2 Optimal Estimators

This chapter addresses the question of how the state $\mathbf{x}_{k+m}$ of a dynamic system can be estimated from an input sequence (u_k) and a measured output sequence (y_k). Depending on the value of m, the estimation process is referred to as

(1) Smoothing for $m < 0$

(2) Filtering for $m = 0$, or

(3) Prediction for $m > 0$

Figure 2.1 provides a graphical illustration of these three cases.

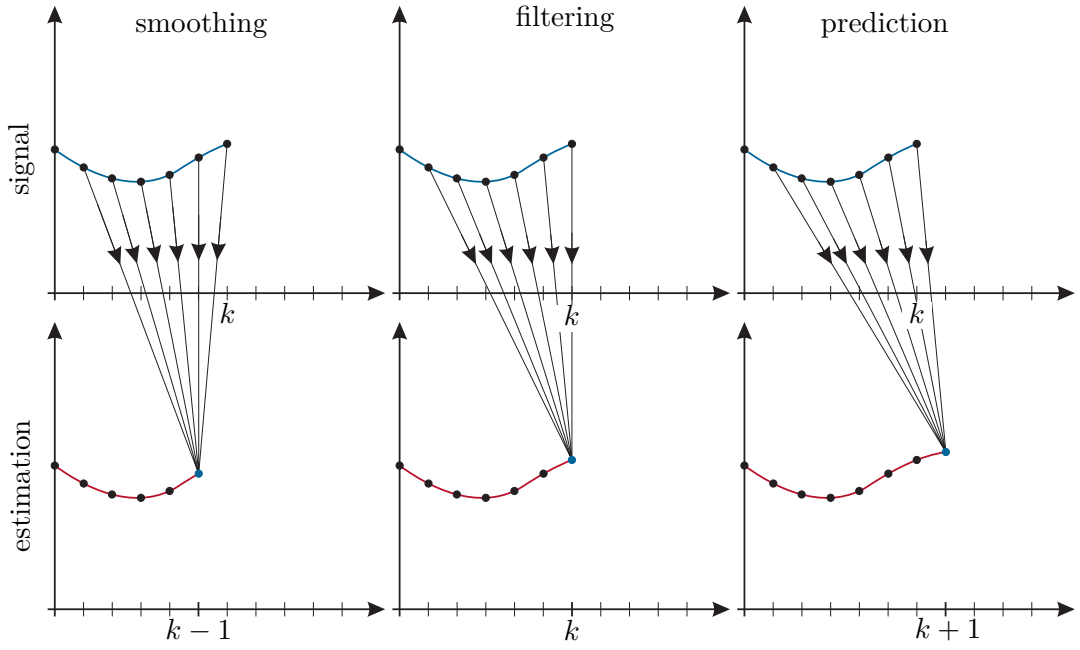


Figure 2.1: Concerning the terms smoothing, filtering, and prediction.

Further considerations will be limited to the last case (3), as this is the most interesting for control engineering applications. As a result of the following considerations, an optimal state observer, the so-called *Kalman filter*, will be determined, which minimizes a quadratic performance criterion. However, this requires an intermediate step of extending the results of the least-squares estimation from the previous chapter. Since the expected value and the covariance of random numbers are used repeatedly in this chapter, these two concepts will be explained for normally distributed and uniformly distributed random numbers.

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

Note 2.1 (Normally Distributed Random Variables). A scalar normally distributed random variable x is defined by the probability density function (Gaussian distribution)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}, \tag{2.1}$$

with the mean (expected value) m and the variance σ^2. The mean (expected value, first moment) and the variance (second central moment) are, as explained in Appendix A, given by

$$\mathrm{E}(x) = m = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x \tag{2.2a}$$

$$\mathrm{E}((x - \mathrm{E}(x))^2) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mathrm{E}(x))^2 f(x)\,\mathrm{d}x \tag{2.2b}$$

Figure 2.2 shows the probability density function for different parameterizations of a normally distributed random number.

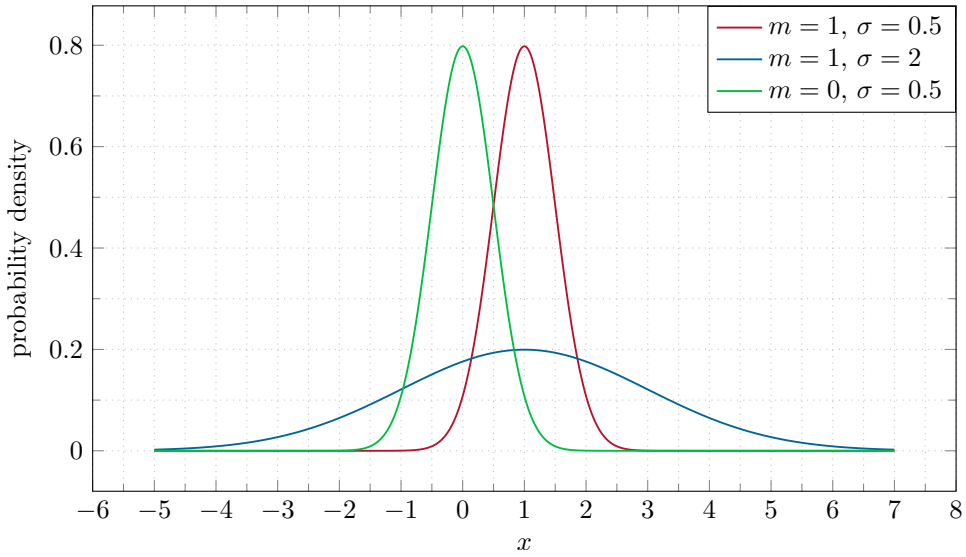


Figure 2.2: Probability density function for different normally distributed random variables.

The probability that a random number x lies in the interval δ around the expected value $\mathrm{E}(x) = m$ is calculated as

$$\mathrm{P}(m-\delta < x \leq m+\delta) = \int_{m-\delta}^{m+\delta} f(x)\,\mathrm{d}x = \mathrm{erf}\left(\frac{\delta}{\sqrt{2}\sigma}\right), \tag{2.3}$$

cf. Exercise A.1 in Appendix A. For example, if $\delta = \sigma$, a random number x lies in this interval with a probability of 0.68. Furthermore, a random number x lies in the interval $\delta = 2\sigma$ with a probability of 0.95. The variance thus represents a measure of the scatter of the random numbers around the expected value. The joint probability density function $f(\mathbf{x})$ of an n-dimensional normally distributed random vector $\mathbf{x}$ is calculated in the form

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathbf{Q})}} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{m})}, \tag{2.4}$$

with the expected value $\mathbf{m}$ and the covariance matrix $\mathbf{Q}$,

$$\mathbf{m} = \mathrm{E}(\mathbf{x}) \tag{2.5a}$$

$$\mathbf{Q} = \mathrm{E}((\mathbf{x}-\mathrm{E}(\mathbf{x}))(\mathbf{x}-\mathrm{E}(\mathbf{x}))^{\mathrm{T}})\ . \tag{2.5b}$$

For the special case $n = 2$, the probability that a random vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ lies in an ellipse of the form

$$(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{m}) = C^2 \tag{2.6}$$

is given by

$$\mathrm{P} = 1 - \mathrm{e}^{-\frac{C^2}{2}} \tag{2.7}$$

The proof of this statement can be found, for example, in [2.1].

Thus, with the help of the covariance matrix $\mathbf{Q}$, the region (i.e., the ellipses for $n = 2$) can be determined in which a random vector $\mathbf{x}$ lies with a probability P. Figure 2.3 shows the distribution of 3000 random vectors $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathrm{T}}$, with the normally distributed, uncorrelated random variables x_1 ($\mathrm{E}(x_1) = m_1 = 1$, $\sigma_1 = 2$) and x_2 ($\mathrm{E}(x_2) = m_2 = 2$, $\sigma_2 = 1$). Furthermore, the ellipses for $\mathrm{P} = 0.5$ and $\mathrm{P} = 0.95$, and the expected value $\mathrm{E}(\mathbf{x}) = \mathbf{m}$ are shown.

This consideration for $n = 2$ can be generalized to n-dimensional normally distributed random variables. In this case, the n-dimensional ellipsoid $(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{m}) = C^2$ describes a measure of the distribution of the random vector. The probability P that a random vector $\mathbf{x}$ lies in this ellipsoid is given by

$$1 - \mathrm{P} = \frac{n}{2^{n/2}\Gamma\left(\frac{n}{2}+1\right)} \int_C^{\infty} \xi^{n-1}\mathrm{e}^{-\frac{\xi^2}{2}}\,\mathrm{d}\xi, \tag{2.8}$$

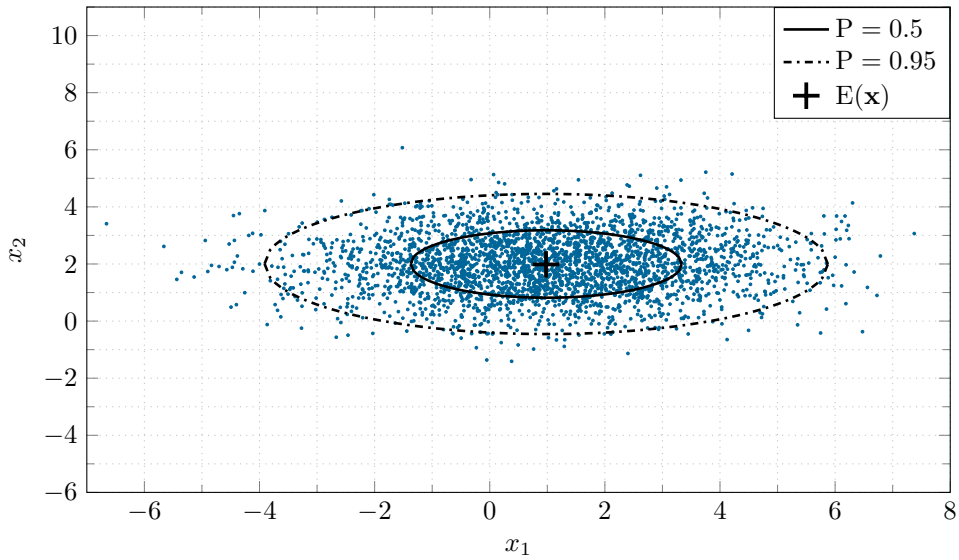


Figure 2.3: Graphical representation of the expected value $\mathrm{E}(\mathbf{x})$ and the covariance matrix $\mathbf{Q}$ for normally distributed random vectors $\mathbf{x}$.

with the Gamma function Γ, see [2.1].

