EE210A: Adaptation and Learning

Professor Ali H. Sayed

LECTURE #07

Rudolph Kalman
(1930-)

KALMAN FILTERING

Sections in order: 7.1-7.7

7.1 INNOVATIONS PROCESS

Consider two zero-mean random variables $\{\boldsymbol{x}, \boldsymbol{y}\}$. We already know from Thm. 3.1 that the linear least-mean-squares estimator of $\boldsymbol{x}$ given $\boldsymbol{y}$ is $\hat{\boldsymbol{x}} = K_o\boldsymbol{y}$, where K_o is any solution to the normal equations

$$K_oR_y = R_{xy} \tag{7.1}$$

In the sequel we assume that R_y is positive-definite so that K_o is uniquely defined as $K_o = R_{xy}R_y^{-1}$.

Usually, the variable $\boldsymbol{y}$ is vector-valued, say $\boldsymbol{y} = \text{col}\{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$, where each $\boldsymbol{y}_i$ is also possibly a vector. Now assume that we could somehow replace $\boldsymbol{y}$ by another vector $\boldsymbol{e}$ of similar dimensions, say

$$\boldsymbol{e} = A\boldsymbol{y} \tag{7.2}$$

for some lower triangular invertible matrix A.

INNOVATIONS

Assume further that the transformation A could be chosen such that the entries of $\boldsymbol{e}$, denoted by $\boldsymbol{e} = \text{col}\{\boldsymbol{e}_0, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}$, are uncorrelated with each other, i.e.,

$$\mathsf{E}\,\boldsymbol{e}_i\boldsymbol{e}_j^* \;\stackrel{\Delta}{=}\; R_{e,i}\delta_{ij}$$

where δ_{ij} denotes the Kronecker delta function that is unity when $i = j$ and zero otherwise, and $R_{e,i}$ denotes the covariance matrix of $\boldsymbol{e}_i$. Then the covariance matrix of $\boldsymbol{e}$ will be block diagonal,

$$R_e \;\stackrel{\Delta}{=}\; \mathsf{E}\,\boldsymbol{e}\boldsymbol{e}^* \;=\; \text{diag}\{R_{e,0}, R_{e,1}, \ldots, R_{e,N}\}$$

and, in addition, the problem of estimating $\boldsymbol{x}$ from $\boldsymbol{y}$ would be equivalent to the problem of estimating $\boldsymbol{x}$ from $\boldsymbol{e}$. To see this, let $\hat{\boldsymbol{x}}_{|\boldsymbol{e}}$ denote the linear least-mean-squares estimator of $\boldsymbol{x}$ given $\boldsymbol{e}$, i.e.,

$$\hat{\boldsymbol{x}}_{|\boldsymbol{e}} = R_{xe}R_e^{-1}\boldsymbol{e} \qquad (7.3)$$

INNOVATIONS

Likewise, let $\hat{x}_{|y}$ denote the estimator of x given y,

$$\hat{x}_{|y} = R_{xy} R_y^{-1} y \tag{7.4}$$

Then since

$$R_e = \mathsf{E}\, ee^* = A\,(\mathsf{E}\, yy^*)\, A^* = A R_y A^*$$

and

$$R_{xe} = \mathsf{E}\, xe^* = (\mathsf{E}\, xy^*)\, A^* = R_{xy} A^*$$

we find that

$$\hat{x}_{|e} \;=\; R_{xe} R_e^{-1} e \;=\; R_{xy} A^* \left(A R_y A^*\right)^{-1} e = R_{xy} R_y^{-1} A^{-1} e = R_{xy} R_y^{-1} y$$

That is,

$$\boxed{\hat{x}_{|e} = \hat{x}_{|y}} \tag{7.5}$$

The key advantage of working with $\boldsymbol{e}$ instead of $\boldsymbol{y}$ is that R_e in (7.3) is block-diagonal and, hence, the estimator $\hat{\boldsymbol{x}}_{|\boldsymbol{e}}$ can be evaluated as the combined sum of individual estimators. Specifically, expression (7.3) gives

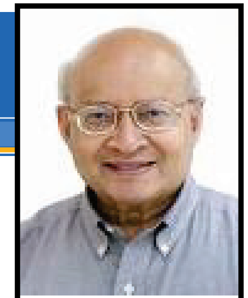
$$\hat{\boldsymbol{x}}_{|\boldsymbol{e}} = \sum_{i=0}^{N} \left(\mathsf{E}\,\boldsymbol{x}\boldsymbol{e}_i^*\right) R_{e,i}^{-1}\boldsymbol{e}_i = \sum_{i=0}^{N} \hat{\boldsymbol{x}}_{|\boldsymbol{e}_i}$$

This result shows that we can estimate $\boldsymbol{x}$ from $\boldsymbol{y}$ by estimating $\boldsymbol{x}$ individually from each $\boldsymbol{e}_i$ and then combining the resulting estimators. In particular, if we replace the notations $\hat{\boldsymbol{x}}_{|\boldsymbol{e}}$ and $\hat{\boldsymbol{x}}_{|\boldsymbol{y}}$ by the more suggestive notation $\hat{\boldsymbol{x}}_{|N}$, in order to indicate that the estimator of $\boldsymbol{x}$ is based on the observations $\boldsymbol{y}_0$ through $\boldsymbol{y}_N$, then the above expression shows that

$$\hat{\boldsymbol{x}}_{|N} = \sum_{i=0}^{N} \hat{\boldsymbol{x}}_{|\boldsymbol{e}_i} = \hat{\boldsymbol{x}}_{|\boldsymbol{e}_N} + \sum_{i=0}^{N-1} \hat{\boldsymbol{x}}_{|\boldsymbol{e}_i}$$

where the last sum on the right-hand side is simply the estimator of $\boldsymbol{x}$ using the observations $\boldsymbol{y}_0$ through $\boldsymbol{y}_{N-1}$.

