned}
e(N+1) &= y(N+1) - E\big[\,y(N+1)|\,y^N\big] = \\
&= Cx(N+1) + v_2(N+1) - E\big[\,Cx(N+1) + v_2(N+1)|\,y^N\big] = \\
&= Cx(N+1) + v_2(N+1) - E\big[\,Cx(N+1)|\,y^N\big] - E\big[\,v_2(N+1)|\,y^N\big] = \\
&= Cx(N+1) + v_2(N+1) - C\,E\big[\,x(N+1)|\,y^N\big] - E[v_2(N+1)] = \\
&= Cx(N+1) - C\,E\big[\,x(N+1)|\,y^N\big] + v_2(N+1) = \\
&= C\,\big(x(N+1) - E\big[\,x(N+1)|\,y^N\big]\big) + v_2(N+1) = \\
&= C\tilde{x}(N+1) + v_2(N+1)
\end{aligned}
$$

where $E\big[v_2(N+1)|y^N\big] = E[v_2(N+1)] = \mathbf{0}$, since $v_2(t) \sim WN(\mathbf{0}, V_2)$, $\forall t$
$\Rightarrow v_2(N+1)$ is a random variable with zero mean value and independent of y^N

- The optimal estimate for the state $x(N+1)$ based on data y^N is given by:

$$\hat{x}(N+1|N) = E\big[\,x(N+1)|\,y^N\,\big] = E\big[\,x(N+1)|\,y^{N-1}, y(N)\big]$$

where $y^{N-1} = \big[y(N-1)^T\ y(N-2)^T\ \cdots\ y(1)^T\big]^T \in \mathbb{R}^{(N-1)q}$.

From the recursive Bayesian estimate formula, it results that:

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

where $e(N) = y(N) - E\big[\,y(N)|\,y^{N-1}\big]$ is the innovation of $y(N)$ given y^{N-1}

- From the state equation of the system $\mathcal{S}$:

$$\begin{aligned}E\big[\,x(N+1)|\,y^{N-1}\big] &= E\big[\,Ax(N) + v_1(N)|\,y^{N-1}\big] = \\ &= A\,E\big[\,x(N)|\,y^{N-1}\big] + E\big[\,v_1(N)|\,y^{N-1}\big] = \\ &= A\,E\big[\,x(N)|\,y^{N-1}\big] = A\hat{x}(N|N-1)\end{aligned}$$

since $v_1(N)$ is independent of $y^{N-1} \Rightarrow E\big[\,v_1(N)|\,y^{N-1}\big] = E[v_1(N)] = \mathbf{0}$

- Since $E[x(t)] = E[e(t)] = \mathbf{0}, \forall t$, then:

$$E[\,x(N+1)|\,e(N)] = \Sigma_{x(N+1)e(N)}\Sigma^{-1}_{e(N)e(N)}e(N) = K(N)e(N)$$

where it can be proved that:

$$\Sigma_{x(N+1)e(N)} = AP(N)C^T$$
$$\Sigma_{e(N)e(N)} = CP(N)C^T + V_2$$
$$K(N) = \Sigma_{x(N+1)e(N)}\Sigma^{-1}_{e(N)e(N)} = AP(N)C^T\left[CP(N)C^T + V_2\right]^{-1}$$
$$P(N) = Var[\tilde{x}(N)] = E\left[\tilde{x}(N)\tilde{x}(N)^T\right]$$

- The recursive form of the one-step state prediction is then:

$$\boxed{\hat{\mathbf{x}}(\mathbf{N}+\mathbf{1}|\mathbf{N}) = \mathbf{A}\hat{\mathbf{x}}(\mathbf{N}|\mathbf{N}-\mathbf{1}) + \mathbf{K}(\mathbf{N})\mathbf{e}(\mathbf{N})}$$

where $K(N) \in \mathbb{R}^{n\times q}$ is called **one-step Kalman predictor gain matrix** and involves the prediction error variance $P(N) \in \mathbb{R}^{n\times n}$ of the state $x(N)$

- The prediction error variance $P(N) \in \mathbb{R}^{n \times n}$ of the state $x(N)$ can be recursively computed using the **Difference Riccati Equation (DRE)**, which can be written in one of the following equivalent forms:

$$P(N+1) = [A - K(N)C]\, P(N)\, [A - K(N)C]^T + V_1 + K(N)V_2K(N)^T$$
$$P(N+1) = AP(N)A^T + V_1 - K(N)\left[CP(N)C^T + V_2\right] K(N)^T$$

assuming $P(1) = Var[x(1)] = P_1$ as starting value

- The update equation of the state prediction

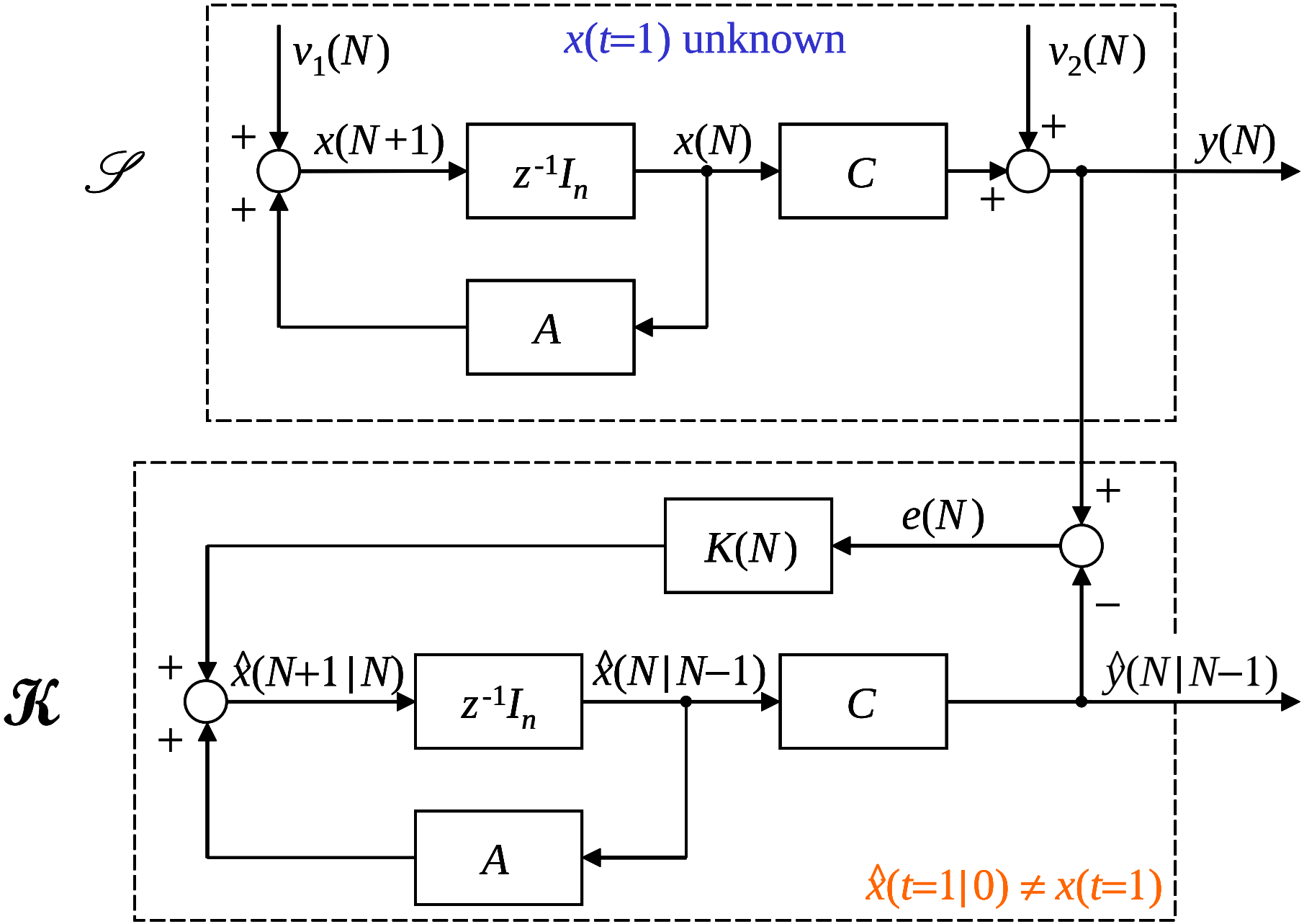
$$\hat{x}(N+1|N) = A\hat{x}(N|N-1) + K(N)e(N)$$

has to be initialized by assuming as starting value:

$$\hat{x}(1|0) = E[x(1)] = \mathbf{0}$$

- The optimal estimate for the output $y(N+1)$ based on data y^N is given by:

$$\begin{aligned}\hat{y}(N+1|N) &= E\left[y(N+1)|\, y^N\right] = E\left[Cx(N+1) + v_2(N+1)|\, y^N\right] = \\ &= E\left[Cx(N+1)|\, y^N\right] + E\left[v_2(N+1)|\, y^N\right] = \\ &= CE\left[x(N+1)|\, y^N\right] + E[v_2(N+1)] = C\hat{x}(N+1|N)\end{aligned}$$


$x(t{=}1)$ unknown
$v_1(N)$
$v_2(N)$
$\mathscr{S}$
$x(N+1)$
$z^{-1}I_n$
$x(N)$
C
$y(N)$
A
$e(N)$
$K(N)$
$\mathscr{K}$
$\hat{x}(N{+}1|N)$
$z^{-1}I_n$
$\hat{x}(N|N{-}1)$
C
$\hat{y}(N|N{-}1)$
A
$\hat{x}(t{=}1|0) \neq x(t{=}1)$

Remarks

Remark #1: the gain matrix $K(N)$ is not constant, since it involves the prediction error variance $P(N)$ of the state that is given by the Difference Riccati equation $\Rightarrow$ even if the dynamical system $\mathcal{S}$ is LTI, the one-step Kalman predictor is **time-variant**

Remark #2: under all the above assumptions, the one-step Kalman predictor is the **optimal linear predictor**. In fact, the Bayesian estimate of $x(N+1)$ given y^N:

$$\hat{x}(N+1|N) = E\left[x(N+1)|\, y^N\right] = \Sigma_{x(N+1)y^N}\Sigma^{-1}_{y^N y^N} y^N$$

is a linear function of y^N and it minimizes the state prediction error variance

Remark #3: under the further assumptions that:

- the noises $v_1(t)$ and $v_2(t)$ are individually and jointly Gaussian, $\forall t$
- the initial state $x(t=1)$ is Gaussian

the state $x(N+1)$ and the data vector y^N are individually and jointly Gaussian, since $\mathcal{S}$ is linear $\Rightarrow$ the Bayesian estimate of $x(N+1)$ given y^N is optimal $\Rightarrow$ the one-step Kalman predictor is the **optimal predictor** in absolute terms

Generalization

Let us consider a discrete-time, linear time-variant, dynamical system $\mathcal{S}$ with an *exogenous* (deterministic and known) input $u(\cdot)$, described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = A(t)x(t) + B(t)u(t) + v_1(t) \\ y(t) = C(t)x(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $u(t) \in \mathbb{R}^p$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- $v_1(t)$ and $v_2(t)$ are white noises with zero mean value and known variance, i.e., $v_1(t) \sim WN(\mathbf{0}, V_1(t))$ and $v_2(t) \sim WN(\mathbf{0}, V_2(t))$, which may be correlated if considered at the same time instant but are uncorrelated at different time instants:

$$E[v_1(t_1)v_1(t_2)^T] = V_1(t_1)\delta(t_2 - t_1)$$
$$E[v_2(t_1)v_2(t_2)^T] = V_2(t_1)\delta(t_2 - t_1)$$
$$E[v_1(t_1)v_2(t_2)^T] = V_{12}(t_1)\delta(t_2 - t_1)$$

- $A(t) \in \mathbb{R}^{n \times n}$, $B(t) \in \mathbb{R}^{n \times p}$, $C(t) \in \mathbb{R}^{q \times n}$, $V_1(t) \in \mathbb{R}^{n \times n}$, $V_2(t) \in \mathbb{R}^{q \times q}$, $V_{12}(t) \in \mathbb{R}^{n \times q}$ are known matrices $\forall t$

- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\bar{x}_1, P_1)$, with known $\bar{x}_1 \in \mathbb{R}^n$ and $P_1 \in \mathbb{R}^{n \times n}$, which is uncorrelated with $v_1(t)$ and $v_2(t)$
- the output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$
- the one-step Kalman predictor $\mathcal{K}$ is described by the state space model:

$$\mathcal{K} : \begin{cases} \hat{x}(N+1|N) = A(N)\hat{x}(N|N-1) + B(N)u(N) + K(N)e(N) \\ \hat{y}(N|N-1) = C(N)\hat{x}(N|N-1) \\ e(N) = y(N) - \hat{y}(N|N-1) \end{cases}$$