An exact relationship between the ellipsoids defined by the covariance matrix $\mathbf{Q}$ and the probability that a random vector $\mathbf{x}$ lies in this region is only defined for normally distributed random variables. For other distributions, these ellipsoids are only a more or less accurate approximation. Nevertheless, the covariance matrix $\mathbf{Q}$ is also a meaningful measure for estimating the distribution of the random vectors for these distributions, which is why the corresponding ellipses or ellipsoids are often displayed here as well.

Note 2.2. A scalar random variable x uniformly distributed in the interval $[a, b]$ is defined by the probability density function

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

The corresponding expected value m and variance σ^2 are calculated as

$$\mathrm{E}(x) = m = \frac{a+b}{2} \tag{2.10a}$$

$$\mathrm{E}((x - \mathrm{E}(x))^2) = \sigma^2 = \frac{1}{12}(b-a)^2 \ . \tag{2.10b}$$

Figure 2.4 shows the probability density functions for different uniformly distributed random variables.

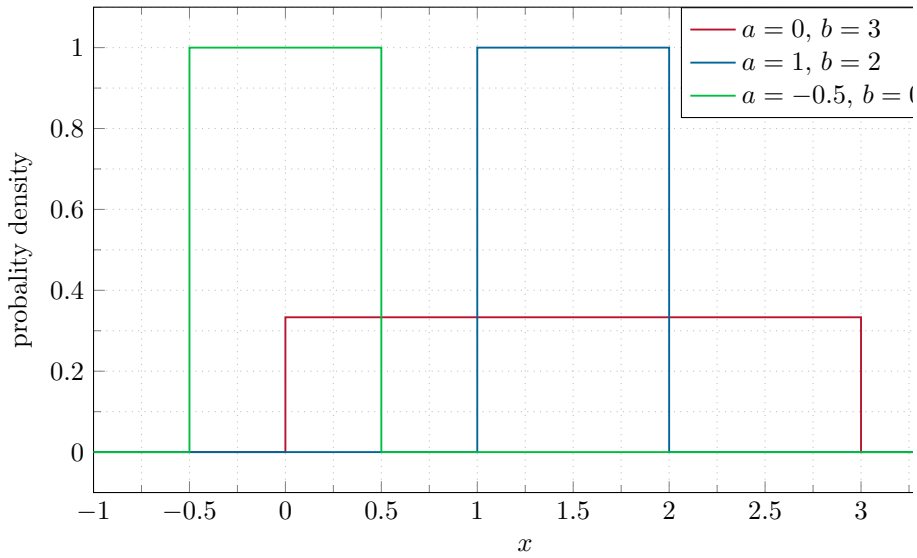


Figure 2.4: Probability density function for different uniformly distributed random variables.

2.1 Gauss-Markov Estimation

Consider the overdetermined linear equation system of (1.60) (compare also (1.87)-(1.90)) augmented by the stochastic disturbance $\mathbf{v}$

$$\mathbf{y} = \mathbf{S}\mathbf{p} + \mathbf{v} \tag{2.11}$$

with the known $(m \times n)$-matrix $\mathbf{S} \in \mathbb{R}^{m \times n}$, the n-dimensional vector of unknowns $\mathbf{p} \in \mathbb{R}^n$, and the m-dimensional measurement vector $\mathbf{y} \in \mathbb{R}^m$. We now assume that the stochastic disturbance $\mathbf{v}$ has the following stochastic properties:

$$\mathrm{E}(\mathbf{v}) = \mathbf{0} \quad \text{and} \quad \mathrm{cov}(\mathbf{v}) = \mathrm{E}(\mathbf{v}\mathbf{v}^{\mathrm{T}}) = \mathbf{Q} \quad \text{with} \quad \mathbf{Q} > 0 \ . \tag{2.12}$$

It should be noted here that $\mathbf{v}$ can also be interpreted as a measurement error. We are now looking for a *linear estimator* of the form

$$\hat{\mathbf{p}} = \mathbf{K}\mathbf{y} \tag{2.13}$$

with a constant $(n \times m)$-matrix $\mathbf{K} \in \mathbb{R}^{n \times m}$. Since $\mathbf{y}$ is the sum of a constant vector $\mathbf{S}\mathbf{p}$ and a stochastic vector (vector with stochastic entries) $\mathbf{v}$, $\mathbf{y}$, the estimated parameter $\hat{\mathbf{p}}$ according to (2.13), and the parameter error $\mathbf{e} = \mathbf{p} - \hat{\mathbf{p}}$ are stochastic quantities. Therefore, it does not make sense to determine the matrix $\mathbf{K}$ such that $\|\mathbf{e}\|_2^2$ is minimized, but the task

$$\min_{\mathbf{K}} \mathrm{E}(\|\mathbf{e}\|_2^2) = \min_{\mathbf{K}} \mathrm{E}((\mathbf{p} - \mathbf{K}\mathbf{y})^{\mathrm{T}}(\mathbf{p} - \mathbf{K}\mathbf{y})) \tag{2.14}$$

must be solved.

If we now substitute the relationship (2.11) into (2.14), then we obtain, taking into

account (2.12) and the results from Appendix A (see Exercise A.3):

$$\min_{\mathbf{K}} \mathrm{E}((\mathbf{p}-\mathbf{KSp}-\mathbf{Kv})^{\mathrm{T}}(\mathbf{p}-\mathbf{KSp}-\mathbf{Kv})) =$$

$$\min_{\mathbf{K}} \left\{ \mathrm{E}((\mathbf{p}-\mathbf{KSp})^{\mathrm{T}}(\mathbf{p}-\mathbf{KSp})) - 2\underbrace{\mathrm{E}((\mathbf{p}-\mathbf{KSp})^{\mathrm{T}}\mathbf{Kv})}_{=0} + \underbrace{\mathrm{E}(\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\mathbf{Kv})}_{=\mathrm{tr}(\mathrm{E}(\mathbf{Kvv}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}))} \right\} =$$

$$\min_{\mathbf{K}} \left\{ \|\mathbf{p}-\mathbf{KSp}\|_2^2 + \mathrm{tr}(\mathbf{KQK}^{\mathrm{T}}) \right\} . \tag{2.15}$$

The matrix $\mathbf{K}$ that minimizes the expression (2.15) is obviously a function of the unknown parameter vector $\mathbf{p}$. Therefore, the solution of this minimization problem is also not suitable for providing an estimator for $\mathbf{p}$ according to (2.13). To circumvent this problem, a substitute problem is solved in the following: It can be seen that with the choice

$$\mathbf{KS} = \mathbf{E} \tag{2.16}$$

and $\mathbf{E}$ as the identity matrix, the solution of the minimization problem (2.15) is independent of the parameter $\mathbf{p}$. This *constraint* (2.16) may seem arbitrary at first glance, but if we calculate the expected value of $\hat{\mathbf{p}}$ of the linear estimator (2.13), then it follows that

$$\mathrm{E}(\hat{\mathbf{p}}) = \mathrm{E}(\mathbf{Ky}) = \mathrm{E}(\mathbf{KSp}+\mathbf{Kv}) = \mathbf{KS}\underbrace{\mathrm{E}(\mathbf{p})}_{=\mathbf{p}} + \mathbf{K}\underbrace{\mathrm{E}(\mathbf{v})}_{=\mathbf{0}} . \tag{2.17}$$

I.e., the constraint (2.16) implies that $\mathrm{E}(\hat{\mathbf{p}}) = \mathbf{p}$ and thus the estimator is *unbiased*.

The problem

$$\min_{\mathbf{K}} \left\{ \mathrm{tr}(\mathbf{KQK}^{\mathrm{T}}) \right\} \quad \text{subject to} \quad \mathbf{KS} = \mathbf{E} \tag{2.18}$$

is now equivalent to solving n separate optimization problems of the form

$$\min_{\mathbf{k}_j} \mathbf{k}_j^{\mathrm{T}}\mathbf{Q}\mathbf{k}_j \quad \text{subject to} \quad \mathbf{k}_j^{\mathrm{T}}\mathbf{S} = \mathbf{e}_j^{\mathrm{T}} \quad \text{for} \quad j = 1,\ldots,n . \tag{2.19}$$

To show this, write the matrix $\mathbf{K}$ and the identity matrix $\mathbf{E}$ in the form

$$\mathbf{K} = \begin{bmatrix} \mathbf{k}_1^{\mathrm{T}} \\ \mathbf{k}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{k}_n^{\mathrm{T}} \end{bmatrix} \quad \text{and} \quad \mathbf{E} = \begin{bmatrix} \mathbf{e}_1^{\mathrm{T}} \\ \mathbf{e}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{e}_n^{\mathrm{T}} \end{bmatrix} \tag{2.20}$$

and substitute this into (2.18):

$$\min_{\mathbf{K}} \left\{ \mathrm{tr}(\mathbf{KQK}^{\mathrm{T}}) \right\} = \min_{\mathbf{K}} \left\{ \sum_{j=1}^{n} \mathbf{k}_j^{\mathrm{T}}\mathbf{Q}\mathbf{k}_j \right\} \quad \text{subject to} \quad \mathbf{k}_j^{\mathrm{T}}\mathbf{S} = \mathbf{e}_j^{\mathrm{T}} \tag{2.21}$$

for $j = 1,\ldots,n$. Since the j-th summand in (2.21) only depends on $\mathbf{k}_j$, the minimization problem of (2.21) can be replaced by n minimization problems according to (2.19).