RECURSIVE ESTIMATION



T. Kailath
(1935-)

It follows that

$$\hat{\boldsymbol{x}}_{|N} = \hat{\boldsymbol{x}}_{|N-1} + \hat{\boldsymbol{x}}_{|\boldsymbol{e}_N}$$

i.e.,

$$\boxed{\hat{\boldsymbol{x}}_{|N} = \hat{\boldsymbol{x}}_{|N-1} \;+\; \left(\mathsf{E}\,\boldsymbol{x}\boldsymbol{e}_N^*\right) R_{e,N}^{-1}\boldsymbol{e}_N} \qquad (7.6)$$

This is a useful recursive formula; it shows how the estimator of $\boldsymbol{x}$ can be updated recursively by adding the contribution of the most recent variable $\boldsymbol{e}_N$.

The question now is how to generate the variables $\{\boldsymbol{e}_i\}$ from the $\{\boldsymbol{y}_i\}$. One possible transformation is the so-called Gram-Schmidt procedure. Let $\hat{\boldsymbol{y}}_{i|i-1}$ denote the estimator of $\boldsymbol{y}_i$ that is based on the observations up to time $i-1$, i.e., on $\{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}\}$. The same argument that led to (7.5) shows that $\hat{\boldsymbol{y}}_{i|i-1}$ can be alternatively calculated by estimating $\boldsymbol{y}_i$ from $\{\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{i-1}\}$. Then we can construct $\boldsymbol{e}_i$ as

$$\boxed{\boldsymbol{e}_i \;\stackrel{\Delta}{=}\; \boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i|i-1}} \qquad \textbf{INNOVATIONS} \qquad (7.7)$$

That is, we can choose $\boldsymbol{e}_i$ as the estimation error that results from estimating $\boldsymbol{y}_i$ from the observations $\{\boldsymbol{y}_0, \boldsymbol{y}_1 \ldots, \boldsymbol{y}_{i-1}\}$.

UNCORRELATEDNESS

In order to verify that the resulting $\{\boldsymbol{e}_i\}$ are uncorrelated with each other, we recall that, by virtue of the orthogonality condition of linear least-mean-squares estimation (cf. Thm. 4.1),

$$\boldsymbol{e}_i \perp \{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}\}$$

That is, $\boldsymbol{e}_i$ is uncorrelated with the observations $\{\boldsymbol{y}_0, \boldsymbol{y}_1 \ldots, \boldsymbol{y}_{i-1}\}$. It then follows that $\boldsymbol{e}_i$ should be uncorrelated with any $\boldsymbol{e}_j$ for $j < i$ since, by definition, $\boldsymbol{e}_j$ is a linear combination of the observations $\{\boldsymbol{y}_0, \boldsymbol{y}_1 \ldots, \boldsymbol{y}_j\}$ and, moreover,

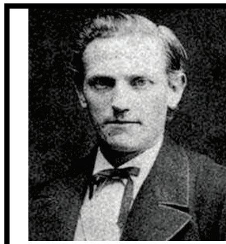
$$\{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_j\} \subset \{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}\} \qquad \text{for} \quad j < i$$

By the same token, $\boldsymbol{e}_i$ is uncorrelated with any $\boldsymbol{e}_j$ for $j > i$.

GRAM-SCHMIDT PROCEDURE

It is instructive to see what choice of a transformation A in (7.2) corresponds to the use of the Gram-Schmidt procedure. Assume, for illustration purposes, that $N = 2$. Then writing (7.7) for $i = 0, 1, 2$ we get

$$\begin{bmatrix} \boldsymbol{e}_0 \\ \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{bmatrix} = \begin{bmatrix} \mathrm{I} & & \\ -(\mathsf{E}\,\boldsymbol{y}_1\boldsymbol{y}_0^*)(\mathsf{E}\,\boldsymbol{y}_0\boldsymbol{y}_0^*)^{-1} & \mathrm{I} & \\ \times & \times & \mathrm{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$$



Jorgen P. Gram
(1850-1916)

Erhard Schmidt
(1876-1959)

where the entries $\times$ arise from the calculation

$$\begin{bmatrix} \times & \times \end{bmatrix} = \left(\mathsf{E}\,\boldsymbol{y}_2 \begin{bmatrix} \boldsymbol{y}_0^* & \boldsymbol{y}_1^* \end{bmatrix}\right) \left(\mathsf{E} \begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \end{bmatrix}^*\right)^{-1}$$

We thus find that A is a lower triangular transformation with unit entries along its diagonal. The lower triangularity of A is relevant since it translates into a causal relationship between the $\{\boldsymbol{e}_i\}$ and the $\{\boldsymbol{y}_i\}$. By causality we mean that each $\boldsymbol{e}_i$ can be computed from $\{\boldsymbol{y}_j, j \leq i\}$ and, similarly, each $\boldsymbol{y}_i$ can be recovered from $\{\boldsymbol{e}_j, j \leq i\}$. We also see from the construction (7.7) that we can regard $\boldsymbol{e}_i$ as the "new information" in $\boldsymbol{y}_i$ given $\{\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1}\}$. Therefore, it is customary to refer to the $\{\boldsymbol{e}_i\}$ as the innovations process associated with the $\{\boldsymbol{y}_i\}$.

7.2 STATE-SPACE MODEL

As we now proceed to show, the Kalman filter is an efficient procedure for determining the innovations when the observation process $\{\boldsymbol{y}_i\}$ arises from a finite-dimensional linear state-space model.

What we mean by a state-space model for $\{\boldsymbol{y}_i\}$ is the following. We assume that $\boldsymbol{y}_i$ satisfies an equation of the form

$$\boxed{\boldsymbol{y}_i = H_i \boldsymbol{x}_i + \boldsymbol{v}_i, \quad i \geq 0} \tag{7.8}$$

in terms of an $n \times 1$ so-called state-vector $\boldsymbol{x}_i$, which in turn obeys a recursion of the form

$$\boxed{\boldsymbol{x}_{i+1} = F_i \boldsymbol{x}_i + G_i \boldsymbol{n}_i, \quad i \geq 0} \tag{7.9}$$

The processes $\boldsymbol{v}_i$ and $\boldsymbol{n}_i$ are assumed to be $p \times 1$ and $m \times 1$ zero-mean white noise processes, respectively, with covariances and cross-covariances denoted by

$$\mathsf{E}\begin{bmatrix} \boldsymbol{n}_i \\ \boldsymbol{v}_i \end{bmatrix}\begin{bmatrix} \boldsymbol{n}_j \\ \boldsymbol{v}_j \end{bmatrix}^* \triangleq \begin{bmatrix} Q_i & S_i \\ S_i^* & R_i \end{bmatrix}\delta_{ij}$$