- the one-step Kalman predictor gain matrix $K(N)$ is equal to:

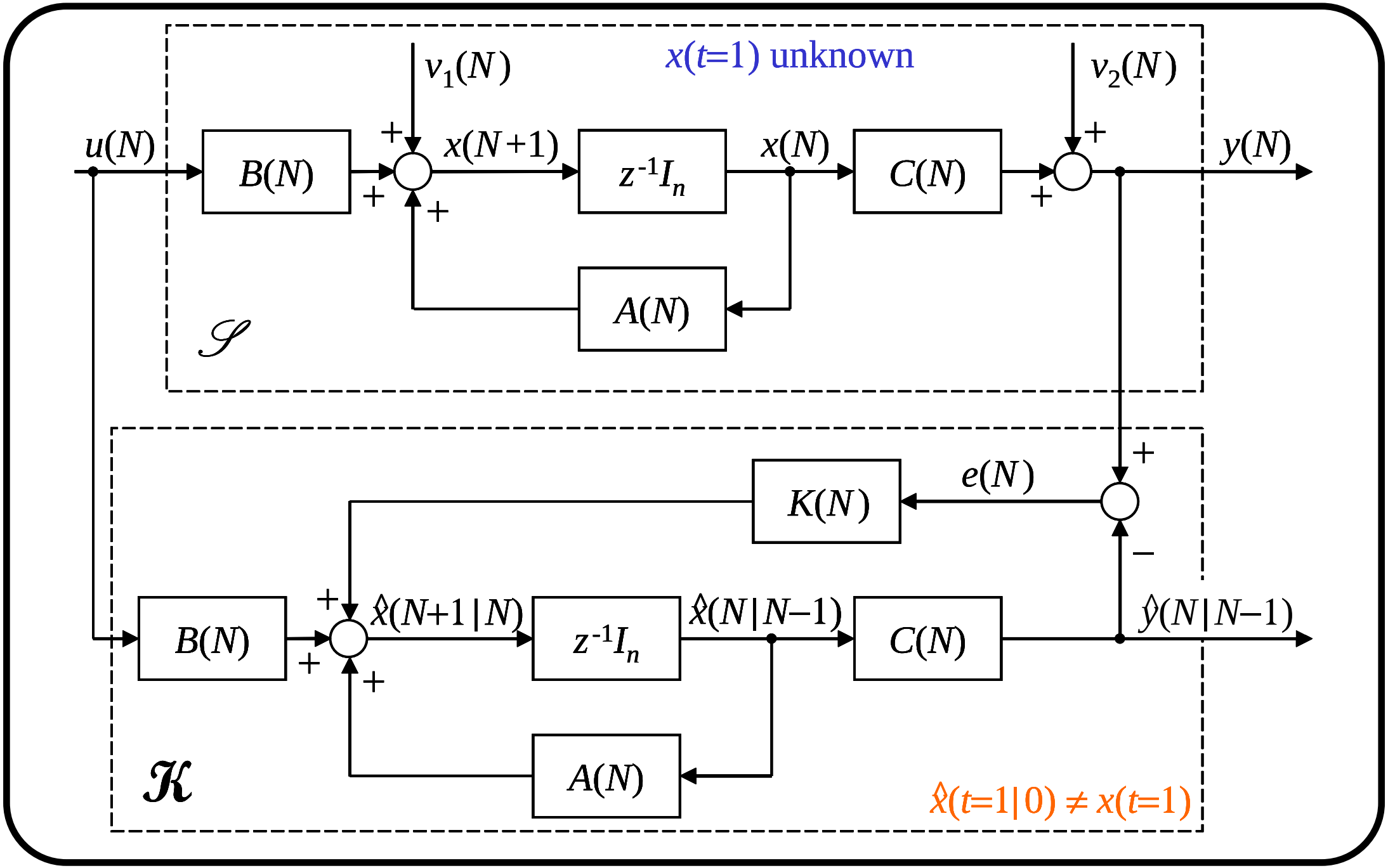
$$K(N) = \left[A(N)P(N)C(N)^T + V_{12}(N)\right]\left[C(N)P(N)C(N)^T + V_2(N)\right]^{-1}$$

where the state prediction error variance $P(N)$ is given by the following DRE:

$$\begin{aligned} P(N+1) = \; & A(N)P(N)A(N)^T + V_1(N) + \\ & -K(N)\left[C(N)P(N)C(N)^T + V_2(N)\right]K(N)^T \end{aligned}$$

- the initialization is:

$$\begin{aligned} \hat{x}(1|0) &= E[x(1)] = \bar{x}_1 \\ P(1) &= E[(x(1) - \bar{x}_1)(x(1) - \bar{x}_1)^T] = P_1 \end{aligned}$$


x(t=1) unknown
v1(N)
v2(N)
u(N)
B(N)
x(N+1)
z-1In
x(N)
C(N)
y(N)
A(N)
S
K(N)
e(N)
B(N)
x̂(N+1|N)
z-1In
x̂(N|N−1)
C(N)
ŷ(N|N−1)
A(N)
K
x̂(t=1|0) ≠ x(t=1)

Multi-step Kalman predictor

Let us consider a discrete-time, linear time-invariant (LTI), dynamical system $\mathcal{S}$ without any exogenous input, described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = Ax(t) + v_1(t) \\ y(t) = Cx(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- $v_1(t)$ and $v_2(t)$ are white noises with zero mean value that may be correlated if considered at the same time instant but are uncorrelated at different time instants:
$$E[v_i(t_1)v_j(t_2)^T] = V_{ij}\delta(t_2 - t_1), \quad i = 1, 2; \quad j = 1, 2;$$
- $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{q \times n}$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{q \times q}$, $V_{12} \in \mathbb{R}^{n \times q}$ are known matrices
- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\bar{x}_1, P_1)$, with known $\bar{x}_1 \in \mathbb{R}^n$ and $P_1 \in \mathbb{R}^{n \times n}$, which is uncorrelated with $v_1(t)$ and $v_2(t)$
- the output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$
- **Goal:** estimate the state $x(N + r)$, $r > 1$

- The optimal estimate for the state $x(N+r)$ based on the N measurements collected in the vector $y^N = \left[y(N)^T\, y(N-1)^T \cdots y(1)^T\right]^T \in \mathbb{R}^{Nq}$ is given by:

$$\begin{aligned}\hat{x}(N+r|N) &= E\left[x(N+r)|\, y^N\right] = E\left[Ax(N+r-1) + v_1(N+r-1)|\, y^N\right] \\ &= E\left[Ax(N+r-1)|\, y^N\right] + E\left[v_1(N+r-1)|\, y^N\right] = \\ &= AE\left[x(N+r-1)|\, y^N\right] + E\left[v_1(N+r-1)\right] = A\hat{x}(N+r-1|N)\end{aligned}$$

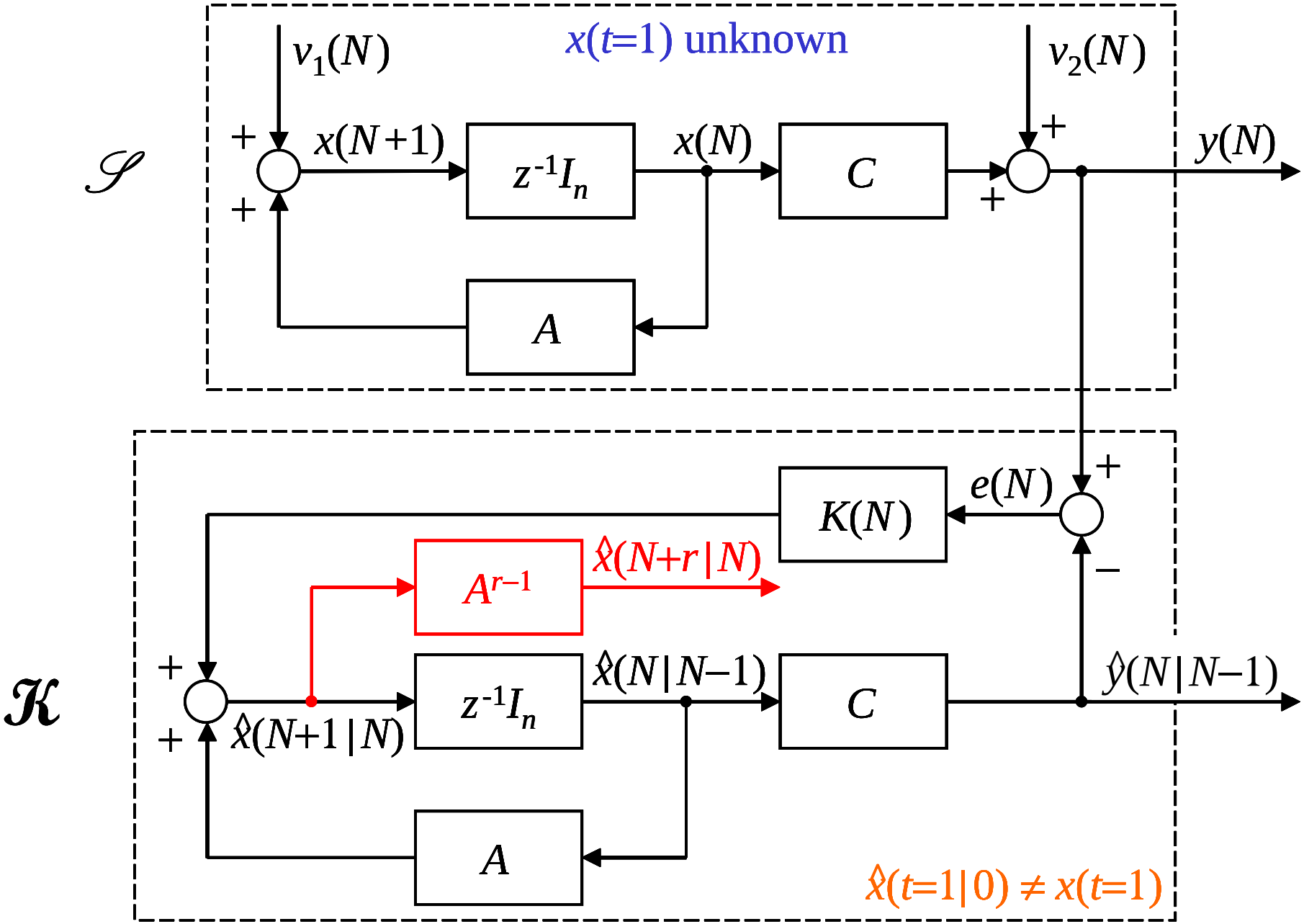
since $v_1(N+r-1)$ is independent of y^N, $\forall r > 1 \Rightarrow$
$E\left[v_1(N+r-1)|\, y^N\right] = E[v_1(N+r-1)] = \mathbf{0} \Rightarrow$ by iterating, it results that:

$$\boxed{\hat{\mathbf{x}}(\mathbf{N}+\mathbf{r}|\mathbf{N}) = \mathbf{A}^{r-1}\hat{\mathbf{x}}(\mathbf{N}+\mathbf{1}|\mathbf{N})}$$

- The optimal estimate for the output $y(N+r)$ based on data y^N is given by:

$$\begin{aligned}\hat{y}(N+r|N) &= E\left[y(N+r)|\, y^N\right] = E\left[Cx(N+r) + v_2(N+r)|\, y^N\right] = \\ &= E\left[Cx(N+r)|\, y^N\right] + E\left[v_2(N+r)|\, y^N\right] = \\ &= CE\left[x(N+r)|\, y^N\right] + E[v_2(N+r)] = \\ &= C\hat{x}(N+r|N) = CA^{r-1}\hat{x}(N+1|N)\end{aligned}$$

since $v_2(N+r)$ is independent of $y^N \Rightarrow E\left[v_2(N+r)|\, y^N\right] = E[v_2(N+r)] = \mathbf{0}$


$x(t{=}1)$ unknown
$v_1(N)$
$v_2(N)$
$\mathscr{S}$
$x(N+1)$
$z^{-1}I_n$
$x(N)$
C
$y(N)$
A
$e(N)$
$K(N)$
A^{r-1}
$\hat{x}(N{+}r\,|\,N)$
$\mathscr{K}$
$\hat{x}(N|N{-}1)$
$\hat{y}(N|N{-}1)$
$\hat{x}(N{+}1\,|\,N)$
$\hat{x}(t{=}1|0) \neq x(t{=}1)$