For the following, the solution of a problem of type (2.19) for a fixed j is therefore of interest. If we now consider the Hilbert space $\mathcal{H} = \mathbb{R}^m$ with the inner product $\langle \mathbf{x}, \mathbf{z} \rangle = \mathbf{x}^{\mathrm{T}}\mathbf{z} = \sum_{j=1}^{m} x_j z_j$, then we see that the column vectors $\mathbf{k}_k$, $k = 1, \ldots, n$, of the matrix $\mathbf{K}$, i.e., $\operatorname{span}\{\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_n\}$, which satisfy the constraints of (2.19), do not form a subspace of the Hilbert space $\mathcal{H}$ and thus the projection theorem of Theorem 1.1 is not directly applicable.

> *Exercise* 2.1. Show that the sum $\mathbf{k}_j + \mathbf{k}_k$ does not satisfy the constraint of (2.21), even if $\mathbf{k}_j$ and $\mathbf{k}_k$ individually satisfy this constraint.

2.1.1 Quadratic Minimization with Affine Constraints

To solve the above problem, an extension of the projection theorem of Theorem 1.1 is required:

Theorem 2.1 (Extension of the Projection Theorem)**.** *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{U}$ a closed subspace of $\mathcal{H}$. The translational shift of $\mathcal{U}$ in the form $\mathcal{A} = \mathbf{x} + \mathcal{U}$ for a fixed $\mathbf{x} \in \mathcal{H}$ is called a* linear variety *or* affine subspace*. Then there exists a unique vector $\mathbf{x}_0 \in \mathcal{A}$ of minimal norm, and this is orthogonal to $\mathcal{U}$ (see Figure 2.5).*

> *Proof of Theorem 2.1.* Shift the affine subspace $\mathcal{A}$ by $-\mathbf{x}$ so that it becomes a closed subspace and then apply the projection theorem of Theorem 1.1. Note that the optimal solution $\mathbf{x}_0$ is *not* orthogonal to the affine subspace $\mathcal{A}$ but orthogonal to $\mathcal{U}$. □

Before the problem of (2.19) can be solved, some theoretical foundations should be explained:

Definition 2.1 (Orthogonal Complement)**.** If $\mathcal{U}$ is a subspace of a Hilbert space $\mathcal{H}$ with the inner product $\langle \cdot, \cdot \rangle$, then the set of all vectors orthogonal to $\mathcal{U}$ is called the *orthogonal complement of* $\mathcal{U}$, and we write $\mathcal{U}^{\perp}$ for it.

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

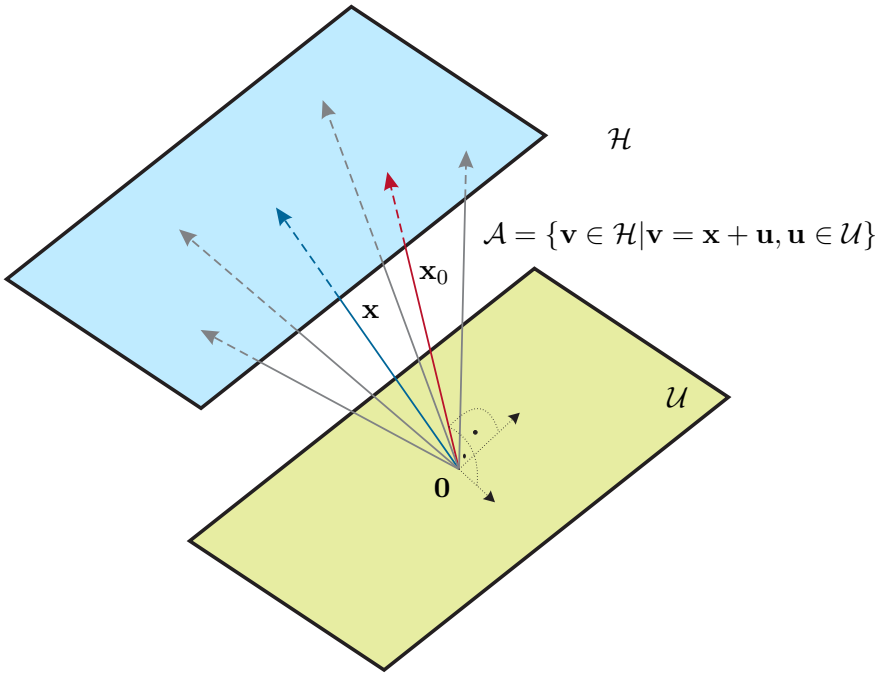


Figure 2.5: On the projection theorem for affine subspaces.

For the subspaces $\mathcal{U}$ and $\mathcal{V}$ of a Hilbert space, the following properties now hold:

(1) The orthogonal complement $\mathcal{U}^\perp$ is a closed subspace,

(2) $\mathcal{U} \subseteq \mathcal{U}^{\perp\perp}$,

(3) If $\mathcal{U} \subset \mathcal{V}$, then $\mathcal{V}^\perp \subset \mathcal{U}^\perp$,

(4) $\mathcal{U}^{\perp\perp\perp} = \mathcal{U}^\perp$, and

(5) $\mathcal{U}^{\perp\perp}$ is the smallest closed subspace containing $\mathcal{U}$.

Proof of (1). Since the linear combination of orthogonal vectors is again orthogonal, it follows immediately that $\mathcal{U}^\perp$ is a subspace. The closedness of $\mathcal{U}^\perp$ follows from the fact that, due to the continuity of the inner product $\langle \cdot, \cdot \rangle$, for the limit $\mathbf{x}$ of a convergent sequence $(\mathbf{x}_k)$ in $\mathcal{U}^\perp$, we have

$$\mathbf{0} = \langle \mathbf{y}, \mathbf{x}_k \rangle = \langle \mathbf{y}, \mathbf{x} \rangle \tag{2.22}$$

for all $\mathbf{y} \in \mathcal{U}$ and thus also $\mathbf{x} \in \mathcal{U}^\perp$. □

Exercise 2.2. Prove the above properties (2) to (4).

Definition 2.2 (Direct Sum)**.** A vector space $\mathcal{X}$ is called the direct sum of two subspaces $\mathcal{U}$ and $\mathcal{V}$, and we write $\mathcal{X} = \mathcal{U} \oplus \mathcal{V}$, if every vector $\mathbf{x} \in \mathcal{X}$ can be *uniquely* represented as the sum $\mathbf{x} = \mathbf{u} + \mathbf{v}$ with $\mathbf{u} \in \mathcal{U}$ and $\mathbf{v} \in \mathcal{V}$.

Without proof, the following theorem holds as a corollary of the projection theorem of Theorem 1.1:

Theorem 2.2. *If $\mathcal{U}$ is a closed linear subspace of a Hilbert space $\mathcal{H}$, then $\mathcal{H} = \mathcal{U} \oplus \mathcal{U}^\perp$ and $\mathcal{U} = \mathcal{U}^{\perp\perp}$.*

The following theorem can now be given for the solution of the minimization problem with affine constraints from (2.19):

Theorem 2.3 (Minimization with Affine Constraints)**.** *Let $\mathcal{H}$ be a Hilbert space with the linearly independent vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$. Among all possible vectors $\mathbf{x} \in \mathcal{H}$ that satisfy the affine equation system*

$$\begin{aligned} \langle \mathbf{x}, \mathbf{s}_1 \rangle &= c_1 \\ \langle \mathbf{x}, \mathbf{s}_2 \rangle &= c_2 \\ &\vdots \\ \langle \mathbf{x}, \mathbf{s}_n \rangle &= c_n \end{aligned} \tag{2.23}$$

with the constant coefficients $c_1, c_2, \ldots, c_n$, let the vector $\mathbf{x}_0$ have the minimal norm. Then $\mathbf{x}_0$ can be written in the form

$$\mathbf{x}_0 = \sum_{j=1}^{n} p_{0,j} \mathbf{s}_j = \mathbf{S}\mathbf{p}_0 \tag{2.24}$$

where $\mathbf{p}_0^T = \begin{bmatrix} p_{0,1} & p_{0,2} & \ldots & p_{0,n} \end{bmatrix}$ is calculated from the relationship

$$\underbrace{\begin{bmatrix} \langle \mathbf{s}_1, \mathbf{s}_1 \rangle & \cdots & \langle \mathbf{s}_n, \mathbf{s}_1 \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{s}_1, \mathbf{s}_n \rangle & \cdots & \langle \mathbf{s}_n, \mathbf{s}_n \rangle \end{bmatrix}}_{\mathbf{G}=\mathbf{S}^\mathrm{T}\mathbf{S}} \underbrace{\begin{bmatrix} p_{0,1} \\ \vdots \\ p_{0,n} \end{bmatrix}}_{\mathbf{p}_0} = \underbrace{\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}}_{\mathbf{c}} \tag{2.25}$$

with the Gramian matrix $\mathbf{G}$.