INITIAL CONDITIONS

whereas the initial state $\boldsymbol{x}_0$ is assumed to have zero mean, covariance matrix Π_0, and to be uncorrelated with $\{\boldsymbol{n}_i\}$ and $\{\boldsymbol{v}_i\}$, i.e.,

$$\mathsf{E}\,\boldsymbol{x}_0\boldsymbol{x}_0^* = \Pi_0, \quad \mathsf{E}\,\boldsymbol{n}_i\boldsymbol{x}_0^* = 0, \quad \text{and} \quad \mathsf{E}\,\boldsymbol{v}_i\boldsymbol{x}_0^* = 0 \quad \text{for all} \quad i \geq 0$$

The assumptions on $\{\boldsymbol{x}_0, \boldsymbol{n}_i, \boldsymbol{v}_i\}$ can be compactly restated as

$$\mathsf{E}\begin{bmatrix} \boldsymbol{n}_i \\ \boldsymbol{v}_i \\ \boldsymbol{x}_0 \\ 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{n}_j \\ \boldsymbol{v}_j \\ \boldsymbol{x}_0 \end{bmatrix}^* = \begin{bmatrix} \begin{bmatrix} Q_i & S_i \\ S_i^* & R_i \end{bmatrix}\delta_{ij} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \Pi_0 \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix} \qquad (7.10)$$

It is also assumed that the matrices

$$F_i\ (n\times n), \quad G_i\ (n\times m), \quad H_i\ (p\times n), \quad Q_i\ (m\times m), \quad R_i\ (p\times p), \quad S_i\ (m\times p), \quad \Pi_0\ (n\times n)$$

are known *a priori*. The process $\boldsymbol{v}_i$ is called measurement noise and the process $\boldsymbol{n}_i$ is called process noise. We now examine how the innovations $\{\boldsymbol{e}_i\}$ of a process $\{\boldsymbol{y}_i\}$ satisfying a state-space model of the form (7.8)–(7.10) can be evaluated.

7.3 RECURSION FOR THE STATE ESTIMATOR

Let $\{\hat{\boldsymbol{y}}_{i|i-1}, \hat{\boldsymbol{x}}_{i|i-1}, \hat{\boldsymbol{v}}_{i|i-1}\}$ denote the estimators of the variables $\{\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{v}_i\}$ from the observations $\{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}\}$, respectively. Then using $\boldsymbol{y}_i = H_i \boldsymbol{x}_i + \boldsymbol{v}_i$, and appealing to linearity, we have

$$\hat{\boldsymbol{y}}_{i|i-1} = H_i \hat{\boldsymbol{x}}_{i|i-1} + \hat{\boldsymbol{v}}_{i|i-1} \tag{7.11}$$

Now the assumptions on our state-space model imply that

$$\boldsymbol{v}_i \perp \boldsymbol{y}_j \quad \text{for} \quad j \leq i-1$$

i.e., $\boldsymbol{v}_i$ is uncorrelated with the observations $\{\boldsymbol{y}_j, j \leq i-1\}$, so that

$$\hat{\boldsymbol{v}}_{i|i-1} = 0$$

This is because from the model (7.8)–(7.9), $\boldsymbol{y}_j$ is a linear combination of the variables $\{\boldsymbol{v}_j, \boldsymbol{n}_{j-1}, \ldots, \boldsymbol{n}_0, \boldsymbol{x}_0\}$, all of which are uncorrelated with $\boldsymbol{v}_i$ for $j \leq i-1$.

STATE ESTIMATOR

Consequently,

$$\boldsymbol{e}_i = \boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i|i-1} = \boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_{i|i-1} \tag{7.12}$$

Therefore, the problem of finding the innovations reduces to one of finding $\hat{\boldsymbol{x}}_{i|i-1}$. For this purpose, we can use (7.6) to write

$$\begin{aligned} \hat{\boldsymbol{x}}_{i+1|i} &= \hat{\boldsymbol{x}}_{i+1|i-1} + \left(\mathsf{E}\,\boldsymbol{x}_{i+1}\boldsymbol{e}_i^*\right) R_{e,i}^{-1} \boldsymbol{e}_i \\ &= \hat{\boldsymbol{x}}_{i+1|i-1} + \left(\mathsf{E}\,\boldsymbol{x}_{i+1}\boldsymbol{e}_i^*\right) R_{e,i}^{-1} (\boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_{i|i-1}) \end{aligned} \tag{7.13}$$

where

$$\boxed{R_{e,i} \triangleq \mathsf{E}\,\boldsymbol{e}_i \boldsymbol{e}_i^*} \tag{7.14}$$

But since $\boldsymbol{x}_{i+1}$ obeys the state equation $\boldsymbol{x}_{i+1} = F_i \boldsymbol{x}_i + G_i \boldsymbol{n}_i$, we also obtain, again by linearity, that

$$\hat{\boldsymbol{x}}_{i+1|i-1} = F_i \hat{\boldsymbol{x}}_{i|i-1} + G_i \hat{\boldsymbol{n}}_{i|i-1} = F_i \hat{\boldsymbol{x}}_{i|i-1} + 0 \tag{7.15}$$

since $\boldsymbol{n}_i \perp \boldsymbol{y}_j, j \leq i-1$.