Generalization

Let us consider a discrete-time, linear time-invariant (LTI), dynamical system $\mathcal{S}$ with an *exogenous* (deterministic and known) input $u(\cdot)$, described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = Ax(t) + Bu(t) + v_1(t) \\ y(t) = Cx(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $u(t) \in \mathbb{R}^p$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- $u(\cdot)$ may possibly depend on the output $y(\cdot)$ through a causal feedback as
$$u(t) = f(y(t), y(t-1), y(t-2), \ldots), \quad \forall t$$
- $v_1(t)$ and $v_2(t)$ are white noises with zero mean value that may be correlated if considered at the same time instant but are uncorrelated at different time instants:
$$E[v_i(t_1)v_j(t_2)^T] = V_{ij}\delta(t_2 - t_1), \quad i = 1, 2; \quad j = 1, 2;$$
- $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{q \times n}$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{q \times q}$, $V_{12} \in \mathbb{R}^{n \times q}$ are known
- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\bar{x}_1, P_1)$, with known $\bar{x}_1 \in \mathbb{R}^n$ and $P_1 \in \mathbb{R}^{n \times n}$, which is uncorrelated with $v_1(t)$ and $v_2(t)$
- the output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$

- The optimal estimate for $x(N+r)$, $r>1$, based on the data vector y^N is given by:

$$\begin{aligned}\hat{x}(N+r|N) &= E\left[x(N+r)|\,y^N\right] = \\ &= E\left[Ax(N+r-1) + Bu(N+r-1) + v_1(N+r-1)|\,y^N\right] = \\ &= E\left[Ax(N+r-1)|\,y^N\right] + E\left[Bu(N+r-1)|\,y^N\right] + E\left[v_1(N+r-1)|\,y^N\right] \\ &= AE\left[x(N+r-1)|\,y^N\right] + BE\left[u(N+r-1)|\,y^N\right] + E[v_1(N+r-1)] = \\ &= A\hat{x}(N+r-1|N) + B\hat{u}(N+r-1|N)\end{aligned}$$

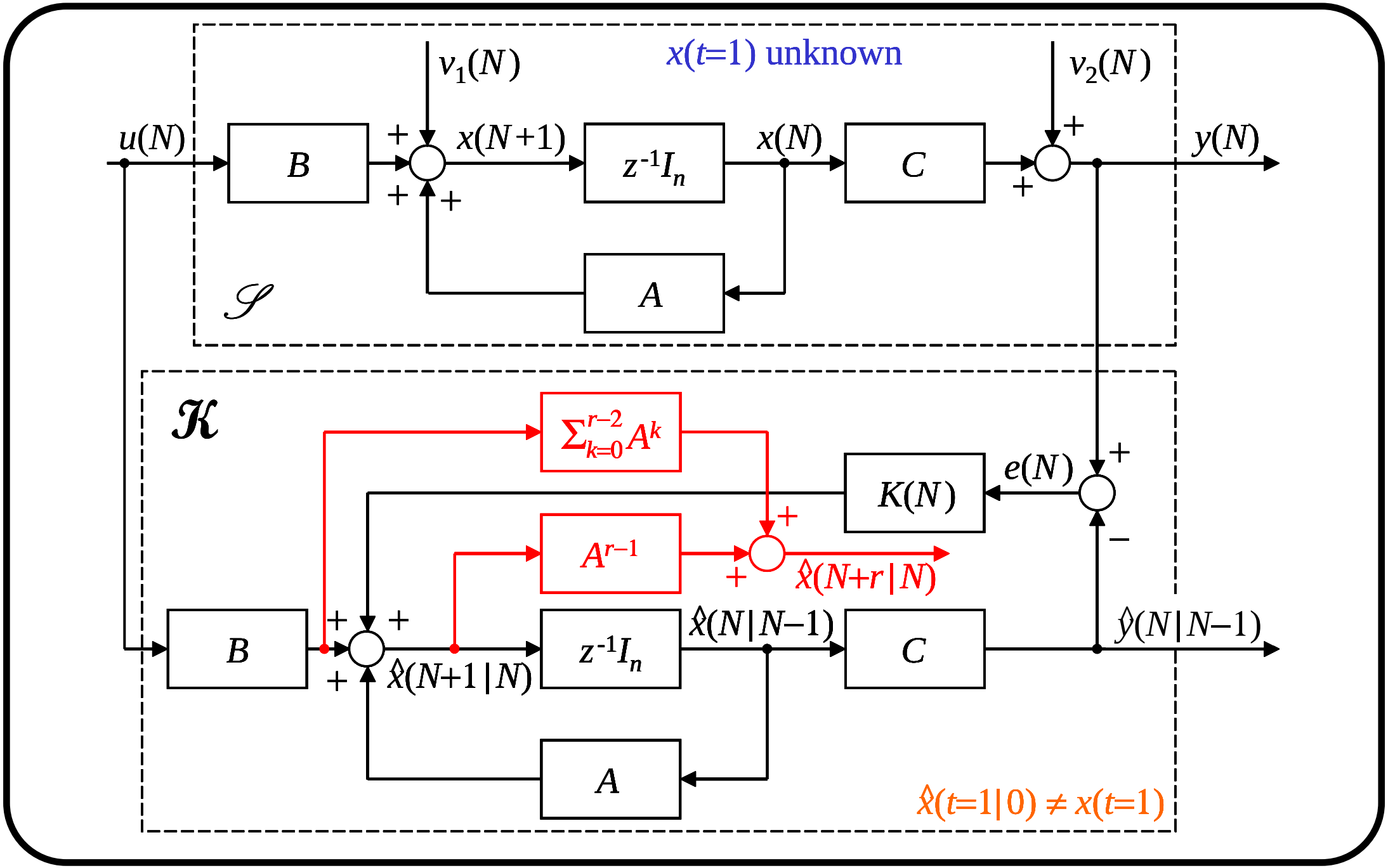
($v_1(N+r-1)$ is independent of $y^N \Rightarrow E\left[v_1(N+r-1)|\,y^N\right] = E[v_1(N+r-1)] = \mathbf{0}$)

$$\Downarrow$$

by iterating, it results that:

$$\hat{x}(N+r|N) = A^{r-1}\hat{x}(N+1|N) + \left[B\ AB\ \cdots\ A^{r-2}B\right]\begin{bmatrix}\hat{u}(N+r-1|N)\\ \hat{u}(N+r-2|N)\\ \vdots \\ \hat{u}(N+1|N)\end{bmatrix}$$

where $u(\cdot)$ is predicted up to $r-1$ steps ahead to have $\hat{u}(N+i|N)$ or, if it cannot be predicted, then $\hat{u}(N+i|N) = u(N)$ is often assumed, for $1 \le i \le r-1$


x(t=1) unknown
v1(N)
v2(N)
u(N)
B
x(N+1)
z-1In
x(N)
C
y(N)
A
𝒮
𝒦
Σk=0r−2 Ak
K(N)
e(N)
Ar−1
x̂(N+r|N)
B
x̂(N+1|N)
z-1In
x̂(N|N−1)
C
ŷ(N|N−1)
A
x̂(t=1|0) ≠ x(t=1)

Kalman filter

Let us consider a discrete-time, linear time-invariant (LTI), dynamical system $\mathcal{S}$ without any exogenous input, described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = Ax(t) + v_1(t) \\ y(t) = Cx(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- $v_1(t)$ and $v_2(t)$ are white noises with zero mean value that may be correlated if considered at the same time instant but are uncorrelated at different time instants:
$$E[v_i(t_1)v_j(t_2)^T] = V_{ij}\delta(t_2 - t_1), \quad i = 1, 2; \quad j = 1, 2;$$
- $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{q \times n}$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{q \times q}$, $V_{12} \in \mathbb{R}^{n \times q}$ are known matrices
- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\mathbf{0}, P_1)$, with known $P_1 \in \mathbb{R}^{n \times n}$, which is uncorrelated with $v_1(t)$ and $v_2(t)$
- the output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$
- **Goal:** estimate the state $x(N)$

- The optimal estimate for the state $x(N)$ based on data y^N is given by:

$$\hat{x}(N|N) = E\left[x(N)|\, y^N\right] = E\left[x(N)|\, y^{N-1}, y(N)\right]$$

where $y^{N-1} = \left[y(N-1)^T\ y(N-2)^T\ \cdots\ y(1)^T\right]^T \in \mathbb{R}^{(N-1)q}$.

From the recursive Bayesian estimate formula, it results that:

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

where $e(N) = y(N) - \hat{y}(N|N-1)$ is the innovation of $y(N)$ given y^{N-1}

- Since $E[x(t)] = E[e(t)] = \mathbf{0}, \forall t$, then:

$$E[x(N)|\, e(N)] = \Sigma_{x(N)e(N)}\Sigma^{-1}_{e(N)e(N)}e(N) = K_0(N)e(N)$$

where it can be proved that:

$$\begin{aligned}\Sigma_{x(N)e(N)} &= P(N)C^T \\ \Sigma_{e(N)e(N)} &= CP(N)C^T + V_2 \\ K_0(N) &= \Sigma_{x(N)e(N)}\Sigma^{-1}_{e(N)e(N)} = P(N)C^T\left[CP(N)C^T + V_2\right]^{-1} \\ P(N) &= Var[\tilde{x}(N)] = E\left[\tilde{x}(N)\tilde{x}(N)^T\right]\end{aligned}$$

- The optimal estimate for the state $x(N)$ based on data y^N can be obtained from the one-step Kalman predictor as:

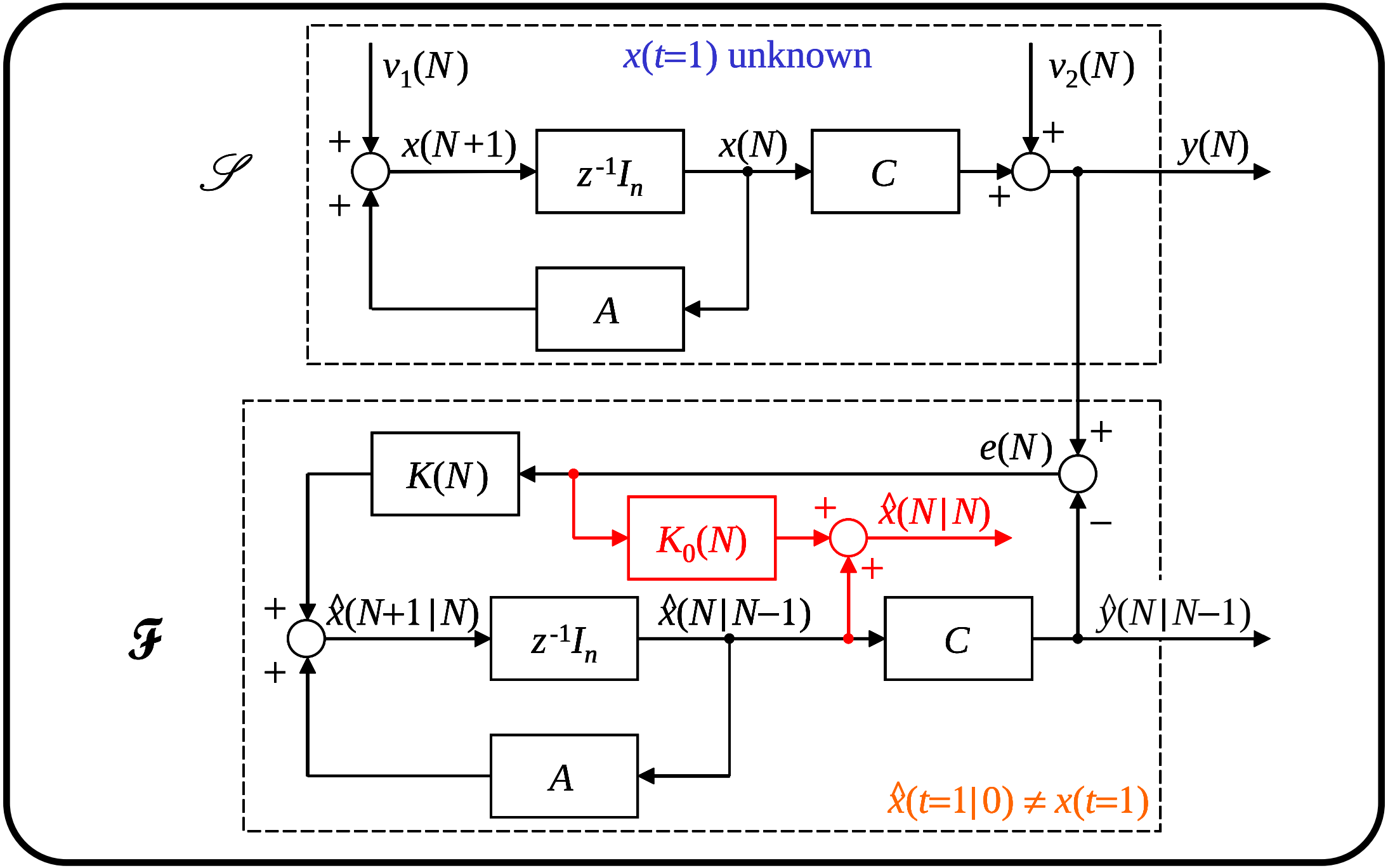
$$\boxed{\hat{\mathbf{x}}(\mathbf{N}|\mathbf{N}) = \hat{\mathbf{x}}(\mathbf{N}|\mathbf{N}-\mathbf{1}) + \mathbf{K}_0(\mathbf{N})\mathbf{e}(\mathbf{N})}$$

  where $K_0(N) \in \mathbb{R}^{n\times q}$ is called **Kalman filter gain matrix** and involves the variance $P(N) \in \mathbb{R}^{n\times n}$ of the one-step prediction error of the state $x(N)$, i.e., $\tilde{x}(N) = x(N) - \hat{x}(N|N-1)$. Note that, if $V_{12} = \mathbf{0}_{n\times q}$, then $K(N) = AK_0(N)$