Proof of Theorem 2.3. Let $\mathcal{U} = \operatorname{span}\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ be a closed linear subspace of the Hilbert space $\mathcal{H}$. The set of all possible vectors $\mathbf{x} \in \mathcal{H}$ that satisfy the affine equation system (2.23) form an affine subspace of $\mathcal{H}$, namely the translational shift of $\mathcal{U}^\perp$ by a vector $\boldsymbol{\gamma}$. Since the orthogonal complement $\mathcal{U}^\perp$ is a closed subspace, we can apply Theorem 2.1 and thus know that the optimal solution $\mathbf{x}_0$ exists, is unique, and is orthogonal to $\mathcal{U}^\perp$. However, this implies $\mathbf{x}_0 \in \mathcal{U}^{\perp\perp}$, and due to the closedness of $\mathcal{U}$ and Theorem 2.2, we obtain $\mathcal{U}^{\perp\perp} = \mathcal{U}$. Since $\mathbf{x}_0 \in \mathcal{U}$ holds, $\mathbf{x}_0$ must be representable

as a linear combination of the $\mathbf{s}_j$, $j = 1, \ldots, n$, according to (2.24). Substituting (2.24) into the affine constraint (2.23), we obtain the result (2.25). □

Applying Theorem 2.3 to the minimization problem (2.19) with $\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \ldots & \mathbf{s}_n \end{bmatrix}$ and the inner product $\langle \mathbf{x}, \mathbf{z} \rangle_Q = \mathbf{x}^{\mathrm{T}} \mathbf{Q} \mathbf{z}$ in the Hilbert space $\mathcal{H} = \mathbb{R}^m$, i.e.,

$$\min_{\mathbf{k}_j} \langle \mathbf{k}_j, \mathbf{k}_j \rangle_Q \quad \text{subject to} \quad \left\langle \mathbf{k}_j, \mathbf{Q}^{-1} \mathbf{s}_l \right\rangle_Q = \delta_{jl} = \begin{cases} 1 & \text{for } j = l \\ 0 & \text{otherwise} \end{cases} \tag{2.26}$$

for $j = 1, \ldots, n$, then, replacing $\mathbf{s}_l$ with $\mathbf{Q}^{-1}\mathbf{s}_l$ in (2.24) and (2.25), we obtain the result

$$\mathbf{k}_{j,0} = \mathbf{Q}^{-1}\mathbf{S}\mathbf{p}_0 \quad \text{and} \quad (\mathbf{Q}^{-1}\mathbf{S})^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{S}\mathbf{p}_0 = \mathbf{e}_j \tag{2.27}$$

or

$$\mathbf{k}_{j,0} = \mathbf{Q}^{-1}\mathbf{S}(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{e}_j \ . \tag{2.28}$$

According to (2.20), the optimal solution $\mathbf{K}_0$ for the matrix $\mathbf{K}$ is calculated as

$$\mathbf{K}_0^{\mathrm{T}} = \mathbf{Q}^{-1}\mathbf{S}(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S})^{-1} \tag{2.29}$$

and thus the desired linear *Gauss-Markov estimator* according to (2.13) is

$$\hat{\mathbf{p}} = (\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{y} \ . \tag{2.30}$$

Comparing (2.30) with (1.128), we see that the result is identical to the result of the least-squares identification with weighted least squares with the weighting matrix $\mathbf{Q}^{-1}$, where $\mathbf{Q} = \mathrm{cov}(\mathbf{v})$ holds.

For the expected value of the estimation error $\mathbf{e} = \mathbf{p} - \hat{\mathbf{p}}$, we obtain $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ according to (2.17), and the covariance matrix of the estimation error is given by (2.16):

$$\begin{aligned} \mathrm{E}(\mathbf{e}\mathbf{e}^{\mathrm{T}}) = \mathrm{E}((\mathbf{p} - \mathbf{K}\mathbf{y})(\mathbf{p} - \mathbf{K}\mathbf{y})^{\mathrm{T}}) \quad &= \mathrm{E}\Big([\mathbf{p} - \mathbf{K}(\mathbf{S}\mathbf{p} + \mathbf{v})][\mathbf{p} - \mathbf{K}(\mathbf{S}\mathbf{p} + \mathbf{v})]^{\mathrm{T}}\Big) \\ = \mathrm{E}\Big([\mathbf{K}\mathbf{v}][\mathbf{K}\mathbf{v}]^{\mathrm{T}}\Big) &= \mathbf{K}\mathbf{Q}\mathbf{K}^{\mathrm{T}} \end{aligned} \tag{2.31}$$

or, with (2.29),

$$\mathrm{E}(\mathbf{e}\mathbf{e}^{\mathrm{T}}) = (\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{S}(\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S})^{-1} = (\mathbf{S}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{S})^{-1} \ . \tag{2.32}$$

In the literature, the linear estimator (2.30) is often also called *BLUE (best linear unbiased estimate)*.

If the covariance matrix $\mathrm{cov}(\mathbf{v}) = \mathrm{E}(\mathbf{v}\mathbf{v}^{\mathrm{T}}) = \mathbf{Q}$ is a diagonal matrix of the form $\mathbf{Q} = \mathrm{diag}(q_0, q_1, \ldots)$ with $q_j > 0$ for all $j \geq 0$, then for the Gauss-Markov estimator (2.30), according to the recursive method of weighted least squares (1.121), with $\alpha_j = 1/q_j$, a recursive version can be immediately given. This also clarifies the question of what the optimal choice of the sequence of positive weighting coefficients α_j ($\alpha_j > 0$ for all j) looks like in (1.121).

2.2 Minimum-Variance Estimation

It has been assumed so far that no information is available about the n-dimensional vector of unknowns $\mathbf{p}$ in (2.11). However, in some cases, *a priori information about* $\mathbf{p}$ in the form of stochastic characteristics (expected value, covariance matrix) is known. Therefore, it is assumed that for the system of equations (compare also (2.11))

$$\mathbf{y} = \mathbf{S}\mathbf{p} + \mathbf{v} \tag{2.33}$$

with the stochastic perturbation $\mathbf{v}$, the known $(m \times n)$-matrix $\mathbf{S} \in \mathbb{R}^{m \times n}$, the n-dimensional random vector $\mathbf{p} \in \mathbb{R}^n$, and the m-dimensional measurement vector $\mathbf{y} \in \mathbb{R}^m$, the following holds:

$$\begin{aligned} \mathrm{E}(\mathbf{v}) = \mathbf{0}, \qquad & \mathrm{cov}(\mathbf{v}) = \mathrm{E}\Big(\mathbf{v}\mathbf{v}^{\mathrm{T}}\Big) = \mathbf{Q} \qquad \text{with} \quad \mathbf{Q} \geq 0 \\ \mathrm{E}(\mathbf{p}) = \mathbf{0}, \qquad & \mathrm{cov}(\mathbf{p}) = \mathrm{E}\Big(\mathbf{p}\mathbf{p}^{\mathrm{T}}\Big) = \mathbf{R} \qquad \text{with} \quad \mathbf{R} \geq 0 \\ & \mathrm{E}\Big(\mathbf{p}\mathbf{v}^{\mathrm{T}}\Big) = \mathbf{N} \ . \end{aligned} \tag{2.34}$$

Furthermore, it is assumed that the matrix $\Big(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}} + \mathbf{Q} + \mathbf{S}\mathbf{N} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\Big)$ is non-singular. A linear estimator is sought

$$\hat{\mathbf{p}} = \mathbf{K}\mathbf{y} \tag{2.35}$$

with a constant $(n \times m)$-matrix $\mathbf{K} \in \mathbb{R}^{n \times m}$ such that the following minimization problem

$$\min_{\mathbf{K}} \mathrm{E}\Big(\|\mathbf{e}\|_2^2\Big) = \min_{\mathbf{K}} \mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{y}]^{\mathrm{T}}[\mathbf{p} - \mathbf{K}\mathbf{y}]\Big) \tag{2.36}$$

is solved. By expanding (2.36) and using the relationship $\mathrm{spur}(\mathbf{KSR}) = \mathrm{spur}\Big(\mathbf{R}(\mathbf{KS})^{\mathrm{T}}\Big)$ we obtain

$$\begin{aligned} &\min_{\mathbf{K}} \mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{y}]^{\mathrm{T}}[\mathbf{p} - \mathbf{K}\mathbf{y}]\Big) = \\ &\min_{\mathbf{K}} \left\{ \underbrace{\mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{S}\mathbf{p}]^{\mathrm{T}}[\mathbf{p} - \mathbf{K}\mathbf{S}\mathbf{p}]\Big)}_{\mathrm{spur}\left(\mathrm{E}\left([\mathbf{E}-\mathbf{K}\mathbf{S}]\mathbf{p}\mathbf{p}^{\mathrm{T}}[\mathbf{E}-\mathbf{K}\mathbf{S}]^{\mathrm{T}}\right)\right)} - 2 \underbrace{\mathrm{E}\Big([\mathbf{p} - \mathbf{K}\mathbf{S}\mathbf{p}]^{\mathrm{T}}\mathbf{K}\mathbf{v}\Big)}_{\mathrm{spur}\left(\mathrm{E}\left([\mathbf{E}-\mathbf{K}\mathbf{S}]\mathbf{p}\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)\right)} + \underbrace{\mathrm{E}\Big(\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\mathbf{K}\mathbf{v}\Big)}_{\mathrm{spur}\left(\mathrm{E}\left(\mathbf{K}\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)\right)} \right\} = \\ &\min_{\mathbf{K}} \Big\{ \mathrm{spur}\Big([\mathbf{E} - \mathbf{K}\mathbf{S}]\mathbf{R}[\mathbf{E} - \mathbf{K}\mathbf{S}]^{\mathrm{T}} - 2\mathbf{K}\mathbf{N}^{\mathrm{T}} - \mathbf{K}\Big[\mathbf{S}\mathbf{N} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\Big]\mathbf{K}^{\mathrm{T}} + \mathbf{K}\mathbf{Q}\mathbf{K}^{\mathrm{T}} \Big) \Big\} = \\ &\min_{\mathbf{K}} \Big\{ \mathrm{spur}\Big(\mathbf{K}\Big(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}} + \mathbf{Q} + \mathbf{S}\mathbf{N} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\Big)\mathbf{K}^{\mathrm{T}} - 2\mathbf{K}\Big(\mathbf{S}\mathbf{R} + \mathbf{N}^{\mathrm{T}}\Big) \Big) \Big\} \ . \end{aligned} \tag{2.37}$$

Writing the matrix $\mathbf{K}$ and the identity matrix $\mathbf{E}$ as in (2.20)

$$\mathbf{K} = \begin{bmatrix} \mathbf{k}_1^{\mathrm{T}} \\ \mathbf{k}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{k}_n^{\mathrm{T}} \end{bmatrix} \quad \text{and} \quad \mathbf{E} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix} \tag{2.38}$$

then (2.37) becomes

$$\min_{\mathbf{K}}\left\{\sum_{j=1}^{n}\left(\mathbf{k}_j^{\mathrm{T}}\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)\mathbf{k}_j-2\mathbf{k}_j^{\mathrm{T}}\left(\mathbf{S}\mathbf{R}+\mathbf{N}^{\mathrm{T}}\right)\mathbf{e}_j\right)\right\}. \tag{2.39}$$