STATE ESTIMATOR

By combining Eqs. (7.12)–(7.15) we arrive at the following recursive equations for determining the innovations:

$$\boxed{\begin{array}{rcl} \boldsymbol{e}_i & = & \boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_{i|i-1} \\ \hat{\boldsymbol{x}}_{i+1|i} & = & F_i \hat{\boldsymbol{x}}_{i|i-1} + K_{p,i} \boldsymbol{e}_i, \quad i \geq 0 \end{array}} \tag{7.16}$$

with initial conditions

$$\hat{\boldsymbol{x}}_{0|-1} = 0, \qquad \boldsymbol{e}_0 = \boldsymbol{y}_0 \tag{7.17}$$

and where we have defined the gain matrix

$$\boxed{K_{p,i} \triangleq \left(\mathsf{E}\, \boldsymbol{x}_{i+1} \boldsymbol{e}_i^*\right) R_{e,i}^{-1}} \tag{7.18}$$

The subscript "p" indicates that $K_{p,i}$ is used to update a predicted estimator of the state vector. By combining the equations in (7.16) we also find that

$$\boxed{\hat{\boldsymbol{x}}_{i+1|i} = F_{p,i} \hat{\boldsymbol{x}}_{i|i-1} + K_{p,i} \boldsymbol{y}_i, \quad F_{p,i} \triangleq F_i - K_{p,i} H_i, \quad \hat{\boldsymbol{x}}_{0|-1} = 0, \; i \geq 0} \tag{7.19}$$

which shows that in finding the innovations, we actually also have a complete recursion for the state-estimator $\{\hat{\boldsymbol{x}}_{i|i-1}\}$.

7.4 COMPUTING THE GAIN MATRIX

We still need to evaluate $K_{p,i}$ and $R_{e,i}$. To do so, we introduce the state-estimation error $\tilde{\boldsymbol{x}}_{i|i-1} = \boldsymbol{x}_i - \hat{\boldsymbol{x}}_{i|i-1}$, and denote its covariance matrix by

$$\boxed{P_{i|i-1} \triangleq \mathsf{E}\,\tilde{\boldsymbol{x}}_{i|i-1}\tilde{\boldsymbol{x}}^*_{i|i-1}} \tag{7.20}$$

Then, as we are going to see, the $\{K_{p,i}, R_{e,i}\}$ can be expressed in terms of $P_{i|i-1}$ and, in addition, the evaluation of $P_{i|i-1}$ will require propagating a so-called Riccati recursion.

To see this, note first that

$$\boldsymbol{e}_i = \boldsymbol{y}_i - H_i\hat{\boldsymbol{x}}_{i|i-1} = H_i\boldsymbol{x}_i - H_i\hat{\boldsymbol{x}}_{i|i-1} + \boldsymbol{v}_i = H_i\tilde{\boldsymbol{x}}_{i|i-1} + \boldsymbol{v}_i \tag{7.21}$$

Moreover, $\boldsymbol{v}_i \perp \tilde{\boldsymbol{x}}_{i|i-1}$. This is because $\tilde{\boldsymbol{x}}_{i|i-1}$ is a linear combination of the variables $\{\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{i-1}, \boldsymbol{x}_0, \boldsymbol{n}_0, \ldots, \boldsymbol{n}_{i-1}\}$, all of which are uncorrelated with $\boldsymbol{v}_i$. This claim follows from the definition $\tilde{\boldsymbol{x}}_{i|i-1} = \boldsymbol{x}_i - \hat{\boldsymbol{x}}_{i|i-1}$ and from the fact that $\hat{\boldsymbol{x}}_{i|i-1}$ is a linear combination of $\{\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{i-1}\}$ and $\boldsymbol{x}_i$ is a linear combination of $\{\boldsymbol{x}_0, \boldsymbol{n}_0, \ldots, \boldsymbol{n}_{i-1}\}$. Therefore, we get

$$\boxed{R_{e,i} = \mathsf{E}\,\boldsymbol{e}_i\boldsymbol{e}_i^* = R_i + H_iP_{i|i-1}H_i^*} \tag{7.22}$$

GAIN MATRIX

Likewise, since

$$\mathsf{E}\,\boldsymbol{x}_{i+1}\boldsymbol{e}_i^* = F_i\left(\mathsf{E}\,\boldsymbol{x}_i\boldsymbol{e}_i^*\right) + G_i\left(\mathsf{E}\,\boldsymbol{n}_i\boldsymbol{e}_i^*\right) \tag{7.23}$$

with the terms $\mathsf{E}\,\boldsymbol{x}_i\boldsymbol{e}_i^*$ and $\mathsf{E}\,\boldsymbol{n}_i\boldsymbol{e}_i^*$ given by

$$\begin{aligned}
\mathsf{E}\,\boldsymbol{x}_i\boldsymbol{e}_i^* &= \mathsf{E}\,\left(\hat{\boldsymbol{x}}_{i|i-1} + \tilde{\boldsymbol{x}}_{i|i-1}\right)\boldsymbol{e}_i^* \\
&= \mathsf{E}\,\tilde{\boldsymbol{x}}_{i|i-1}\boldsymbol{e}_i^*, \quad \text{since } \ \boldsymbol{e}_i \perp \hat{\boldsymbol{x}}_{i|i-1} \\
&= \mathsf{E}\,\tilde{\boldsymbol{x}}_{i|i-1}(H_i\tilde{\boldsymbol{x}}_{i|i-1} + \boldsymbol{v}_i)^* \\
&= \mathsf{E}\,\tilde{\boldsymbol{x}}_{i|i-1}(H_i\tilde{\boldsymbol{x}}_{i|i-1} + 0, \quad \text{since } \ \boldsymbol{v}_i \perp \tilde{\boldsymbol{x}}_{i|i-1} \\
&= P_{i|i-1}H_i^*
\end{aligned}$$

and

$$\begin{aligned}
\mathsf{E}\,\boldsymbol{n}_i\boldsymbol{e}_i^* &= \mathsf{E}\,\boldsymbol{n}_i(H_i\tilde{\boldsymbol{x}}_{i|i-1} + \boldsymbol{v}_i)^* \\
&= 0 + \mathsf{E}\,\boldsymbol{n}_i\boldsymbol{v}_i^*, \quad \text{since } \ \boldsymbol{n}_i \perp \tilde{\boldsymbol{x}}_{i|i-1} \\
&= S_i
\end{aligned}$$

we get

$$\boxed{K_{p,i} = \left(\mathsf{E}\,\boldsymbol{x}_{i+1}\boldsymbol{e}_i^*\right)R_{e,i}^{-1} = (F_iP_{i|i-1}H_i^* + G_iS_i)R_{e,i}^{-1}} \tag{7.24}$$

RICCATI RECURSION

7.5 RICCATI RECURSION

Jacopo F. Riccati
(1676-1754)