- The variance of the filtering error $x(N) - \hat{x}(N|N)$ involves the variance $P(N)$ of the one-step prediction error as well, since it can be proved that:

$$Var[x(N) - \hat{x}(N|N)] = P(N) - P(N)C^T[CP(N)C^T + V_2]^{-1}CP(N)$$
$$\Downarrow$$
$$Var[x(N) - \hat{x}(N|N)] \leq P(N)$$

  since the estimate $\hat{x}(N|N)$ provided by the Kalman filter is based also on the data sample $y(N)$ with respect to the estimate $\hat{x}(N|N-1)$ provided by the one-step Kalman predictor $\Rightarrow$ the uncertainty on $\hat{x}(N|N)$ has to be lower


$x(t{=}1)$ unknown
$v_1(N)$
$v_2(N)$
$\mathscr{S}$
$x(N+1)$
$z^{-1}I_n$
$x(N)$
C
$y(N)$
A
$e(N)$
$K(N)$
$K_0(N)$
$\hat{x}(N|N)$
$\mathscr{F}$
$\hat{x}(N+1|N)$
$z^{-1}I_n$
$\hat{x}(N|N-1)$
C
$\hat{y}(N|N-1)$
A
$\hat{x}(t{=}1|0) \neq x(t{=}1)$

Generalization

Let us consider a discrete-time, linear time-invariant (LTI), dynamical system $\mathcal{S}$ with an *exogenous* (deterministic and known) input $u(\cdot)$, described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = Ax(t) + Bu(t) + v_1(t) \\ y(t) = Cx(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $u(t) \in \mathbb{R}^p$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- $u(\cdot)$ may possibly depend on the output $y(\cdot)$ through a causal feedback as
$$u(t) = f(y(t), y(t-1), y(t-2), \ldots), \quad \forall t$$
- $v_1(t)$ and $v_2(t)$ are white noises with zero mean value that may be correlated if considered at the same time instant but are uncorrelated at different time instants:
$$E[v_i(t_1)v_j(t_2)^T] = V_{ij}\delta(t_2 - t_1), \quad i = 1, 2; \quad j = 1, 2;$$
- $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{q \times n}$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{q \times q}$, $V_{12} \in \mathbb{R}^{n \times q}$ are known
- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\bar{x}_1, P_1)$, with known $\bar{x}_1 \in \mathbb{R}^n$ and $P_1 \in \mathbb{R}^{n \times n}$, which is uncorrelated with $v_1(t)$ and $v_2(t)$
- the output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$

- The optimal estimate for the state $x(N)$ based on the data vector y^N can still be obtained from the one-step Kalman predictor as:

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

where $K_0(N) = P(N)C^T\left[CP(N)C^T + V_2\right]^{-1} \in \mathbb{R}^{n\times q}$ is the Kalman filter gain matrix involving the state prediction error variance $P(N) \in \mathbb{R}^{n\times n}$ given by:

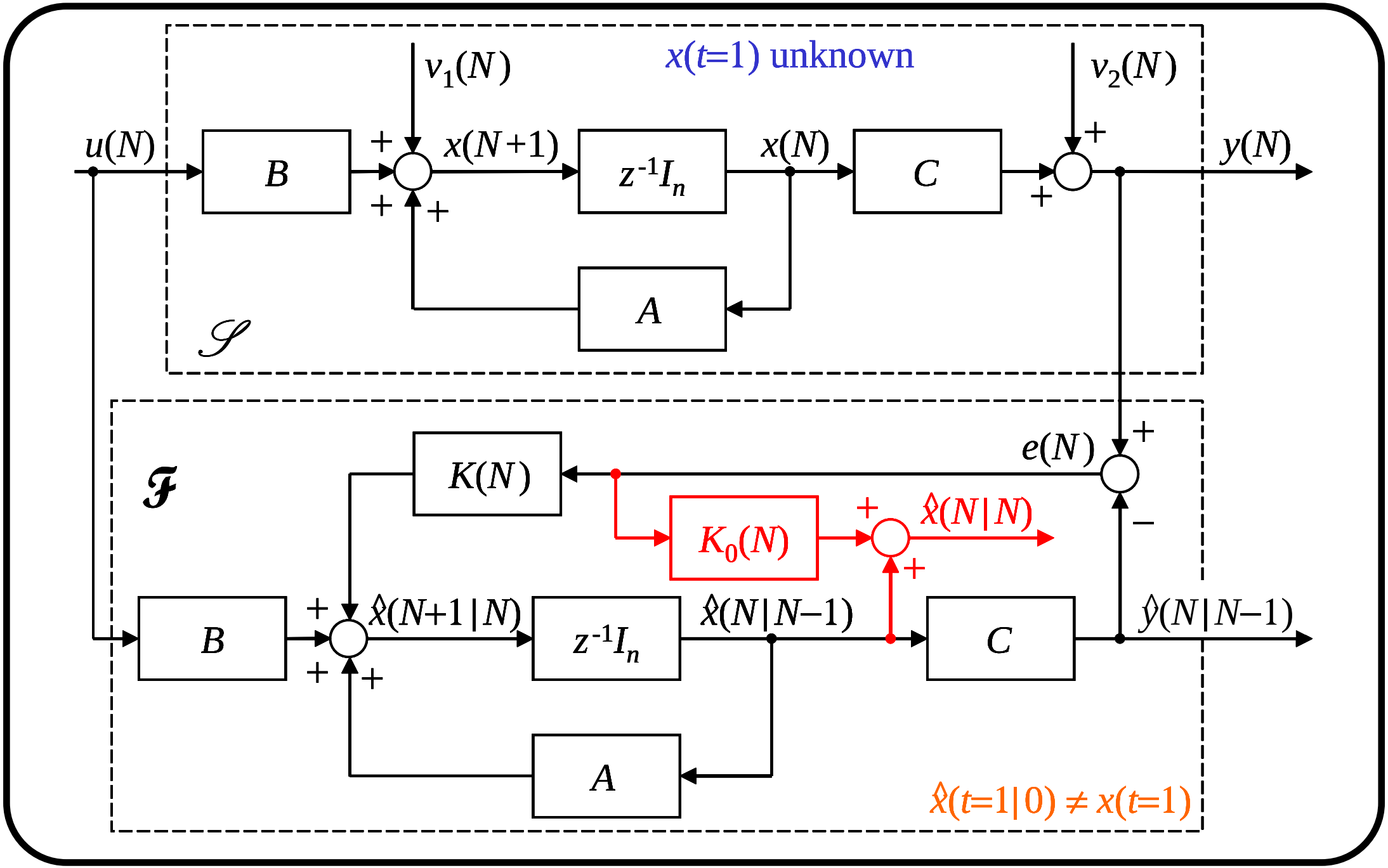
$$P(N+1) = AP(N)A^T + V_1 - K(N)\left[CP(N)C^T + V_2\right]K(N)^T$$

with the one-step Kalman predictor gain matrix $K(N) \in \mathbb{R}^{n\times q}$ equal to:

$$K(N) = \left[AP(N)C^T + V_{12}\right]\left[CP(N)C^T + V_2\right]^{-1}$$

- The overall equations of the Kalman filter are:

$$\mathcal{F}: \begin{cases} \hat{x}(N+1|N) = A\hat{x}(N|N-1) + Bu(N) + K(N)e(N) \\ \hat{y}(N|N-1) = C\hat{x}(N|N-1) \\ \hat{x}(N|N) = \hat{x}(N|N-1) + K_0(N)e(N) \\ e(N) = y(N) - \hat{y}(N|N-1) \end{cases}$$


x(t=1) unknown
v1(N)
v2(N)
u(N)
B
x(N+1)
z-1In
x(N)
C
y(N)
A
𝒮
ℱ
K(N)
e(N)
K0(N)
x̂(N|N)
B
x̂(N+1|N)
z-1In
x̂(N|N−1)
C
ŷ(N|N−1)
A
x̂(t=1|0) ≠ x(t=1)

Predictor/corrector form of one-step Kalman predictor

The so-called predictor/corrector form provides a numerically reliable formulation of the equations involved in the one-step Kalman predictor, where the transition from $\hat{x}(N|N-1)$ to $\hat{x}(N+1|N)$ is performed in two steps:

- first, the filtered estimate $\hat{x}(N|N)$ is derived from $\hat{x}(N|N-1)$ making use of the Kalman filter gain matrix $K_0(N)$
- then, the one-step prediction $\hat{x}(N+1|N)$ is derived by updating $\hat{x}(N|N)$

In fact:

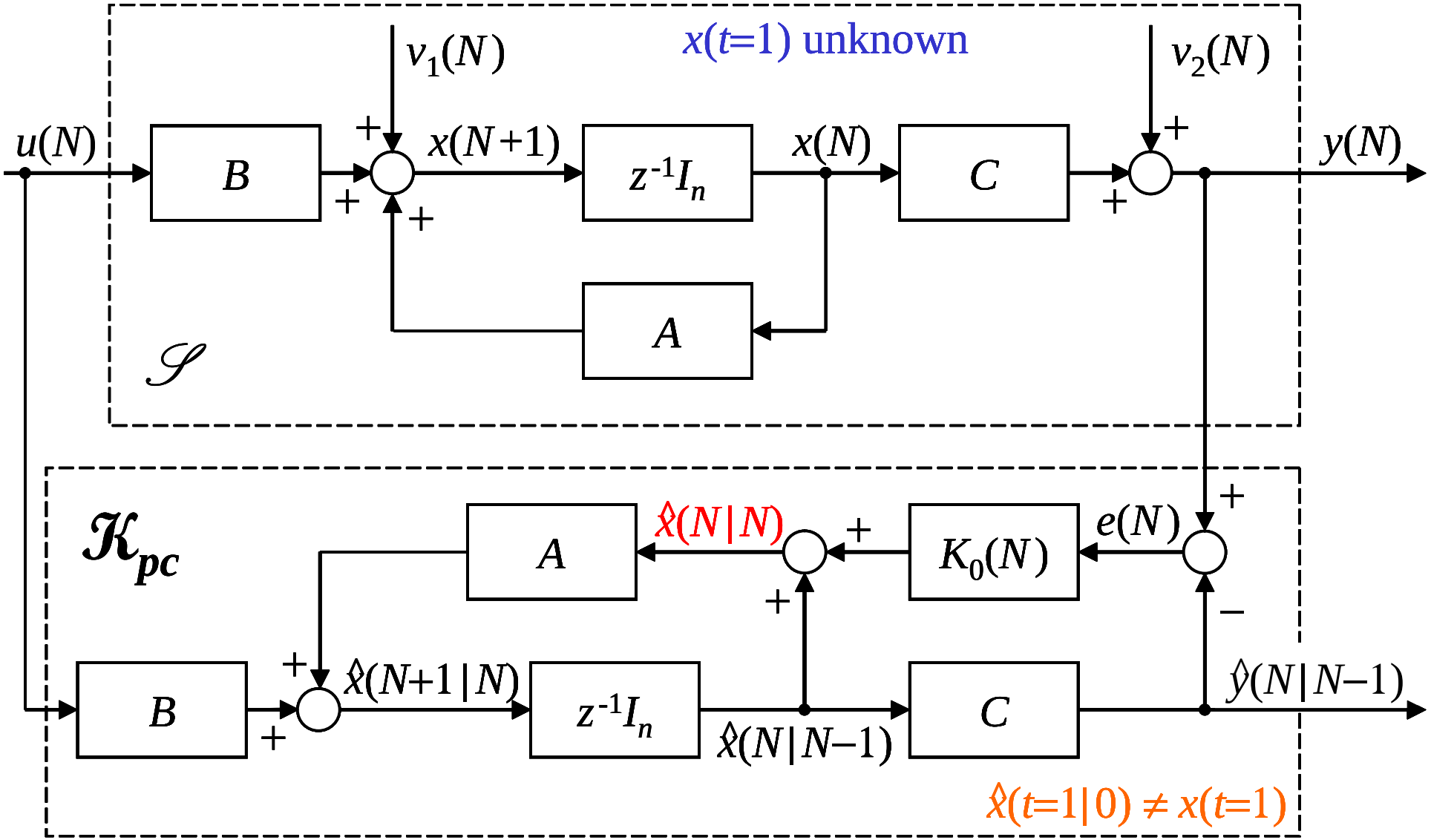
$$\begin{aligned}
\hat{x}(N+1|N) &= A\hat{x}(N|N-1)+Bu(N)+K(N)e(N) = \\
&= A\hat{x}(N|N-1)+Bu(N)+\left[AP(N)C^T+V_{12}\right]\left[CP(N)C^T+V_2\right]^{-1}e(N) = \\
&= A\hat{x}(N|N-1)+Bu(N)+AP(N)C^T\left[CP(N)C^T+V_2\right]^{-1}e(N)+ \\
&\qquad +V_{12}\left[CP(N)C^T+V_2\right]^{-1}e(N) = \\
&= A\hat{x}(N|N-1)+AP(N)C^T\left[CP(N)C^T+V_2\right]^{-1}e(N)+Bu(N)+ \\
&\qquad +V_{12}\left[CP(N)C^T+V_2\right]^{-1}e(N) = \\
&= A\left\{\hat{x}(N|N-1)+P(N)C^T\left[CP(N)C^T+V_2\right]^{-1}e(N)\right\}+Bu(N)+ \\
&\qquad +V_{12}\left[CP(N)C^T+V_2\right]^{-1}e(N) = \\
&= A\left[\hat{x}(N|N-1)+K_0(N)e(N)\right]+Bu(N)+V_{12}\left[CP(N)C^T+V_2\right]^{-1}e(N) = \\
&= A\hat{x}(N|N)+Bu(N)+V_{12}\left[CP(N)C^T+V_2\right]^{-1}e(N)
\end{aligned}$$

Note that, in the case $V_{12} = \mathbf{0}_{n\times q}$ and $V_2 > 0$, where $K(N) = AK_0(N)$:

$$\hat{x}(N+1|N) = A\left[\hat{x}(N|N-1) + K_0(N)e(N)\right] + Bu(N) =$$
$$= A\hat{x}(N|N) + Bu(N)$$

and then the equations are:

$$K_0(N) = P(N)C^T\left[CP(N)C^T + V_2\right]^{-1} \qquad \text{(filter gain)}$$
$$P_0(N) = \left[I_n - K_0(N)C\right]P(N)\left[I_n - K_0(N)C\right]^T + K_0(N)V_2K_0(N)^T$$
$$= \left[I_n - K_0(N)C\right]P(N) \qquad \text{(measurement update)}$$
$$e(N) = y(N) - C\hat{x}(N|N-1) \qquad \text{(innovation)}$$
$$\hat{x}(N|N) = \hat{x}(N|N-1) + K_0(N)e(N) \qquad \text{(corrector)}$$
$$P(N+1) = AP_0(N)A^T + V_1 \qquad \text{(time update)}$$
$$\hat{x}(N+1|N) = A\hat{x}(N|N) + Bu(N) \qquad \text{(predictor)}$$


x(t=1) unknown
v1(N)
v2(N)
u(N)
B
x(N+1)
z-1In
x(N)
C
y(N)
A
𝒮
𝒦pc
A
x̂(N|N)
K0(N)
e(N)
B
x̂(N+1|N)
z-1In
x̂(N|N−1)
C
ŷ(N|N−1)
x̂(t=1|0) ≠ x(t=1)

Steady-state Kalman predictor

Let us consider a discrete-time, linear time-invariant (LTI), dynamical system $\mathcal{S}$ without any exogenous input, described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = Ax(t) + v_1(t) \\ y(t) = Cx(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- $v_1(t)$ and $v_2(t)$ are white noises with zero mean value that are uncorrelated with each other, i.e., $v_1(t) \sim WN(\mathbf{0}, V_1)$ and $v_2(t) \sim WN(\mathbf{0}, V_2)$ are such that:

$$E[v_1(t_1)v_1(t_2)^T] = V_1\delta(t_2 - t_1) \qquad \text{(whiteness of } v_1(t)\text{)}$$
$$E[v_2(t_1)v_2(t_2)^T] = V_2\delta(t_2 - t_1) \qquad \text{(whiteness of } v_2(t)\text{)}$$
$$E[v_1(t_1)v_2(t_2)^T] = \mathbf{0}_{n\times q}, \quad \forall t_1, t_2 \qquad \text{(uncorrelation of } v_1(t) \text{ and } v_2(t)\text{)}$$

- $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{q\times n}$, $V_1 = B_v B_v^T \in \mathbb{R}^{n\times n}$, $V_2 \in \mathbb{R}^{q\times q}$ are known matrices
- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\bar{x}_1, P_1)$, with known $\bar{x}_1 \in \mathbb{R}^n$ and $P_1 \in \mathbb{R}^{n\times n}$, which is uncorrelated with $v_1(t)$ and $v_2(t)$

- The output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$
- **Goal:** estimate the state $x(N+1)$ using a linear **time-invariant** predictor
- Question: under which conditions the Kalman predictor gain $K(N)$ converges to a constant matrix $\bar{K}$ called **steady-state gain**, i.e., $\lim_{N\to\infty} K(N) = \bar{K}$? Since $K(N)$ involves the variance $P(N)$ computed using the Difference Riccati Equation (DRE), under which conditions $P(N)$ converges to a constant matrix $\bar{P}$? In this case, the DRE becomes static or algebraic and may be written as:

$$P = APA^T + V_1 - APC^T(CPC^T + V_2)^{-1}CPA^T$$

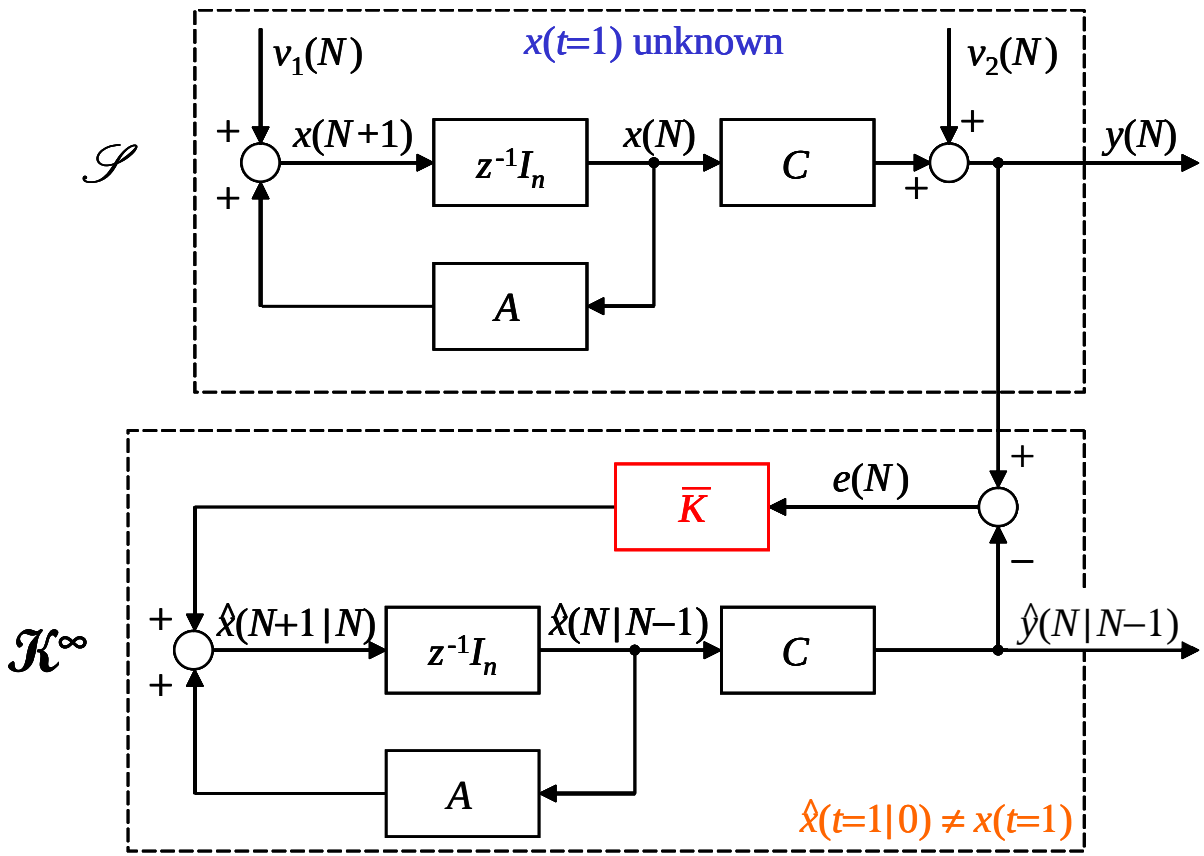
known as **Algebraic Riccati Equation (ARE)** , whose solution is $\bar{P}$

- The corresponding **steady-state Kalman predictor** $\mathcal{K}^\infty$ is described by the model:

$$\mathcal{K}^\infty : \begin{cases} \hat{x}(N+1|N) = A\hat{x}(N|N-1) + \bar{K}e(N) \\ \hat{y}(N|N-1) = C\hat{x}(N|N-1) \\ e(N) = y(N) - \hat{y}(N|N-1) \end{cases}$$

where the steady-state gain matrix $\bar{K} \in \mathbb{R}^{n\times q}$ is given by:

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



- The predictor $\mathcal{K}^\infty$ is a LTI dynamical system with input $y(\cdot)$ and state equation:
$$\hat{x}(N+1|N) = A\hat{x}(N|N-1) + \bar{K}e(N) = A\hat{x}(N|N-1) + \bar{K}[y(N) - \hat{y}(N|N-1)]$$
$$= A\hat{x}(N|N-1) + \bar{K}[y(N) - C\hat{x}(N|N-1)] = (A - \bar{K}C)\hat{x}(N|N-1) + \bar{K}y(N)$$
the internal asymptotic stability of $\mathcal{K}^\infty$ depends on the eigenvalues of $A - \bar{K}C$:
$\mathcal{K}^\infty$ is asymptotically stable if and only if $\left|\lambda_i(A - \bar{K}C)\right| < 1, \forall i$

The asymptotic stability of $\mathcal{K}^\infty$ preserves the predictor performances from disturbances and uncertainties, as well as it guarantees the boundedness of the state prediction error

In fact, in absence of noise (i.e., assuming $v_1(t) = \mathbf{0}$ and $v_2(t) = \mathbf{0},\ \forall t \geq 1$), the prediction error of the state $\tilde{x}(N+1)$ is given by:

$$\begin{aligned}
\tilde{x}(N+1) &= x(N+1) - \hat{x}(N+1|N) = \\
&= Ax(N) + v_1(N) - \left[(A - \bar{K}C)\hat{x}(N|N-1) + \bar{K}y(N)\right] = \\
&= Ax(N) - (A - \bar{K}C)\hat{x}(N|N-1) - \bar{K}\left[Cx(N) + v_2(N)\right] = \\
&= (A - \bar{K}C)x(N) - (A - \bar{K}C)\hat{x}(N|N-1) = \\
&= (A - \bar{K}C)\left[x(N) - \hat{x}(N|N-1)\right] = (A - \bar{K}C)\tilde{x}(N)
\end{aligned}$$

$$\Downarrow$$

$$\tilde{x}(t) = \left(A - \bar{K}C\right)^{t-1}\tilde{x}(t=1) = \left(A - \bar{K}C\right)^{t-1}\left[x(t=1) - \hat{x}(t=1|0)\right], \forall t \geq 1$$