Comparing the minimization problem (2.39) with that of (1.61), i.e.,

$$\min_{\mathbf{p}}(\mathbf{y}-\mathbf{S}\mathbf{p})^{\mathrm{T}}(\mathbf{y}-\mathbf{S}\mathbf{p})=\min_{\mathbf{p}}\left(\mathbf{y}^{\mathrm{T}}\mathbf{y}-2\mathbf{p}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{y}+\mathbf{p}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\mathbf{p}\right), \tag{2.40}$$

it is seen that the two problems are equivalent. Thus, the solution of (2.39) can be directly given by the solution of (2.40) (compare (1.63))

$$\mathbf{p}_0=\left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{y} \tag{2.41}$$

by setting $\mathbf{S}^{\mathrm{T}}\mathbf{S}=\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}$ and $\mathbf{S}^{\mathrm{T}}\mathbf{y}=\left(\mathbf{S}\mathbf{R}+\mathbf{N}^{\mathrm{T}}\right)\mathbf{e}_j$ in (2.41). This yields the optimal solution $\mathbf{K}_0$ for the matrix $\mathbf{K}$ of (2.35) as

$$\mathbf{K}_0^{\mathrm{T}}=\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{S}\mathbf{R}+\mathbf{N}^{\mathrm{T}}\right) \tag{2.42}$$

and the sought linear *minimum-variance estimator* according to (2.35) is

$$\hat{\mathbf{p}}=\left(\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{N}\right)\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\mathbf{y}\,. \tag{2.43}$$

Note that at the beginning of this section it was assumed that $\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)$ is non-singular.

The covariance matrix of the estimation error is calculated as

$$\begin{aligned}
\mathrm{cov}(\mathbf{e}) &= \mathrm{E}\left([\mathbf{p}-\mathbf{K}(\mathbf{S}\mathbf{p}+\mathbf{v})][\mathbf{p}-\mathbf{K}(\mathbf{S}\mathbf{p}+\mathbf{v})]^{\mathrm{T}}\right)=\mathrm{E}\left([\mathbf{E}-\mathbf{K}\mathbf{S}]\mathbf{p}\mathbf{p}^{\mathrm{T}}[\mathbf{E}-\mathbf{K}\mathbf{S}]^{\mathrm{T}}\right)-\\
&\quad \mathrm{E}\left(\mathbf{K}\mathbf{v}\mathbf{p}^{\mathrm{T}}\left(\mathbf{E}-\mathbf{S}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)+(\mathbf{E}-\mathbf{K}\mathbf{S})\mathbf{p}\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)+\mathrm{E}\left(\mathbf{K}\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)\\
&= (\mathbf{E}-\mathbf{K}\mathbf{S})\mathbf{R}(\mathbf{E}-\mathbf{K}\mathbf{S})^{\mathrm{T}}-\mathbf{K}\mathbf{N}^{\mathrm{T}}\left(\mathbf{E}-\mathbf{S}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)-(\mathbf{E}-\mathbf{K}\mathbf{S})\mathbf{N}\mathbf{K}^{\mathrm{T}}+\mathbf{K}\mathbf{Q}\mathbf{K}^{\mathrm{T}}\\
&= \mathbf{R}-\mathbf{K}\left(\mathbf{S}\mathbf{R}+\mathbf{N}^{\mathrm{T}}\right)-\left(\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{N}\right)\mathbf{K}^{\mathrm{T}}+\mathbf{K}\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)\mathbf{K}^{\mathrm{T}}\\
&= \mathbf{R}-\left(\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{N}\right)\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{S}\mathbf{R}+\mathbf{N}^{\mathrm{T}}\right)-\\
&\quad \left(\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{N}\right)\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{S}\mathbf{R}+\mathbf{N}^{\mathrm{T}}\right)+\left(\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{N}\right)\\
&\quad \left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)\\
&\quad \left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{S}\mathbf{R}+\mathbf{N}^{\mathrm{T}}\right)\\
&= \mathbf{R}-\left(\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{N}\right)\left(\mathbf{S}\mathbf{R}\mathbf{S}^{\mathrm{T}}+\mathbf{Q}+\mathbf{S}\mathbf{N}+\mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\right)^{-1}\left(\mathbf{S}\mathbf{R}+\mathbf{N}^{\mathrm{T}}\right).
\end{aligned} \tag{2.44}$$

Exercise 2.3. If you have the time, verify (2.44).

In contrast to the Gauss-Markov estimator (2.30), the minimum-variance estimator (2.43) yields meaningful results even if fewer measurements m than unknowns n are available, provided that the matrix $\mathbf{SRS}^{\mathrm{T}} + \mathbf{Q} + \mathbf{SN} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}$ is non-singular. The reason for this property is obviously that, in the minimum-variance estimator, stochastic information about the parameter vector $\mathbf{p}$ is available, and thus a meaningful estimation is possible with fewer measurements, or even with no measurements for $\mathbf{Q} > 0$.

Applying now the matrix inversion lemma, Theorem 1.3,

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\Big(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}\Big)^{-1}\mathbf{DA}^{-1} \tag{2.45}$$

to the covariance matrix of the estimation error $\mathrm{cov}(\mathbf{e})$ from (2.44) with $\mathbf{A} = \mathbf{R}^{-1}$, $\mathbf{B} = \mathbf{S}^{\mathrm{T}} + \mathbf{R}^{-1}\mathbf{N}$, $\mathbf{C} = \Big(\mathbf{Q} - \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{N}\Big)^{-1}$ and $\mathbf{D} = \mathbf{S} + \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}$, we obtain

$$\begin{aligned}\mathrm{cov}(\mathbf{e}) &= \mathbf{R} - \Big(\mathbf{RS}^{\mathrm{T}} + \mathbf{N}\Big)\Big(\mathbf{SRS}^{\mathrm{T}} + \mathbf{Q} + \mathbf{SN} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\Big)^{-1}\Big(\mathbf{SR} + \mathbf{N}^{\mathrm{T}}\Big) \\ &= \Big(\mathbf{R}^{-1} + \Big(\mathbf{S}^{\mathrm{T}} + \mathbf{R}^{-1}\mathbf{N}\Big)\Big(\mathbf{Q} - \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{N}\Big)^{-1}\Big(\mathbf{S} + \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}\Big)\Big)^{-1}.\end{aligned} \tag{2.46}$$

Furthermore, it is easy to verify that the minimum-variance estimator (2.43) can be written in the form

$$\hat{\mathbf{p}} = \mathrm{cov}(\mathbf{e})\Big(\mathbf{S}^{\mathrm{T}} + \mathbf{R}^{-1}\mathbf{N}\Big)\Big(\mathbf{Q} - \mathbf{N}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{N}\Big)^{-1}\mathbf{y}. \tag{2.47}$$

Exercise 2.4. Show the validity of (2.47).

Equation (2.47) now shows that for $\mathbf{R}^{-1} = \mathbf{0}$, i.e., infinitely high variance of the parameter vector $\mathbf{p}$ – thus no meaningful a priori information for $\mathbf{p}$ is available – and $\mathbf{Q} > 0$, the minimum-variance estimator (2.43) or (2.47) transitions into the Gauss-Markov estimator (2.30).

The results (2.43) and (2.44) in combination with the relationships

$$\mathrm{E}\Big(\mathbf{py}^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{pp}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}} + \mathbf{pv}^{\mathrm{T}}\Big) = \Big(\mathbf{RS}^{\mathrm{T}} + \mathbf{N}\Big) \tag{2.48}$$

and

$$\mathrm{E}\Big(\mathbf{yy}^{\mathrm{T}}\Big) = \mathrm{E}\Big([\mathbf{Sp} + \mathbf{v}][\mathbf{Sp} + \mathbf{v}]^{\mathrm{T}}\Big) = \Big(\mathbf{SRS}^{\mathrm{T}} + \mathbf{Q} + \mathbf{SN} + \mathbf{N}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\Big) \tag{2.49}$$

can be summarized in the following theorem:

Theorem 2.4 (Minimum-Variance Estimator)**.** *For the system of equations* (2.33)

$$\mathbf{y} = \mathbf{Sp} + \mathbf{v} \tag{2.50}$$

with the stochastic quantities $\mathbf{p}$*,* $\mathbf{v}$ *and* $\mathbf{y}$*, it is assumed that* $\mathrm{E}\Big(\mathbf{yy}^{\mathrm{T}}\Big)$ *is invertible. The optimal linear estimate* $\hat{\mathbf{p}}$ *of* $\mathbf{p}$*, which minimizes the expected value of the quadratic*

error $\mathrm{E}\Big([\mathbf{p}-\hat{\mathbf{p}}]^{\mathrm{T}}[\mathbf{p}-\hat{\mathbf{p}}]\Big)$, *is given by*

$$\hat{\mathbf{p}} = \mathrm{E}\Big(\mathbf{p}\mathbf{y}^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\mathbf{y}\mathbf{y}^{\mathrm{T}}\Big)\Big]^{-1}\mathbf{y} \tag{2.51}$$

with the associated error covariance matrix

$$\begin{aligned}\mathrm{cov}(\mathbf{e}) &= \mathrm{E}\Big([\mathbf{p}-\hat{\mathbf{p}}][\mathbf{p}-\hat{\mathbf{p}}]^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{p}\mathbf{p}^{\mathrm{T}}\Big) - \mathrm{E}\Big(\hat{\mathbf{p}}\hat{\mathbf{p}}^{\mathrm{T}}\Big)\\ &= \mathrm{E}\Big(\mathbf{p}\mathbf{p}^{\mathrm{T}}\Big) - \mathrm{E}\Big(\mathbf{p}\mathbf{y}^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\mathbf{y}\mathbf{y}^{\mathrm{T}}\Big)\Big]^{-1}\mathrm{E}\Big(\mathbf{y}\mathbf{p}^{\mathrm{T}}\Big).\end{aligned} \tag{2.52}$$

Note the similarity of (2.51) to the optimal solution in the sense of least squares from (1.63).