Since $\boldsymbol{n}_i \perp \boldsymbol{x}_i$, it can be easily seen from $\boldsymbol{x}_{i+1} = F_i\boldsymbol{x}_i + G_i\boldsymbol{n}_i$ that the covariance matrix of $\boldsymbol{x}_i$ obeys the recursion

$$\Pi_{i+1} = F_i\Pi_iF_i^* + G_iQ_iG_i^*, \quad \Pi_i \triangleq \mathsf{E}\,\boldsymbol{x}_i\boldsymbol{x}_i^* \tag{7.25}$$

Likewise, since $\boldsymbol{e}_i \perp \hat{\boldsymbol{x}}_{i|i-1}$, it can be seen from $\hat{\boldsymbol{x}}_{i+1|i} = F_i\hat{\boldsymbol{x}}_{i|i-1} + K_{p,i}\boldsymbol{e}_i$ that the covariance matrix of $\hat{\boldsymbol{x}}_{i|i-1}$ satisfies the recursion

$$\Sigma_{i+1} = F_i\Sigma_iF_i^* + K_{p,i}R_{e,i}K_{p,i}^*, \quad \Sigma_i \triangleq \mathsf{E}\,\hat{\boldsymbol{x}}_{i|i-1}\hat{\boldsymbol{x}}_{i|i-1}^* \tag{7.26}$$

with initial condition $\Sigma_0 = 0$. Now the orthogonal decomposition

$$\boldsymbol{x}_i = \hat{\boldsymbol{x}}_{i|i-1} + \tilde{\boldsymbol{x}}_{i|i-1} \quad \text{with} \quad \hat{\boldsymbol{x}}_{i|i-1} \perp \tilde{\boldsymbol{x}}_{i|i-1}$$

shows that $\Pi_i = \Sigma_i + P_{i|i-1}$. It is then immediate to conclude that the matrix $P_{i+1|i} = \Pi_{i+1} - \Sigma_{i+1}$ satisfies the recursion

$$\boxed{P_{i+1|i} = F_iP_{i|i-1}F_i^* + G_iQ_iG_i^* - K_{p,i}R_{e,i}K_{p,i}^*, \quad P_{0|-1} = \Pi_0} \tag{7.27}$$

7.6 COVARIANCE FORM

In summary, we arrive at the following statement of the Kalman filter, also known as the covariance form of the filter.

Algorithm 7.1 (The Kalman filter) Given observations $\{\boldsymbol{y}_i\}$ that satisfy the state-space model (7.8)–(7.10), the innovations process $\{\boldsymbol{e}_i\}$ can be recursively computed as follows. Start with $\hat{\boldsymbol{x}}_{0|-1} = 0$, $P_{0|-1} = \Pi_0$, and repeat for $i \geq 0$:

$$
\begin{aligned}
R_{e,i} &= R_i + H_i P_{i|i-1} H_i^* \\
K_{p,i} &= (F_i P_{i|i-1} H_i^* + G_i S_i) R_{e,i}^{-1} \\
\boldsymbol{e}_i &= \boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_{i|i-1} \\
\hat{\boldsymbol{x}}_{i+1|i} &= F_i \hat{\boldsymbol{x}}_{i|i-1} + K_{p,i} \boldsymbol{e}_i \\
P_{i+1|i} &= F_i P_{i|i-1} F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*
\end{aligned}
$$

TIME AND MEASUREMENT-UPDATES

Algorithm 7.2 (Time- and measurement-update forms) Given observations $\{\boldsymbol{y}_i\}$ that satisfy the state-space model (7.8), (7.9), and (7.10), the innovations process $\{\boldsymbol{e}_i\}$ can be recursively computed as follows. Start with $\hat{\boldsymbol{x}}_{0|-1} = 0$, $P_{0|-1} = \Pi_0$, and repeat for $i \geq 0$:

$$\begin{aligned}
R_{e,i} &= R_i + H_i P_{i|i-1} H_i^* \\
K_{f,i} &= P_{i|i-1} H_i^* R_{e,i}^{-1} \\
\boldsymbol{e}_i &= \boldsymbol{y}_i - H_i \hat{\boldsymbol{x}}_{i|i-1} \\
\hat{\boldsymbol{x}}_{i|i} &= \hat{\boldsymbol{x}}_{i|i-1} + K_{f,i} \boldsymbol{e}_i \\
\hat{\boldsymbol{x}}_{i+1|i} &= F_i \hat{\boldsymbol{x}}_{i|i} + G_i S_i R_{e,i}^{-1} \boldsymbol{e}_i \\
P_{i|i} &= P_{i|i-1} - P_{i|i-1} H_i^* R_{e,i}^{-1} H_i P_{i|i-1} \\
P_{i+1|i} &= F_i P_{i|i} F_i^* + G_i (Q_i - S_i R_{e,i}^{-1} S_i^*) G_i^* - F_i K_{f,i} S_i^* G_i^* - G_i S_i K_{f,i}^* F_i^*
\end{aligned}$$

EXAMPLE

(From this point onwards, material is from handout on Kalman Filtering)

Consider the first-order model

$$\begin{aligned} \boldsymbol{x}(n+1) &= \frac{1}{4}\boldsymbol{x}(n) + \boldsymbol{u}(n) \\ \boldsymbol{y}(n) &= \frac{1}{2}\boldsymbol{x}(n) + \boldsymbol{v}(n), \quad n \geq 0 \end{aligned} \tag{36.68}$$

Comparing with the standard model (36.29)–(36.30), we find that all variables are now scalars (and we writing $\boldsymbol{x}(n)$ instead of $\boldsymbol{x}_n$ to emphasize that the variable is scalar). The model coefficients are time-invariant and given by

$$F = \frac{1}{4}, \;\; G = 1, \;\; H = \frac{1}{2}, \;\; Q = 1, \;\; R = \frac{1}{2}, \;\; S = 0, \;\; \Pi_o = 1 \tag{36.69}$$