If $A - \bar{K}C$ is such that $\left|\lambda_i(A - \bar{K}C)\right| < 1, \forall i \Rightarrow \lim_{t\to\infty} \tilde{x}(t) = \mathbf{0}, \forall \hat{x}(t=1|0)$, i.e., $\hat{x}(t)$ asymptotically converges to $x(t)$ independently of $\hat{x}(t=1|0)$

Moreover, if $\mathcal{K}^\infty$ is asymptotically stable, then $\mathcal{K}^\infty$ is also BIBO-stable and $\tilde{x}(t)$ is bounded for any bounded noise $v_1(t)$ and $v_2(t)$

Result #1: Let A be asymptotically stable, i.e., $|\lambda_i(A)| < 1, \forall i$. Then:

- for any positive semidefinite initial condition, the DRE solution asymptotically converges to the same matrix $\bar{P}$ that solves the ARE
- the steady-state Kalman predictor $\mathcal{K}^\infty$ is asymptotically stable

Result #2: Let the data generating system $\mathcal{S}$ be such that:

- the couple (A,C) is observable, i.e., the rank of the observability matrix of (A,C) defined as $\left[C^T \;\; A^T C^T \;\; (A^T)^2 C^T \;\; \cdots \;\; (A^T)^{n-1} C^T\right]$ is equal to $n = \dim(x)$
- the couple (A,B_v) is reachable, i.e., the rank of the reachability matrix of (A,B_v) defined as $\left[B_v \;\; AB_v \;\; A^2 B_v \;\; \cdots \;\; A^{n-1} B_v\right]$ is equal to n, where B_v is any matrix such that the variance of the process noise v_1 can be written as $V_1 = B_v B_v^T$

Then:

- for any positive semidefinite initial condition, the DRE solution asymptotically converges to the same matrix $\bar{P}$ that solves the ARE
- the matrix $\bar{P}$ is positive definite
- the steady-state Kalman predictor $\mathcal{K}^\infty$ is asymptotically stable

Generalization

Let us consider a discrete-time, linear time-invariant (LTI), dynamical system $\mathcal{S}$ with an *exogenous* (deterministic and known) input $u(\cdot)$, described by the state space model:

$$\mathcal{S}: \begin{cases} x(t+1) = Ax(t) + Bu(t) + v_1(t) \\ y(t) = Cx(t) + Du(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $u(t) \in \mathbb{R}^p$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{q \times n}$, $D \in \mathbb{R}^{q \times p}$ are known matrices
- $u(\cdot)$ may possibly depend on the output $y(\cdot)$ through a causal feedback as
$$u(t) = f(y(t), y(t-1), y(t-2), \ldots), \quad \forall t$$
- $v_1(t)$ and $v_2(t)$ are white noises with zero mean value that may be correlated if considered at the same time instant but are uncorrelated at different time instants:
$$E[v_i(t_1) v_j(t_2)^T] = V_{ij} \delta(t_2 - t_1), \quad i = 1, 2; \quad j = 1, 2;$$
- $V_1 = B_v B_v^T \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{q \times q}$, $V_{12} \in \mathbb{R}^{n \times q}$ are known matrices
- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\bar{x}_1, P_1)$, with known $\bar{x}_1 \in \mathbb{R}^n$ and $P_1 \in \mathbb{R}^{n \times n}$, which is uncorrelated with $v_1(t)$ and $v_2(t)$
- the output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$

- The optimal estimate for $x(N+1)$ based on the data vector y^N is given by:

$$\hat{x}(N+1|N) = A\hat{x}(N|N-1) + Bu(N) + \bar{K}e(N)$$

with $\bar{K} \in \mathbb{R}^{n\times q}$ = steady-state predictor gain matrix, defined as:

$$\bar{K} = \left[A\bar{P}C^T + V_{12}\right]\left[C\bar{P}C^T + V_2\right]^{-1}$$

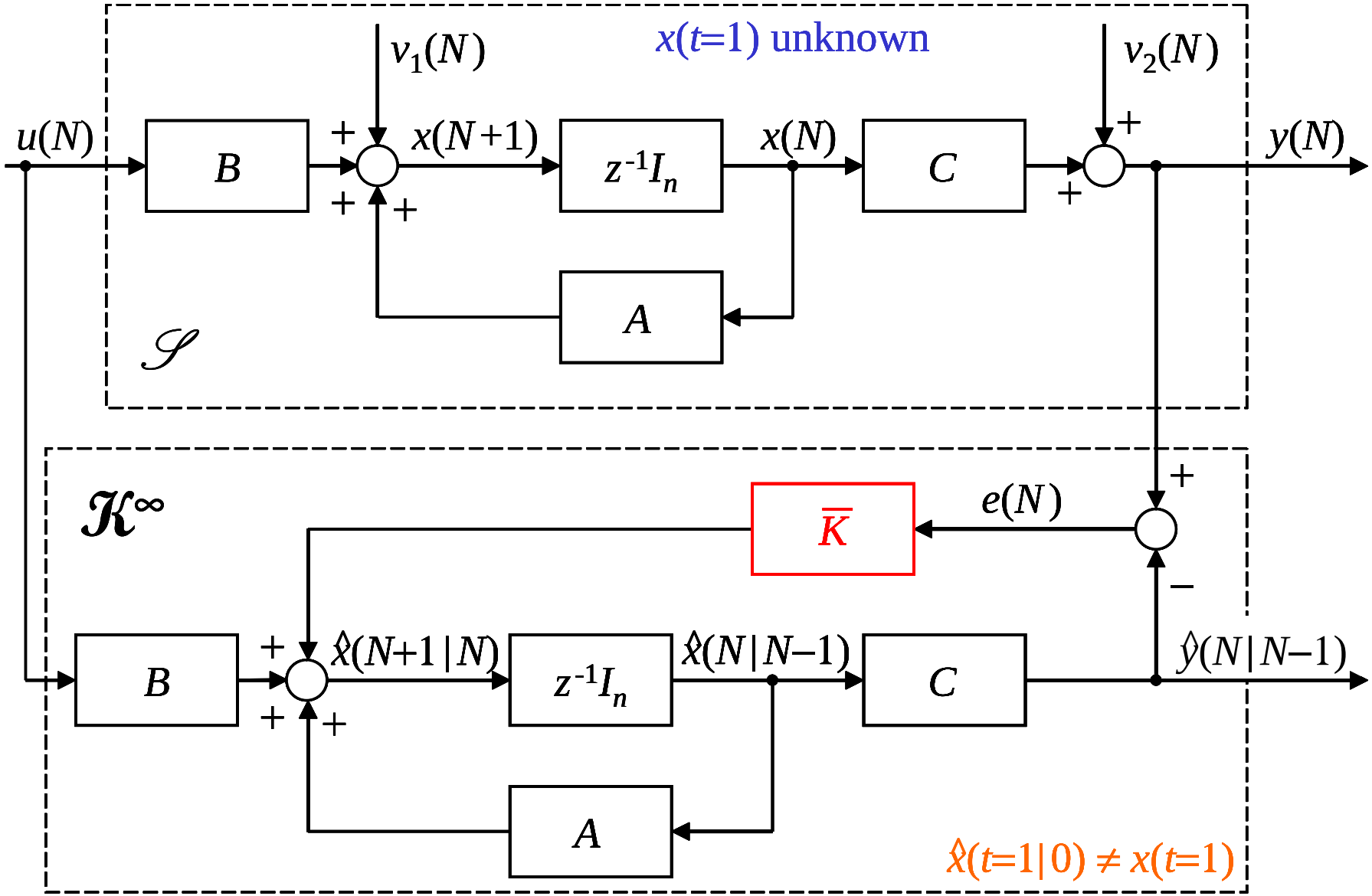
and $\bar{P} \in \mathbb{R}^{n\times n}$ = steady-state variance of the state prediction error, solution of:

$$\bar{P} = A\bar{P}A^T + V_1 - \bar{K}\left[C\bar{P}C^T + V_2\right]\bar{K}^T$$

- The overall equations of the steady-state one-step Kalman predictor are:

$$\mathcal{K}^\infty : \begin{cases} \hat{x}(N+1|N) = A\hat{x}(N|N-1) + Bu(N) + \bar{K}e(N) \\ \hat{y}(N|N-1) = C\hat{x}(N|N-1) + Du(N) \\ e(N) = y(N) - \hat{y}(N|N-1) \end{cases}$$

If $D = \mathbf{0}_{q \times p}$, then:



- Note that $\mathcal{K}^\infty$ is a LTI dynamical system described by the state space model:

$$\begin{aligned}
\hat{x}(N+1|N) &= A\hat{x}(N|N-1)+Bu(N)+\bar{K}e(N) = \\
&= A\hat{x}(N|N-1)+Bu(N)+\bar{K}[y(N)-\hat{y}(N|N-1)] = \\
&= A\hat{x}(N|N-1)+Bu(N)+\bar{K}y(N)-\bar{K}[C\hat{x}(N|N-1)+Du(N)] \\
&= (A-\bar{K}C)\hat{x}(N|N-1)+(B-\bar{K}D)u(N)+\bar{K}y(N) = \\
&= (A-\bar{K}C)\hat{x}(N|N-1)+\begin{bmatrix} B-\bar{K}D & \bar{K} \end{bmatrix}\begin{bmatrix} u(N) \\ y(N) \end{bmatrix} = \\
&= A_{\mathcal{K}^\infty}\hat{x}(N|N-1) + B_{\mathcal{K}^\infty}\begin{bmatrix} u(N) \\ y(N) \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\hat{y}(N|N-1) &= C\hat{x}(N|N-1)+Du(N) = \\
&= C\hat{x}(N|N-1)+\begin{bmatrix} D & \mathbf{0}_{q\times q} \end{bmatrix}\begin{bmatrix} u(N) \\ y(N) \end{bmatrix} = \\
&= C_{\mathcal{K}^\infty}\hat{x}(N|N-1) + D_{\mathcal{K}^\infty}\begin{bmatrix} u(N) \\ y(N) \end{bmatrix}
\end{aligned}$$

- Under MATLAB, the LTI dynamic system $\mathcal{S}$ is described by the state space model:

$$\mathcal{S}: \begin{cases} \mathtt{x}[\mathtt{n}+1] = \mathtt{A}\,\mathtt{x}[\mathtt{n}] + \mathtt{B}\,\mathtt{u}[\mathtt{n}] + \mathtt{G}\,\mathtt{w}[\mathtt{n}] \\ \mathtt{y}[\mathtt{n}] = \mathtt{C}\,\mathtt{x}[\mathtt{n}] + \mathtt{D}\,\mathtt{u}[\mathtt{n}] + \mathtt{H}\,\mathtt{w}[\mathtt{n}] + \mathtt{v}[\mathtt{n}] \end{cases}$$

that corresponds to the standard state space model of $\mathcal{S}$ with:

$$\mathtt{w} \longleftrightarrow v_1, \quad \mathtt{v} \longleftrightarrow v_2, \quad \mathtt{G} \longleftrightarrow I_n, \quad \mathtt{H} \longleftrightarrow \mathbf{0}_{q\times n}$$

$\Rightarrow$ it can be rewritten as:

$$\mathtt{x}[\mathtt{n}+1] = \mathtt{A}\,\mathtt{x}[\mathtt{n}] + \mathtt{B}\,\mathtt{u}[\mathtt{n}] + \mathtt{G}\,\mathtt{w}[\mathtt{n}] = \mathtt{A}\,\mathtt{x}[\mathtt{n}] + \begin{bmatrix} \mathtt{B}, & \mathtt{G} \end{bmatrix} \begin{bmatrix} \mathtt{u}[\mathtt{n}] \\ \mathtt{w}[\mathtt{n}] \end{bmatrix} =$$

$$= \mathtt{A}\,\mathtt{x}[\mathtt{n}] + \begin{bmatrix} \mathtt{B}, & I_n \end{bmatrix} \begin{bmatrix} \mathtt{u}[\mathtt{n}] \\ \mathtt{w}[\mathtt{n}] \end{bmatrix} = \mathtt{A}_{\mathcal{S}}\,\mathtt{x}[\mathtt{n}] + \mathtt{B}_{\mathcal{S}} \begin{bmatrix} \mathtt{u}[\mathtt{n}] \\ \mathtt{w}[\mathtt{n}] \end{bmatrix}$$

$$\mathtt{y}[\mathtt{n}] = \mathtt{C}\,\mathtt{x}[\mathtt{n}] + \mathtt{D}\,\mathtt{u}[\mathtt{n}] + \mathtt{H}\,\mathtt{w}[\mathtt{n}] + \mathtt{v}[\mathtt{n}] = \mathtt{C}\,\mathtt{x}[\mathtt{n}] + \begin{bmatrix} \mathtt{D}, & \mathtt{H} \end{bmatrix} \begin{bmatrix} \mathtt{u}[\mathtt{n}] \\ \mathtt{w}[\mathtt{n}] \end{bmatrix} + \mathtt{v}[\mathtt{n}]$$

$$= \mathtt{C}\,\mathtt{x}[\mathtt{n}] + \begin{bmatrix} \mathtt{D}, & \mathbf{0}_{q\times n} \end{bmatrix} \begin{bmatrix} \mathtt{u}[\mathtt{n}] \\ \mathtt{w}[\mathtt{n}] \end{bmatrix} + \mathtt{v}[\mathtt{n}] = \mathtt{C}_{\mathcal{S}}\,\mathtt{x}[\mathtt{n}] + \mathtt{D}_{\mathcal{S}} \begin{bmatrix} \mathtt{u}[\mathtt{n}] \\ \mathtt{w}[\mathtt{n}] \end{bmatrix} + \mathtt{v}[\mathtt{n}]$$