Exercise 2.5. Show the validity of the relationship (2.52). Further show that

$$\mathrm{E}\Big([\mathbf{p}-\hat{\mathbf{p}}][\mathbf{p}-\hat{\mathbf{p}}]^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{p}[\mathbf{p}-\hat{\mathbf{p}}]^{\mathrm{T}}\Big) \quad \text{or} \quad \mathrm{E}\Big(\hat{\mathbf{p}}\hat{\mathbf{p}}^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{p}\hat{\mathbf{p}}^{\mathrm{T}}\Big)\ . \tag{2.53}$$

Remark: (to Exercise 2.5) Simply substitute the expression from (2.51) for $\hat{\mathbf{p}}$.

Exercise 2.6. Show that the relationships (2.43) and (2.44) can be derived directly using Theorem 2.4.

Exercise 2.7. Assume that the expected values $\mathrm{E}(\mathbf{y})$ and $\mathrm{E}(\mathbf{p})$ are not zero as in (2.34), but $\mathrm{E}(\mathbf{y}) = \mathbf{y}_0 \neq \mathbf{0}$ and $\mathrm{E}(\mathbf{p}) = \mathbf{p}_0 \neq \mathbf{0}$. Show that the minimum-variance estimate of the form

$$\hat{\mathbf{p}} = \mathbf{K}\mathbf{y} + \mathbf{b}$$

with the constant vector $\mathbf{b}$ is given by

$$\hat{\mathbf{p}} = \mathbf{p}_0 + \mathrm{E}\Big([\mathbf{p}-\mathbf{p}_0][\mathbf{y}-\mathbf{y}_0]^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big([\mathbf{y}-\mathbf{y}_0][\mathbf{y}-\mathbf{y}_0]^{\mathrm{T}}\Big)\Big]^{-1}(\mathbf{y}-\mathbf{y}_0)$$

By the *minimum-variance estimation of a linear function*

$$\mathbf{z} = \mathbf{C}\mathbf{p} \tag{2.54}$$

with the optimal estimator

$$\hat{\mathbf{z}} = \mathbf{K}_{\mathbf{z}}\mathbf{y} \tag{2.55}$$

based on the measurements

$$\mathbf{y} = \mathbf{S}\mathbf{p} + \mathbf{v} \tag{2.56}$$

we understand the solution of the minimization problem

$$\min_{\mathbf{K}_{\mathbf{z}}} \mathrm{E}\Big([\mathbf{z}-\hat{\mathbf{z}}]^{\mathrm{T}}[\mathbf{z}-\hat{\mathbf{z}}]\Big)\ . \tag{2.57}$$

The following theorem now applies:

Theorem 2.5 (Minimum-Variance Estimator of a Linear Function)**.** *The linear minimum-variance estimate* (2.55) *of a linear function* $\mathbf{Cp}$ *based on the measurements* (2.56) *is equivalent to the linear function of the minimum-variance estimate* $\hat{\mathbf{p}}$ *itself, i.e., the best estimate of* $\mathbf{Cp}$ *is* $\mathbf{C}\hat{\mathbf{p}}$*.*

Exercise 2.8. Prove Theorem 2.5.

2.2.1 Recursive Minimum-Variance Estimation

In the next step, we will investigate how the optimal estimate $\hat{\mathbf{p}}$ from (2.47) can be improved by adding new measurements. This is of particular importance for on-line applications. The method is again based on the properties of the projection theorem in a Hilbert space. If $\mathcal{U}_1$ and $\mathcal{U}_2$ denote two subspaces of a Hilbert space, then the projection of a vector $\mathbf{p}$ onto the subspace $\mathcal{U}_1 + \mathcal{U}_2$ is identical to the projection of $\mathbf{p}$ onto $\mathcal{U}_1$ plus the projection onto $\mathcal{U}_2^*$, where $\mathcal{U}_2^*$ is orthogonal to $\mathcal{U}_1$ and satisfies the relationship $\mathcal{U}_1 \oplus \mathcal{U}_2^* = \mathcal{U}_1 + \mathcal{U}_2$. If $\mathcal{U}_2$ is spanned by a finite number of vectors, then the differences between these vectors and their projections onto $\mathcal{U}_1$ span the subspace $\mathcal{U}_2^*$. Figure 2.6 illustrates this situation.

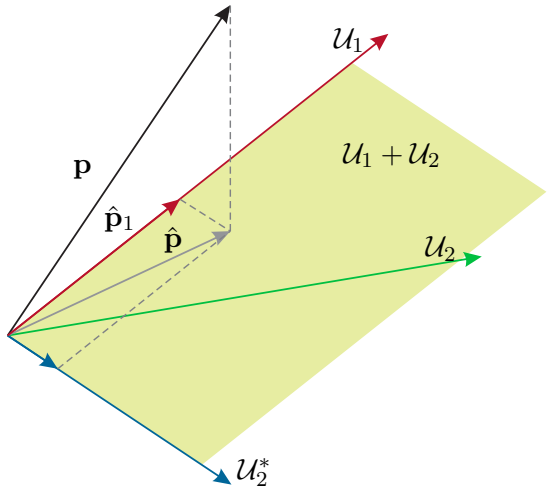


Figure 2.6: On the projection onto the sum of orthogonal subspaces.

Thus, the following theorem can be stated:

Theorem 2.6 (Recursive Minimum-Variance Estimation)**.** *Let* $\mathbf{p}$ *be a random vector of a Hilbert space* $\mathcal{H}$ *of random variables and let* $\hat{\mathbf{p}}_1$ *denote the orthogonal projection of* $\mathbf{p}$ *onto a closed subspace* $\mathcal{U}_1$ *of* $\mathcal{H}$*. According to the projection theorem,* $\hat{\mathbf{p}}_1$ *is thus the best estimate of* $\mathbf{p}$ *in* $\mathcal{U}_1$*. Furthermore, let* $\mathbf{y}_2$ *describe all those random vectors that span the subspace* $\mathcal{U}_2$ *of* $\mathcal{H}$*, and let* $\hat{\mathbf{y}}_2$ *be the orthogonal projection of* $\mathbf{y}_2$ *onto* $\mathcal{U}_1$*. According to the projection theorem,* $\hat{\mathbf{y}}_2$ *is thus the best estimate of* $\mathbf{y}_2$ *in* $\mathcal{U}_1$*. With*

$\tilde{\mathbf{y}}_2 = \mathbf{y}_2 - \hat{\mathbf{y}}_2$, *the projection* $\hat{\mathbf{p}}$ *of* $\mathbf{p}$ *onto* $\mathcal{U}_1 + \mathcal{U}_2$ *is*

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}_1 + \mathrm{E}\Big(\mathbf{p}\tilde{\mathbf{y}}_2^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\tilde{\mathbf{y}}_2\tilde{\mathbf{y}}_2^{\mathrm{T}}\Big)\Big]^{-1}\tilde{\mathbf{y}}_2 \ . \tag{2.58}$$

Thus, the best estimate $\hat{\mathbf{p}}$ *on* $\mathcal{U}_1 + \mathcal{U}_2$ *is composed of the sum of the best estimate of* $\mathbf{p}$ *on* $\mathcal{U}_1$ *(*$\hat{\mathbf{p}}_1$*) and the best estimate of* $\mathbf{p}$ *on* $\mathcal{U}_2^*$ *(the subspace generated by* $\tilde{\mathbf{y}}_2$*).*

Proof of Theorem 2.6. It is easily verified that $\mathcal{U}_1 + \mathcal{U}_2 = \mathcal{U}_1 \oplus \mathcal{U}_2^*$ and that $\mathcal{U}_2^*$ is orthogonal to $\mathcal{U}_1$. Equation (2.58) then follows from the fact that the projection onto a sum of subspaces is equal to the sum of the projections onto the individual subspaces, provided that these are orthogonal. □

The result of Theorem 2.6 can also be interpreted as follows: If $\hat{\mathbf{p}}_1$ denotes the optimal estimate based on measurements that span the subspace $\mathcal{U}_1$, then when receiving new measurements that span the subspace $\mathcal{U}_2$, only that part needs to be considered that is not yet described by the measurements in $\mathcal{U}_1$, i.e., that part of the new data that is orthogonal to the old data and thus lies in the subspace $\mathcal{U}_2^*$.