EXAMPLE

Writing down the Kalman recursions we get the following. Start with $\widehat{\boldsymbol{x}}(0|-1)=0$, $p(0|-1)=1$, and repeat for $n\geq 0$:

$$
\begin{aligned}
r_e(n) &= \frac{1}{2}+\frac{1}{4}p(n|n-1)\\
k_p(n) &= \frac{\frac{1}{8}p(n|n-1)}{\frac{1}{2}+\frac{1}{4}p(n|n-1)}\\
\boldsymbol{e}(n) &= \boldsymbol{y}(n)-\frac{1}{2}\hat{\boldsymbol{x}}(n|n-1)\\
\widehat{\boldsymbol{x}}(n+1|n) &= \frac{1}{4}\widehat{\boldsymbol{x}}(n|n-1)+k_p(n)\boldsymbol{e}(n)\\
&= \left[\frac{1}{4}-\frac{k_p(n)}{2}\right]\widehat{\boldsymbol{x}}(n|n-1)\ +\ k_p(n)\boldsymbol{y}(n)\\
p(n+1|n) &= \frac{1}{16}p(n|n-1)\ +\ 1\ -\ \frac{\frac{1}{64}p^2(n|n-1)}{\frac{1}{2}+\frac{1}{4}p(n|n-1)}
\end{aligned}
$$

EXAMPLE

In particular, these recursions allow us to evaluate the predictors of $\boldsymbol{y}(n)$ given all prior observations from time 0 up to and including time $n-1$:

$$\widehat{\boldsymbol{y}}(n|n-1) = \frac{1}{2}\widehat{\boldsymbol{x}}(n|n-1)$$

Substituting into (36.70d) we get

$$2\widehat{\boldsymbol{y}}(n+1|n) \;=\; \left[\frac{1}{2} - k_p(n)\right]\widehat{\boldsymbol{y}}(n|n-1) \;+\; k_p(n)\boldsymbol{y}(n)$$

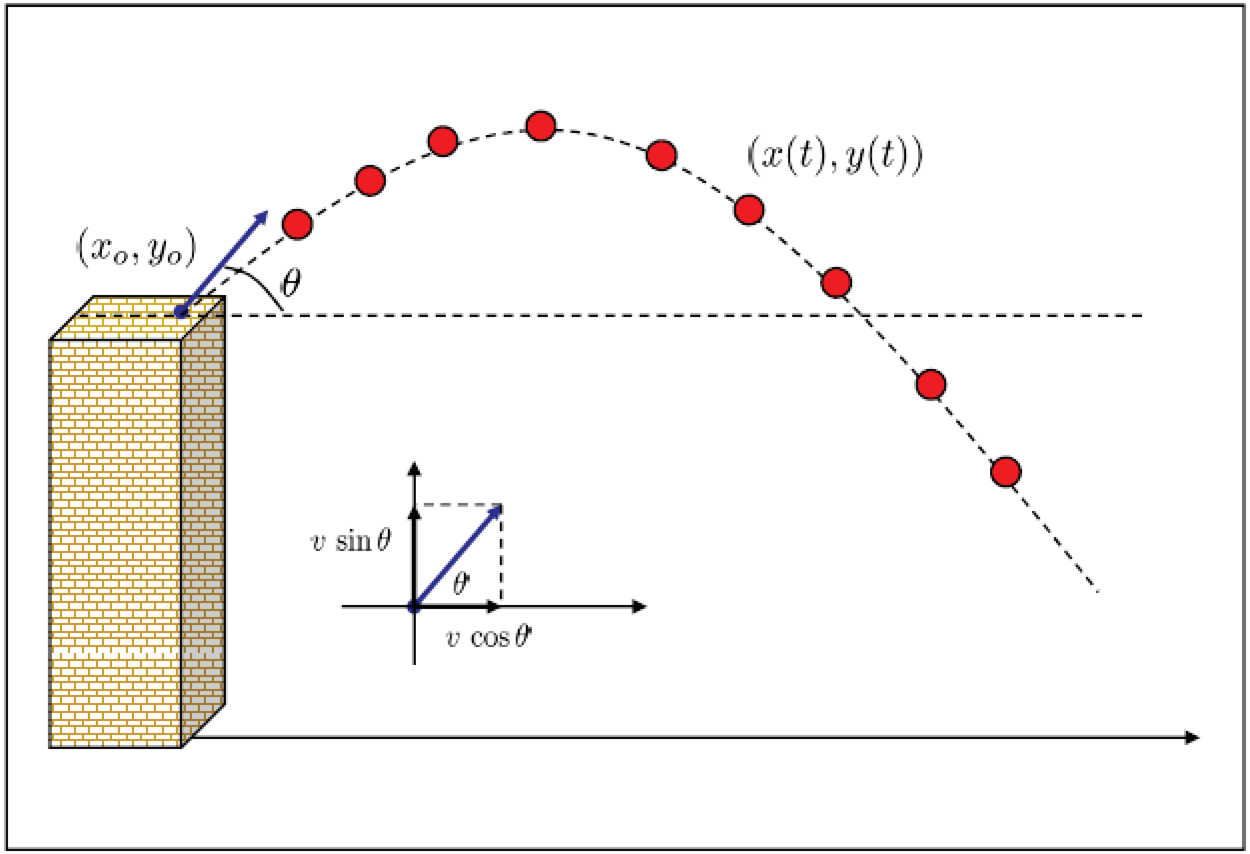
or, delaying by one time unit,

$$\boxed{\widehat{\boldsymbol{y}}(n|n-1) \;=\; \left[\frac{1}{4} - \frac{k_p(n-1)}{2}\right]\widehat{\boldsymbol{y}}(n-1|n-2) \;+\; \frac{k_p(n-1)}{2}\,\boldsymbol{y}(n-1)}$$

This is a first-order difference recursion with time-variant coefficients: the input is $\boldsymbol{y}(n-1)$ and the output is $\widehat{\boldsymbol{y}}(n|n-1)$.