$\Rightarrow$ under MATLAB, $\mathcal{S}$ can be defined as:

```
S=ss(A,[B,eye(n)],C,[D,zeros(q,n)],1)
```

- Under MATLAB, the steady-state one-step predictor $\mathcal{K}^\infty$ can be computed as:

```
[Kalman_pred,Kbar,Pbar,K0bar]=kalman(S,V1,V2,V12)
```

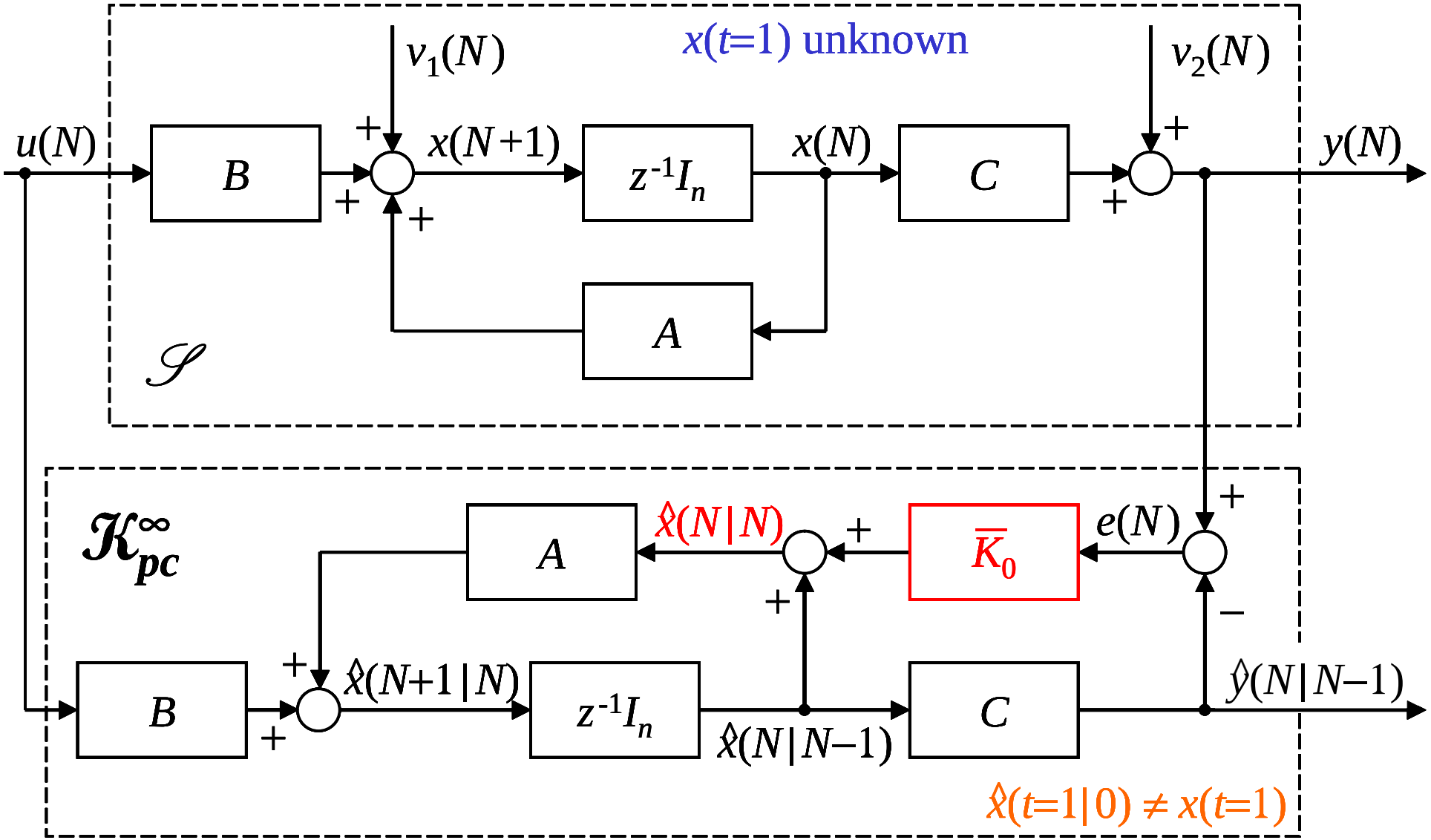
where:

$$\begin{aligned}
\mathtt{S} &= \text{state space model of } \mathcal{S} = \mathtt{ss}(\mathtt{A}_{\mathcal{S}}, \mathtt{B}_{\mathcal{S}}, \mathtt{C}_{\mathcal{S}}, \mathtt{D}_{\mathcal{S}}, 1) \\
\mathtt{V1} &= V_1 = B_v B_v^T \in \mathbb{R}^{n\times n} = \text{variance matrix of } v_1 \\
\mathtt{V2} &= V_2 \in \mathbb{R}^{q\times q} = \text{variance matrix of } v_2 \\
\mathtt{V12} &= V_{12} \in \mathbb{R}^{n\times q} = \text{covariance matrix of } v_1 \text{ and } v_2
\end{aligned}$$

and:

$$\begin{aligned}
\mathtt{Kalman_pred} &= \text{state space model of } \mathcal{K}^\infty = \mathtt{ss}(\mathtt{A}_{\mathcal{K}^\infty}, \mathtt{B}_{\mathcal{K}^\infty}, \mathtt{C}_{\mathcal{K}^\infty}, \mathtt{D}_{\mathcal{K}^\infty}, 1) \\
\mathtt{Kbar} &= \bar{K} = (A\bar{P}C^T + V_{12})(C\bar{P}C^T + V_2)^{-1} \in \mathbb{R}^{n\times q} \\
\mathtt{Pbar} &= \bar{P} \in \mathbb{R}^{n\times n}, \text{ solution of the ARE} \\
& \bar{P} = A\bar{P}A^T + V_1 - \bar{K}(C\bar{P}C^T + V_2)\bar{K}^T \\
\mathtt{K0bar} &= \bar{K}_0 = \bar{P}C^T\left[C\bar{P}C^T + V_2\right]^{-1} \in \mathbb{R}^{n\times q}
\end{aligned}$$

Predictor/corrector form of steady-state predictor $(D=\mathbf{0}_{q\times p}, V_{12}=\mathbf{0}_{n\times q})$



Linearization of nonlinear dynamic systems

Given a discrete-time, nonlinear, time-variant, dynamical system $\mathcal{S}$ described by:

$$\mathcal{S}: \begin{cases} x(t+1) = f(t, x(t), u(t)) \\ y(t) = h(t, x(t), u(t)) \end{cases} \qquad t = 1, 2, \ldots$$

and the nominal movement $\bar{x}(t)$ corresponding to nominal $\bar{u}(t)$ and $\bar{y}(t)$, define:

- the state perturbation $\delta x(t) = x(t) - \bar{x}(t) \;\Rightarrow\; x(t) = \bar{x}(t) + \delta x(t)$
- the input perturbation $\delta u(t) = u(t) - \bar{u}(t) \;\Rightarrow\; u(t) = \bar{u}(t) + \delta u(t)$
- the output perturbation $\delta y(t) = y(t) - \bar{y}(t) \;\Rightarrow\; y(t) = \bar{y}(t) + \delta y(t)$

The perturbation dynamics are well approximated by the **linearized dynamic system** $\mathcal{S}_\ell$:

$$\mathcal{S}_\ell: \begin{cases} \delta x(t+1) = \bar{A}(t)\delta x(t) + \bar{B}(t)\delta u(t) \\ \delta y(t) = \bar{C}(t)\delta x(t) + \bar{D}(t)\delta u(t) \end{cases} \qquad t = 1, 2, \ldots$$

$\bar{A}(t) = \left.\dfrac{\partial f(\cdot)}{\partial x}\right|_{\substack{x(t)=\bar{x}(t)\\ u(t)=\bar{u}(t)}}, \bar{B}(t) = \left.\dfrac{\partial f(\cdot)}{\partial u}\right|_{\substack{x(t)=\bar{x}(t)\\ u(t)=\bar{u}(t)}}$: Jacobians of f with respect to x and u

$\bar{C}(t) = \left.\dfrac{\partial h(\cdot)}{\partial x}\right|_{\substack{x(t)=\bar{x}(t)\\ u(t)=\bar{u}(t)}}, \bar{D}(t) = \left.\dfrac{\partial h(\cdot)}{\partial u}\right|_{\substack{x(t)=\bar{x}(t)\\ u(t)=\bar{u}(t)}}$: Jacobians of h with respect to x and u

if $\delta x(t), \delta u(t)$ and $\delta y(t)$ are <u>sufficiently small</u> (note that $\mathcal{S}_\ell$ is linear and time-variant)

Nonlinear Kalman filtering

Let us consider a discrete-time, nonlinear, time-variant, dynamic system $\mathcal{S}$ described by:

$$\mathcal{S}: \begin{cases} x(t+1) = f(t, x(t)) + v_1(t) \\ y(t) = h(t, x(t)) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $v_1(t) \in \mathbb{R}^n$, $v_2(t) \in \mathbb{R}^q$, and assume that:

- $v_1(t)$ and $v_2(t)$ are white noises with zero mean value such that:

  $E[v_1(t_1)v_1(t_2)^T] = V_1\delta(t_2 - t_1)$ (whiteness of $v_1(t)$)

  $E[v_2(t_1)v_2(t_2)^T] = V_2\delta(t_2 - t_1)$ (whiteness of $v_2(t)$)

  $E[v_1(t_1)v_2(t_2)^T] = \mathbf{0}_{n\times q}, \quad \forall t_1, t_2$ (uncorrelation of $v_1(t)$ and $v_2(t)$)
- $f : \mathbb{R}^n \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^q$ are known nonlinear functions, $V_1 \in \mathbb{R}^{n\times n}$ and $V_2 \in \mathbb{R}^{q\times q}$ are known matrices
- the initial state $x(t=1)$ is an unknown random vector: $x(t=1) \sim (\bar{x}_1, P_1)$, with known $\bar{x}_1 \in \mathbb{R}^n$ and $P_1 \in \mathbb{R}^{n\times n}$, which is uncorrelated with $v_1(t)$ and $v_2(t)$
- the output measurements $y(t)$ are available for $t = 1, 2, \ldots, N$
- **Goal:** estimate the state $x(N+1)$

Linearized Kalman predictor

- Let the nominal movement $\bar{x}(\cdot)$ be solution of the deterministic difference equation:
$$\bar{x}(t+1) = f(t, \bar{x}(t)), \qquad \bar{x}(t=1) = E[x(t=1)] = \bar{x}_1$$
and the corresponding nominal output $\bar{y}(\cdot)$ be defined as:
$$\bar{y}(t) = h(t, \bar{x}(t))$$
$\bar{x}(\cdot)$ and $\bar{y}(\cdot)$ can be a priori computed, since they are independent of the data
- The nonlinear system $\mathcal{S}$ can be well approximated in the neighborhood of $\bar{x}(\cdot)$ by the linearized dynamic system $\mathcal{S}_\ell$ described by the state space model:
$$\mathcal{S}_\ell : \begin{cases} \delta x(t+1) = \bar{A}(t)\delta x(t) + v_1(t) \\ \delta y(t) = \bar{C}(t)\delta x(t) + v_2(t) \end{cases} \qquad t = 1, 2, \ldots$$
where
$$\delta x(t) = x(t) - \bar{x}(t), \qquad \delta y(t) = y(t) - \bar{y}(t)$$
$$\bar{A}(t) = \left.\frac{\partial f(t,x)}{\partial x}\right|_{x(t)=\bar{x}(t)}, \quad \bar{C}(t) = \left.\frac{\partial h(t,x)}{\partial x}\right|_{x(t)=\bar{x}(t)}$$
and the initial state $\delta x(t=1)$ is an unknown random vector: $\delta x(t=1) \sim (\mathbf{0}, P_1)$