As an application example, consider a system of equations of the form of (2.33)

$$\mathbf{y}_1 = \mathbf{S}_1\mathbf{p} + \mathbf{v}_1 \ . \tag{2.59}$$

Furthermore, $\hat{\mathbf{p}}_1 = \mathrm{E}\Big(\mathbf{p}\mathbf{y}_1^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\mathbf{y}_1\mathbf{y}_1^{\mathrm{T}}\Big)\Big]^{-1}\mathbf{y}_1$ denotes the optimal minimum-variance estimate of $\mathbf{p}$ according to (2.47) or (2.51) based on $\dim(\mathbf{y}_1)$ measurements with the error covariance matrix

$$\mathrm{cov}(\mathbf{p} - \hat{\mathbf{p}}_1) = \mathrm{E}\Big([\mathbf{p} - \hat{\mathbf{p}}_1][\mathbf{p} - \hat{\mathbf{p}}_1]^{\mathrm{T}}\Big) = \mathbf{P}_1 \ . \tag{2.60}$$

The question now arises how to improve the estimate of $\mathbf{p}$ by adding new measurements

$$\mathbf{y}_2 = \mathbf{S}_2\mathbf{p} + \mathbf{v}_2 \tag{2.61}$$

For the stochastic perturbation $\mathbf{v}_2$ and the random parameter vector $\mathbf{p}$, let

$$\begin{aligned} \mathrm{E}(\mathbf{v}_2) = \mathbf{0}, \qquad & \mathrm{cov}(\mathbf{v}_2) = \mathrm{E}\Big(\mathbf{v}_2\mathbf{v}_2^{\mathrm{T}}\Big) = \mathbf{Q}_2 \quad \text{with} \quad \mathbf{Q}_2 \geq 0 \\ \mathrm{E}(\mathbf{p}) = \mathbf{0}, \qquad & \mathrm{E}\Big(\mathbf{p}\mathbf{v}_2^{\mathrm{T}}\Big) = \mathbf{N}_2 \ . \end{aligned} \tag{2.62}$$

Furthermore, it is reasonable to assume that the perturbation $\mathbf{v}_2$ is not correlated with past measurements $\mathbf{y}_1$, and thus

$$\mathrm{E}\Big(\mathbf{v}_2\mathbf{y}_1^{\mathrm{T}}\Big) = \mathbf{0} \quad \text{or} \quad \mathrm{E}\Big(\mathbf{v}_2\hat{\mathbf{p}}_1^{\mathrm{T}}\Big) = \mathbf{0} \ . \tag{2.63}$$

The best estimate $\hat{\mathbf{y}}_2$ of $\mathbf{y}_2$ based on the past measurements $\mathbf{y}_1$ is

$$\hat{\mathbf{y}}_2 = \mathbf{S}_2\hat{\mathbf{p}}_1 \ . \tag{2.64}$$

Thus, according to Theorem 2.6 with $\tilde{\mathbf{y}}_2 = \mathbf{y}_2 - \hat{\mathbf{y}}_2$, the improved estimate $\hat{\mathbf{p}}_2$ is

$$\hat{\mathbf{p}}_2 = \hat{\mathbf{p}}_1 + \mathrm{E}\Big(\mathbf{p}\tilde{\mathbf{y}}_2^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\tilde{\mathbf{y}}_2\tilde{\mathbf{y}}_2^{\mathrm{T}}\Big)\Big]^{-1}\tilde{\mathbf{y}}_2 \tag{2.65}$$

with

$$\mathrm{E}\Big(\mathbf{p}\tilde{\mathbf{y}}_2^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{p}(\mathbf{p}-\hat{\mathbf{p}}_1)^{\mathrm{T}}\mathbf{S}_2^{\mathrm{T}} + \mathbf{p}\mathbf{v}_2^{\mathrm{T}}\Big) \overset{(2.53)}{=} \mathbf{P}_1\mathbf{S}_2^{\mathrm{T}} + \mathbf{N}_2 \tag{2.66}$$

and

$$\mathrm{E}\Big(\tilde{\mathbf{y}}_2\tilde{\mathbf{y}}_2^{\mathrm{T}}\Big) = \mathrm{E}\Big([\mathbf{S}_2(\mathbf{p}-\hat{\mathbf{p}}_1)+\mathbf{v}_2][\mathbf{S}_2(\mathbf{p}-\hat{\mathbf{p}}_1)+\mathbf{v}_2]^{\mathrm{T}}\Big) = \mathbf{S}_2\mathbf{P}_1\mathbf{S}_2^{\mathrm{T}} + \mathbf{Q}_2 + \mathbf{S}_2\mathbf{N}_2 + \mathbf{N}_2^{\mathrm{T}}\mathbf{S}_2^{\mathrm{T}} \ . \tag{2.67}$$

Exercise 2.9. Show that the error covariance matrix, analogous to (2.44), can be calculated in the form

$$\begin{aligned}\mathbf{P}_2 &= \mathrm{cov}(\mathbf{p}-\hat{\mathbf{p}}_2)\\ &= \mathbf{P}_1 - \Big(\mathbf{P}_1\mathbf{S}_2^{\mathrm{T}} + \mathbf{N}_2\Big)\Big(\mathbf{S}_2\mathbf{P}_1\mathbf{S}_2^{\mathrm{T}} + \mathbf{Q}_2 + \mathbf{S}_2\mathbf{N}_2 + \mathbf{N}_2^{\mathrm{T}}\mathbf{S}_2^{\mathrm{T}}\Big)^{-1}\Big(\mathbf{S}_2\mathbf{P}_1 + \mathbf{N}_2^{\mathrm{T}}\Big)\end{aligned} \tag{2.68}$$

This yields the *recursive minimum-variance estimator*

$$\hat{\mathbf{p}}_k = \hat{\mathbf{p}}_{k-1} + \Big(\mathbf{P}_{k-1}\mathbf{S}_k^{\mathrm{T}} + \mathbf{N}_k\Big)\Big(\mathbf{S}_k\mathbf{P}_{k-1}\mathbf{S}_k^{\mathrm{T}} + \mathbf{Q}_k + \mathbf{S}_k\mathbf{N}_k + \mathbf{N}_k^{\mathrm{T}}\mathbf{S}_k^{\mathrm{T}}\Big)^{-1}(\mathbf{y}_k - \mathbf{S}_k\hat{\mathbf{p}}_{k-1}) \tag{2.69}$$

with

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \Big(\mathbf{P}_{k-1}\mathbf{S}_k^{\mathrm{T}} + \mathbf{N}_k\Big)\Big(\mathbf{S}_k\mathbf{P}_{k-1}\mathbf{S}_k^{\mathrm{T}} + \mathbf{Q}_k + \mathbf{S}_k\mathbf{N}_k + \mathbf{N}_k^{\mathrm{T}}\mathbf{S}_k^{\mathrm{T}}\Big)^{-1}\Big(\mathbf{S}_k\mathbf{P}_{k-1} + \mathbf{N}_k^{\mathrm{T}}\Big) \tag{2.70}$$

and the initial values $\mathbf{P}_{-1}$ and $\hat{\mathbf{p}}_{-1}$.

Assuming now that exactly one new measurement is added in each iteration step, i.e., the quantities y_k and v_k are scalars, then by substituting $\mathbf{S}_k = \mathbf{s}_k^{\mathrm{T}}$, $\mathbf{Q}_k = \mathrm{E}(v_k^2) = q_k$ and $\mathbf{N}_k = \mathrm{E}(\mathbf{p}v_k) = \mathbf{n}_k$ into (2.69), (2.70), the recursive minimum-variance estimator becomes

$$\mathbf{k}_k = \frac{\mathbf{P}_{k-1}\mathbf{s}_k + \mathbf{n}_k}{(q_k + 2\mathbf{s}_k^{\mathrm{T}}\mathbf{n}_k + \mathbf{s}_k^{\mathrm{T}}\mathbf{P}_{k-1}\mathbf{s}_k)} \tag{2.71a}$$

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{k}_k\Big(\mathbf{s}_k^{\mathrm{T}}\mathbf{P}_{k-1} + \mathbf{n}_k^{\mathrm{T}}\Big) \tag{2.71b}$$

$$\hat{\mathbf{p}}_k = \hat{\mathbf{p}}_{k-1} + \mathbf{k}_k\Big(y_k - \mathbf{s}_k^{\mathrm{T}}\hat{\mathbf{p}}_{k-1}\Big) \ . \tag{2.71c}$$

Note also in this context the analogy to the recursive weighted least squares method (1.121) for $q_k = 1/\alpha_k$ and $\mathbf{n}_k = \mathbf{0}$.

2.3 The Kalman Filter

Building on the previous considerations, especially the recursive minimum variance estimation, the next step is to derive the Kalman filter, an *optimal observer* in the sense of control theory. For the fundamentals of observer theory, refer to Chapter 8 of the Automation script. Numerous versions of the Kalman filter exist in the literature. In this lecture, we will initially consider a linear, time-invariant, discrete-time system of the form

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

$$\mathbf{x}_{k+1} = \mathbf{\Phi}\mathbf{x}_k + \mathbf{\Gamma}\mathbf{u}_k + \mathbf{G}\mathbf{w}_k \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.72a}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{v}_k \tag{2.72b}$$

with the n-dimensional state $\mathbf{x} \in \mathbb{R}^n$, the p-dimensional deterministic input $\mathbf{u} \in \mathbb{R}^p$, the q-dimensional output $\mathbf{y} \in \mathbb{R}^q$, the r-dimensional disturbance $\mathbf{w} \in \mathbb{R}^r$, the measurement noise $\mathbf{v}$, and the matrices $\mathbf{\Phi} \in \mathbb{R}^{n\times n}$, $\mathbf{\Gamma} \in \mathbb{R}^{n\times p}$, $\mathbf{G} \in \mathbb{R}^{n\times r}$, $\mathbf{C} \in \mathbb{R}^{q\times n}$, and $\mathbf{D} \in \mathbb{R}^{q\times p}$. It should be noted here, however, that the Kalman filter can also be designed for linear, time-variant, and continuous-time systems. The following assumptions apply:

(1) For the disturbance $\mathbf{w}$ and the measurement noise $\mathbf{v}$, it is assumed that

$$\mathrm{E}(\mathbf{v}_k) = \mathbf{0} \qquad \mathrm{E}\left(\mathbf{v}_k\mathbf{v}_j^{\mathrm{T}}\right) = \mathbf{R}\delta_{kj} \tag{2.73a}$$

$$\mathrm{E}(\mathbf{w}_k) = \mathbf{0} \qquad \mathrm{E}\left(\mathbf{w}_k\mathbf{w}_j^{\mathrm{T}}\right) = \mathbf{Q}\delta_{kj} \tag{2.73b}$$

$$\mathrm{E}\left(\mathbf{w}_k\mathbf{v}_j^{\mathrm{T}}\right) = \mathbf{0} \tag{2.73c}$$

with $\mathbf{Q} \geq 0$, $\mathbf{R} > 0$, and the Kronecker delta $\delta_{kj} = 1$ for $k = j$ and $\delta_{kj} = 0$ otherwise.