$\diamond$

EXAMPLE: TRACKING A TARGET



EXAMPLE: TRACKING A TARGET

We consider a simplified model and assume the target is moving within the plane. The target is launched from location (x_o, y_o) at an angle θ with the horizontal axis at an initial speed v. The initial velocity components along the horizontal and vertical directions are therefore

$$v_x(0) = v\cos\theta, \qquad v_y(0) = v\sin\theta \tag{36.197}$$

The motion of the object is governed by Newton's equations; the acceleration along the vertical direction is downwards and its magnitude is given by $g \approx 10$ m/s^2

$$v_x(t) = v\cos\theta, \quad t \geq 0$$

$$v_y(t) = v\sin\theta - gt, \quad t \geq 0$$

$$\frac{dx(t)}{dt} = v_x(t), \qquad \frac{dy(t)}{dt} = v_y(t)$$

EXAMPLE: TRACKING A TARGET

We sample the equations of motion every T units of time and write

$$
\begin{aligned}
v_x(n) &\stackrel{\Delta}{=} v_x(nT) = v\cos\theta \\
v_y(n) &\stackrel{\Delta}{=} v_y(nT) = v\sin\theta - ngT \\
x(n+1) &= x(n) + Tv_x(n) \\
y(n+1) &= y(n) + Tv_y(n)
\end{aligned}
$$

EXAMPLE: TRACKING A TARGET

$$\underbrace{\begin{bmatrix} x(n+1) \\ y(n+1) \\ v_x(n+1) \\ v_y(n+1) \end{bmatrix}}_{\boldsymbol{x}_{n+1}} = \underbrace{\begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} x(n) \\ y(n) \\ v_x(n) \\ v_y(n) \end{bmatrix}}_{\boldsymbol{x}_n} - \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} gT}_{d_n}$$

$$\boldsymbol{z}_n = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{H} \begin{bmatrix} x(n) \\ y(n) \\ v_x(n) \\ v_y(n) \end{bmatrix} + \boldsymbol{w}_n$$

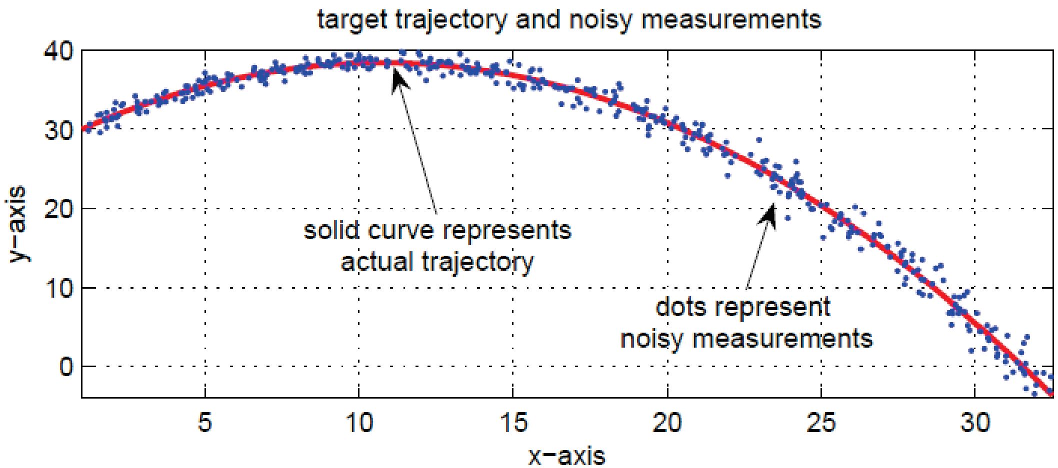
EXAMPLE: TRACKING A TARGET

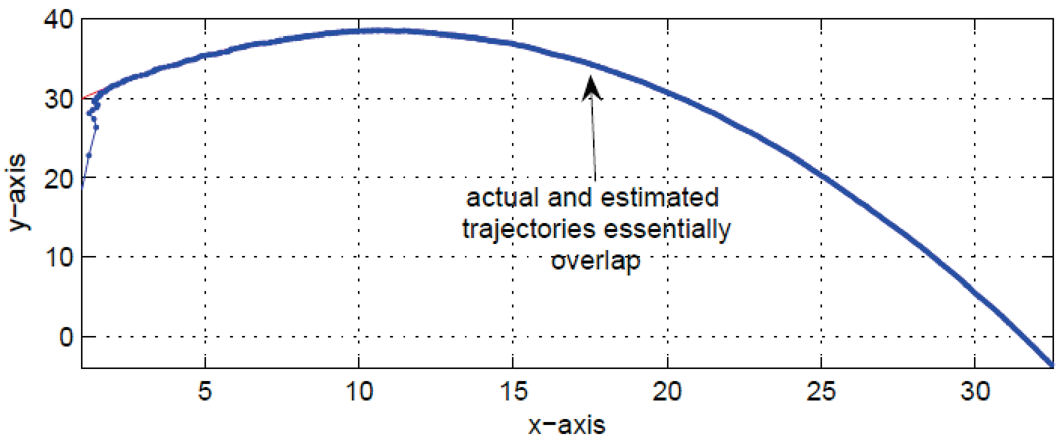
We start with $\widehat{\boldsymbol{x}}_{0|-1} = 0,\ P_{0|-1} = \Pi_0$

$$
\begin{aligned}
R_{e,n} &= R + HP_{n|n-1}H^* \\
K_{p,n} &= FP_{n|n-1}H^*R_{e,n}^{-1} \\
\boldsymbol{e}_n &= \boldsymbol{z}_n - H\widehat{\boldsymbol{x}}_{n|n-1} \\
\widehat{\boldsymbol{x}}_{n+1|n} &= F\widehat{\boldsymbol{x}}_{n|n-1} + K_{p,n}\boldsymbol{e}_n + d_n \\
P_{n+1|n} &= FP_{n|n-1}F^* - K_{p,n}R_{e,n}K_{p,n}^*
\end{aligned}
$$

EXAMPLE: TRACKING A TARGET

$$\Pi_o = I, \quad R = \begin{bmatrix} 0.3 & \\ & 0.3 \end{bmatrix}, \quad (x_o, y_o) = (1, 30), \quad v = 15, \quad T = 0.01, \quad \theta = 60^o$$





MODELING AND WHITENING FILTERS

Whitening Filter:

$$\begin{aligned} \widehat{\boldsymbol{x}}_{n+1|n} &= [F_n - K_{p,n}H_n]\,\widehat{\boldsymbol{x}}_{n|n-1} + K_{p,n}\,\boldsymbol{y}_n \\ \boldsymbol{e}_n &= -H_n\,\widehat{\boldsymbol{x}}_{n|n-1} + \boldsymbol{y}_n \end{aligned}$$

This model has $\boldsymbol{y}_n$ as input and $\boldsymbol{e}_n$ as output. Since $\boldsymbol{e}_n$ is an uncorrelated sequence with variance matrix $R_{e,n}$, we call the above model a whitening filter: it shows how to decorrelate the input sequence $\boldsymbol{y}_n$.