- The optimal estimate for the state perturbation $\delta x(N+1)$ based on N data is provided by the one-step Kalman predictor described by the state space model:

$$\overline{\mathcal{K}}: \begin{cases} \widehat{\delta x}(N+1|N) = \bar{A}(N)\widehat{\delta x}(N|N-1) + \bar{K}(N)e(N) \\ \widehat{\delta y}(N|N-1) = \bar{C}(N)\widehat{\delta x}(N|N-1) \\ e(N) = \delta y(N) - \widehat{\delta y}(N|N-1) \end{cases}$$

where the linearized Kalman predictor gain matrix $\bar{K}(N)$ is equal to:

$$\bar{K}(N) = \bar{A}(N)P(N)\bar{C}(N)^T \left[\bar{C}(N)P(N)\bar{C}(N)^T + V_2\right]^{-1}$$

with the state prediction error variance $P(N)$ given by the following DRE:

$$P(N+1) = \bar{A}(N)P(N)\bar{A}(N)^T + V_1 - \bar{K}(N)\left[\bar{C}(N)P(N)\bar{C}(N)^T + V_2\right]\bar{K}(N)^T$$

assuming as initial values: $\widehat{\delta x}(1|0) = \mathbf{0}$, $P(1) = P_1$

- Since $\delta x(t) = x(t) - \bar{x}(t) \Rightarrow x(t) = \bar{x}(t) + \delta x(t)$, where $\bar{x}(t)$ is deterministic $\Rightarrow$ the estimate for the state $x(N+1)$ based on N data is given by:

$$\hat{x}(N+1|N) = \bar{x}(N+1) + \widehat{\delta x}(N+1|N)$$

where $\widehat{\delta x}(N+1|N)$ is provided by the one-step Kalman predictor $\overline{\mathcal{K}}$

- The estimate for the state $x(N+1)$ based on N data can be approximated as:

$$\hat{x}(N+1|N) = \bar{x}(N+1) + \widehat{\delta x}(N+1|N) =$$
$$= \underbrace{f(N,\bar{x}(N)) + \bar{A}(N)\widehat{\delta x}(N|N-1)}_{\approx f(N,\hat{x}(N|N-1))} + \bar{K}(N)\left[\delta y(N) - \widehat{\delta y}(N|N-1)\right] \approx$$
$$\approx f(N,\hat{x}(N|N-1)) + \bar{K}(N)\left[y(N) - \bar{y}(N) - \bar{C}(N)\widehat{\delta x}(N|N-1)\right] \approx$$
$$\approx f(N,\hat{x}(N|N-1)) + \bar{K}(N)\left\{y(N) - \left[\underbrace{h(N,\bar{x}(N)) + \bar{C}(N)\widehat{\delta x}(N|N-1)}_{\approx h(N,\hat{x}(N|N-1))}\right]\right\} \approx$$
$$\approx f(N,\hat{x}(N|N-1)) + \bar{K}(N)\left[y(N) - h(N,\hat{x}(N|N-1))\right]$$

by exploiting the linearizations of $f(t,x(t))$ and $h(t,x(t))$ around $\bar{x}(t)$, i.e.:

$$f(t,x(t)) \approx f(t,\bar{x}(t)) + \bar{A}(t)\left[x(t) - \bar{x}(t)\right] = f(t,\bar{x}(t)) + \bar{A}(t)\delta x(t)$$
$$h(t,x(t)) \approx h(t,\bar{x}(t)) + \bar{C}(t)\left[x(t) - \bar{x}(t)\right] = h(t,\bar{x}(t)) + \bar{C}(t)\delta x(t)$$

evaluated at $t = N$ and $x(t) = \hat{x}(N|N-1)$:

$$f(N,\hat{x}(N|N-1)) \approx f(N,\bar{x}(N)) + \bar{A}(N)\widehat{\delta x}(N|N-1)$$
$$h(N,\hat{x}(N|N-1)) \approx h(N,\bar{x}(N)) + \bar{C}(N)\widehat{\delta x}(N|N-1)$$

- By summarizing, the estimate for the state $x(N+1)$ of the nonlinear system $\mathcal{S}$ can be provided by the **linearized (or tangent) Kalman predictor** $\mathcal{LK}$ described by the following nonlinear state space model:

$$\mathcal{LK}: \begin{cases} \hat{x}(N+1|N) = f(N, \hat{x}(N|N-1)) + \bar{K}(N)e(N) \\ \hat{y}(N|N-1) = h(N, \hat{x}(N|N-1)) \\ e(N) = y(N) - \hat{y}(N|N-1) \end{cases}$$

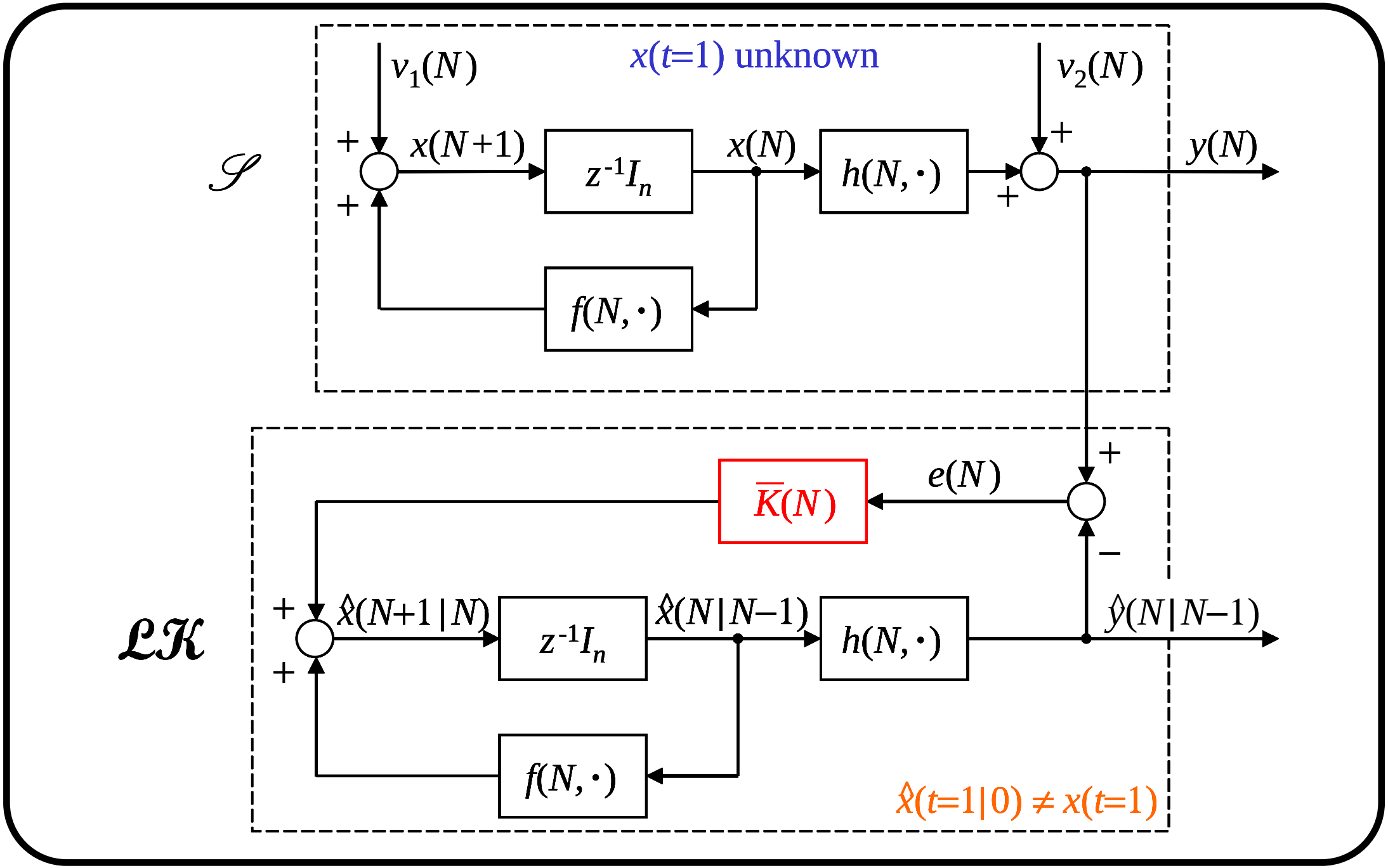
where $\bar{K}(N)$ depends on the matrices $\bar{A}(N)$ and $\bar{C}(N)$ of the linearized system $\mathcal{S}_\ell$, which are a priori computed around the nominal movement $\bar{x}(\cdot)$ and then do not take into account the information on the actual state $x(\cdot)$ provided by the data

$$\Downarrow$$

the perturbations introduced by the noises may lead to huge estimation errors

$$\Downarrow$$

the linearized Kalman predictor is not used in the practical applications


$x(t{=}1)$ unknown
$v_1(N)$
$v_2(N)$
$\mathscr{S}$
$x(N+1)$
$z^{-1}I_n$
$x(N)$
$h(N,\cdot)$
$y(N)$
$f(N,\cdot)$
$\overline{K}(N)$
$e(N)$
$\mathscr{LK}$
$\hat{x}(N+1|N)$
$z^{-1}I_n$
$\hat{x}(N|N-1)$
$h(N,\cdot)$
$\hat{y}(N|N-1)$
$f(N,\cdot)$
$\hat{x}(t{=}1|0) \neq x(t{=}1)$

Extended Kalman predictor

Much better results are obtained by linearizing the functions $f(t, x(t))$ and $h(t, x(t))$ around the last state estimate $\hat{x}(N|N-1)$ instead of the nominal movement $\bar{x}(t)$, i.e., by considering instead of the matrices $\bar{A}(N)$ and $\bar{C}(N)$ the following matrices:

$$\hat{A}(N|N-1) = \left.\frac{\partial f(t,x)}{\partial x}\right|_{\substack{t=N \\ x(N)=\hat{x}(N|N-1)}}, \hat{C}(N|N-1) = \left.\frac{\partial h(t,x)}{\partial x}\right|_{\substack{t=N \\ x(N)=\hat{x}(N|N-1)}}$$

which allow to define the **extended Kalman predictor** $\mathcal{EK}$ described by the model:

$$\mathcal{EK}: \begin{cases} \hat{x}(N+1|N) = f(N, \hat{x}(N|N-1)) + \hat{K}(N)e(N) \\ \hat{y}(N|N-1) = h(N, \hat{x}(N|N-1)) \\ e(N) = y(N) - \hat{y}(N|N-1) \end{cases}$$

where the extended Kalman predictor gain matrix $\hat{K}(N)$ is equal to:

$$\hat{K}(N) = \hat{A}(N|N-1)P(N)\hat{C}(N|N-1)^T[\hat{C}(N|N-1)P(N)\hat{C}(N|N-1)^T + V_2]^{-1}$$

with the state prediction error variance $P(N)$ given by the following DRE:

$$P(N+1) = \hat{A}(N|N-1)P(N)\hat{A}(N|N-1)^T + V_1 - \hat{K}(N)[\hat{C}(N|N-1)P(N)\hat{C}(N|N-1)^T + V_2]\hat{K}(N)^T$$

- From a computational point of view, the extended Kalman predictor is much more demanding than the linearized one, since the linearizations cannot be computed off-line and just once, but have to be performed at the run-time, because they depend on the state estimate provided by the predictor at the previous step
- Extended Kalman predictors are widely used in many practical application fields where state estimates of nonlinear systems are involved, like in mobile robot localization, communication systems (e.g., Global Positioning Systems or GPS), etc.
- Extended Kalman predictors allow also to estimate uncertain physical parameters, which have to be suitably considered as state variables added to the original ones to produce an augmented state vector
- The variance matrices V_1 and V_2 play a crucial role in the Kalman filtering:
  - V_2 is chosen to account for the uncertainty on the measurements
  - V_1 is chosen to allow a suitable trade-off of the confidence level between the measurements (i.e., the *a posteriori* information) and the *a priori* information on the initial state (i.e., $\bar{x}_1$ and P_1)