Modeling Filter:

$$\begin{aligned} \widehat{\boldsymbol{x}}_{n+1|n} &= F_n\,\widehat{\boldsymbol{x}}_{n|n-1} + K_{p,n}\,\boldsymbol{e}_n \\ \boldsymbol{y}_n &= H_n\widehat{\boldsymbol{x}}_{n|n-1} + \boldsymbol{e}_n \end{aligned}$$

This model has $\boldsymbol{e}_n$ as input and $\boldsymbol{y}_n$ as output. Since $\boldsymbol{e}_n$ is an uncorrelated sequence, we therefore say that the above filter is a modeling filter: it shows how to generate the sequence $\boldsymbol{y}_n$ from the uncorrelated input sequence, $\boldsymbol{e}_n$.

STEADY-STATE FILTER

Assume now that the state-space model is time-invariant

$$\begin{aligned} \boldsymbol{x}_{n+1} &= F\boldsymbol{x}_n + G\boldsymbol{u}_n, \quad n > -\infty \\ \boldsymbol{y}_n &= H\boldsymbol{x}_n + \boldsymbol{v}_n \end{aligned}$$

with

$$\mathbb{E}\begin{bmatrix} \boldsymbol{u}_n \\ \boldsymbol{v}_n \\ \boldsymbol{x}_0 \\ 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{u}_m \\ \boldsymbol{v}_m \\ \boldsymbol{x}_0 \end{bmatrix}^* = \begin{bmatrix} \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix}\delta_{nm} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \Pi_0 \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix}$$

It follows from the above state-space model that the transfer matrix function from $\boldsymbol{u}_n$ to $\boldsymbol{y}_n$ is given by

$$H_{uy}(z) = H(zI - F)^{-1}G$$

while the transfer function from $\boldsymbol{v}_n$ to $\boldsymbol{y}_n$ is $H_{vy}(z) = 1$. Assuming F is a stable matrix (meaning all its eigenvalues are inside the unit circle), then $H_{uy}(z)$ will be a BIBO stable mapping. Since the processes $\{\boldsymbol{u}_n, \boldsymbol{v}_n\}$ are wide-sense stationary, it follows that $\{\boldsymbol{y}_n\}$ will also be wide sense stationary process.

STEADY-STATE FILTER

Solutions of DARE

Now, the Kalman recursions that correspond to the above model are given by:

$$
\begin{aligned}
R_{e,n} &= R + HP_{n|n-1}H^* \\
K_{p,n} &= (FP_{n|n-1}H^* + GS)R_{e,n}^{-1} \\
\boldsymbol{e}_n &= \boldsymbol{y}_n - H\hat{\boldsymbol{x}}_{n|n-1} \\
\widehat{\boldsymbol{x}}_{n+1|n} &= F\widehat{\boldsymbol{x}}_{n|n-1} + K_{p,n}\boldsymbol{e}_n \\
P_{n+1|n} &= FP_{n|n-1}F^* + GQG^* - K_{p,n}R_{e,n}K_{p,n}^*
\end{aligned}
$$

STEADY-STATE FILTER

$$P_{n+1|n} \quad = \quad FP_{n|n-1}F^* + GQG^* - K_{p,n}R_{e,n}K_{p,n}^*$$

There are conditions under which the Riccati recursion can be shown to converge to a unique positive-definite matrix P. Discussion of these technical conditions is beyond the scope of this text. It suffices to consider here the case when $S = 0$ and F is a stable matrix. In this case, the Riccati recursion can be shown to converge, as $n \to \infty$, to a unique positive-definite matrix P that satisfies the so-called Discrete Algebraic Riccati Equation (DARE):

$$P = FPF^* + GQG^* - K_pR_eK_p^*$$

where

$$\begin{aligned} R_e &= R + HPH^* \\ K_p &= FPH^*R_e^{-1} \end{aligned}$$

Moreover, the resulting closed-loop matrix $F - K_pH$ will also be stable (i.e., will have all its eigenvalues inside the unit circle).

SPECTRAL FACTORIZATION

Let us now focus on the case in which the sequences $\{\boldsymbol{y}_n, \boldsymbol{v}_n\}$ are scalar sequences, replaced by $\{\boldsymbol{y}(n), \boldsymbol{v}(n)\}$; the discussion can be easily extended to vector processes but it is sufficient for our purposes to study scalar-valued output processes. In this situation, the H matrix becomes a row vector, say h^{T}, and the covariance matrix R becomes a scalar, say r. Additionally, the innovations process $\boldsymbol{e}_n$ becomes scalar-valued, say, $\boldsymbol{e}(n)$, with variance r_e in steady-state, and the gain matrix K_p becomes a column vector, k_p. Writing down the resulting modeling filter (36.84a)–(36.84b) in steady-state, as $n \to \infty$, we have

$$\begin{aligned} \widehat{\boldsymbol{x}}_{n+1|n} &= F\,\widehat{\boldsymbol{x}}_{n|n-1} \,+\, k_p\,\boldsymbol{e}(n) \\ \boldsymbol{y}(n) &= h^{\mathsf{T}}\widehat{\boldsymbol{x}}_{n|n-1} + \boldsymbol{e}(n) \end{aligned}$$

SPECTRAL FACTORIZATION

The transfer function from $\boldsymbol{e}(n)$ to $\boldsymbol{y}(n)$ is the following causal modeling filter $L(z)$:

$$L(z) \;=\; 1 + h^{\mathsf{T}}(zI - F)^{-1}k_p$$

This is a stable transfer function since its poles are given by the eigenvalues of F and these eigenvalues lie inside the unit disc. Accordingly, $\boldsymbol{y}(n)$ is a wide-sense stationary process and its $z-$spectrum is given by

$$S_y(z) \;=\; r_e\, L(z)\left[L\left(\frac{1}{z^*}\right)\right]^*$$

We therefore find that the steady-state Kalman equations allow us to determine the canonical spectral factor, $L(z)$, of the process $\{\boldsymbol{y}(n)\}$ (since it allows us to determine k_p).