(2) The expected value of the initial value and the covariance matrix of the initial error are given by

$$\mathrm{E}(\mathbf{x}_0) = \mathbf{m}_0 \qquad \mathrm{E}\left([\mathbf{x}_0 - \hat{\mathbf{x}}_0][\mathbf{x}_0 - \hat{\mathbf{x}}_0]^{\mathrm{T}}\right) = \mathbf{P}_0 \geq 0 \tag{2.74}$$

with the estimate $\hat{\mathbf{x}}_0$ of the initial value $\mathbf{x}_0$.

(3) The disturbance $\mathbf{w}_k$, $k \geq 0$, and the measurement noise $\mathbf{v}_l$, $l \geq 0$, are uncorrelated with the initial value $\mathbf{x}_0$, i.e.,

$$\mathrm{E}\left(\mathbf{w}_k\mathbf{x}_0^{\mathrm{T}}\right) = \mathbf{0} \tag{2.75a}$$

$$\mathrm{E}\left(\mathbf{v}_l\mathbf{x}_0^{\mathrm{T}}\right) = \mathbf{0}\ . \tag{2.75b}$$

However, due to

$$\mathbf{x}_j = \mathbf{\Phi}^j\mathbf{x}_0 + \sum_{l=0}^{j-1} \mathbf{\Phi}^l(\mathbf{\Gamma}\mathbf{u}_{j-1-l} + \mathbf{G}\mathbf{w}_{j-1-l}) \tag{2.76}$$

and (2.73), the relationship

$$\mathrm{E}\left(\mathbf{w}_k\mathbf{x}_j^{\mathrm{T}}\right) = \mathbf{0} \quad \text{for } k \geq j \tag{2.77a}$$

$$\mathrm{E}\left(\mathbf{v}_l\mathbf{x}_j^{\mathrm{T}}\right) = \mathbf{0} \quad \text{for all } l, j\ . \tag{2.77b}$$

also holds.

For further considerations, the following notation is introduced:

Definition 2.3. The optimal estimate of $\mathbf{x}_k$ considering $0, \dots, j$ measurements is abbreviated as $\hat{\mathbf{x}}(k|j)$.

Theorem 2.7 (Kalman Filter). *The optimal estimate $\hat{\mathbf{x}}(k+1|k)$ of the state $\mathbf{x}_{k+1}$ of the system (2.72) considering $l = 0, \dots, k$ measurements is calculated according to the iteration rule*

$$\begin{aligned}\hat{\mathbf{x}}(k+1|k) &= \mathbf{\Phi}\hat{\mathbf{x}}(k|k-1) + \mathbf{\Gamma}\mathbf{u}_k + \\ &\mathbf{\Phi}\mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}}\Big(\mathbf{C}\mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}} + \mathbf{R}\Big)^{-1}(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}(k|k-1) - \mathbf{D}\mathbf{u}_k)\end{aligned} \tag{2.78}$$

with the covariance matrix of the estimation error

$$\begin{aligned}\mathbf{P}(k+1|k) &= \mathbf{\Phi}\mathbf{P}(k|k-1)\mathbf{\Phi}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}} \\ &- \mathbf{\Phi}\mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}}\Big(\mathbf{C}\mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}} + \mathbf{R}\Big)^{-1}\mathbf{C}\mathbf{P}(k|k-1)\mathbf{\Phi}^{\mathrm{T}}\end{aligned} \tag{2.79}$$

and the initial values $\hat{\mathbf{x}}(0|-1) = \mathbf{m}_0$ and $\mathbf{P}(0|-1) = \mathbf{P}_0$.

Proof of Theorem 2.7. Assume that the measurements $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{k-1}$ have been used for the optimal estimate $\hat{\mathbf{x}}(k|k-1)$ with the error covariance matrix

$$\mathbf{P}(k|k-1) = \mathrm{E}\Big([\mathbf{x}_k - \hat{\mathbf{x}}(k|k-1)][\mathbf{x}_k - \hat{\mathbf{x}}(k|k-1)]^{\mathrm{T}}\Big) \tag{2.80}$$

At time k, the measurement $\mathbf{y}_k$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{v}_k \tag{2.81}$$

is now used to improve the estimate of $\mathbf{x}_k$. According to Theorem 2.6, the estimate $\hat{\mathbf{x}}(k|k)$ of $\mathbf{x}_k$ is

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathrm{E}\Big(\mathbf{x}_k\tilde{\mathbf{y}}_k^{\mathrm{T}}\Big)\Big[\mathrm{E}\Big(\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k^{\mathrm{T}}\Big)\Big]^{-1}\tilde{\mathbf{y}}_k \tag{2.82a}$$

$$\tilde{\mathbf{y}}_k = \mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}(k|k-1) - \mathbf{D}\mathbf{u}_k \tag{2.82b}$$

or with (2.53) in

$$\mathrm{E}\Big(\mathbf{x}_k\tilde{\mathbf{y}}_k^{\mathrm{T}}\Big) = \mathrm{E}\Big(\mathbf{x}_k(\mathbf{C}\mathbf{x}_k + \mathbf{v}_k - \mathbf{C}\hat{\mathbf{x}}(k|k-1))^{\mathrm{T}}\Big) = \mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}} \tag{2.83}$$

and

$$\mathrm{E}\Big(\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k^{\mathrm{T}}\Big) = \mathbf{C}\mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}} + \mathbf{R} \tag{2.84}$$

it follows

$$\begin{aligned}\hat{\mathbf{x}}(k|k) &= \hat{\mathbf{x}}(k|k-1) + \\ &\mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}}\Big(\mathbf{C}\mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}} + \mathbf{R}\Big)^{-1}(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}(k|k-1) - \mathbf{D}\mathbf{u}_k)\ .\end{aligned} \tag{2.85}$$

Thus, the error covariance matrix can be written in the form (compare (2.44), (2.70))

$$\begin{aligned}\mathbf{P}(k|k) &= \mathrm{E}\Big([\mathbf{x}_k - \hat{\mathbf{x}}(k|k)][\mathbf{x}_k - \hat{\mathbf{x}}(k|k)]^{\mathrm{T}}\Big) = \\ &\mathbf{P}(k|k-1) - \mathbf{P}(k|k-1)\mathbf{C}^{\mathrm{T}}\Big(\mathbf{C}\mathbf{P}\mathbf{C}^{\mathrm{T}} + \mathbf{R}\Big)^{-1}\mathbf{C}\mathbf{P}(k|k-1)\end{aligned} \tag{2.86}$$

According to Theorem 2.5, the optimal estimate of $\mathbf{\Phi}\mathbf{x}_k$ is equal to the optimal estimate $\hat{\mathbf{x}}_k$ of $\mathbf{x}_k$ multiplied by $\mathbf{\Phi}$, so it holds that

$$\hat{\mathbf{x}}(k+1|k) = \mathbf{\Phi}\hat{\mathbf{x}}(k|k) + \mathbf{\Gamma}\mathbf{u}_k \ . \tag{2.87}$$

For the covariance matrix of the estimation error, we obtain

$$\begin{aligned}\mathbf{P}(k+1|k) &= \mathrm{E}\Big([\mathbf{x}_{k+1} - \hat{\mathbf{x}}(k+1|k)][\mathbf{x}_{k+1} - \hat{\mathbf{x}}(k+1|k)]^{\mathrm{T}}\Big) \\ &= \mathrm{E}\Big([\mathbf{\Phi}(\mathbf{x}_k - \hat{\mathbf{x}}(k|k)) + \mathbf{G}\mathbf{w}_k][\mathbf{\Phi}(\mathbf{x}_k - \hat{\mathbf{x}}(k|k)) + \mathbf{G}\mathbf{w}_k]^{\mathrm{T}}\Big) \\ &= \mathbf{\Phi}\mathbf{P}(k|k)\mathbf{\Phi}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}}\end{aligned} \tag{2.88}$$

Combining (2.85)–(2.88) directly yields the result of Theorem 2.7. □

The composition of the error covariance matrix (2.79) should be interpreted at this point: The term $\mathbf{\Phi}\mathbf{P}(k|k-1)\mathbf{\Phi}^{\mathrm{T}}$ describes the change in the covariance matrix due to the system dynamics, $\mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}}$ indicates the increase in error variance due to the disturbance $\mathbf{w}$, and the remaining expression with a negative sign describes how the error variance is reduced by adding the information of new measurements.

2.3.1 The Kalman Filter as an Optimal Observer

Introducing the abbreviations $\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}(k+1|k)$, $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}(k|k-1)$, $\mathbf{P}_{k+1} = \mathbf{P}(k+1|k)$, and $\mathbf{P}_k = \mathbf{P}(k|k-1)$, (2.78) and (2.79) can also be represented in the compact form

$$\hat{\mathbf{x}}_{k+1} = \mathbf{\Phi}\hat{\mathbf{x}}_k + \mathbf{\Gamma}\mathbf{u}_k + \hat{\mathbf{K}}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k - \mathbf{D}\mathbf{u}_k) \tag{2.89}$$

with

$$\hat{\mathbf{K}}_k = \mathbf{\Phi}\mathbf{P}_k\mathbf{C}^{\mathrm{T}}\Big(\mathbf{C}\mathbf{P}_k\mathbf{C}^{\mathrm{T}} + \mathbf{R}\Big)^{-1} \tag{2.90}$$

and

$$\mathbf{P}_{k+1} = \mathbf{\Phi}\mathbf{P}_k\mathbf{\Phi}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}} - \mathbf{\Phi}\mathbf{P}_k\mathbf{C}^{\mathrm{T}}\Big(\mathbf{C}\mathbf{P}_k\mathbf{C}^{\mathrm{T}} + \mathbf{R}\Big)^{-1}\mathbf{C}\mathbf{P}_k\mathbf{\Phi}^{\mathrm{T}} \tag{2.91}$$

Equation (2.91) is also called the *discrete Riccati equation*. Comparing (2.89) with a Luenberger observer, it is seen that the Kalman filter is an observer with a *time-varying observer gain matrix* $\hat{\mathbf{K}}_k$. The expected value of the observation error $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

satisfies the iteration rule

$$\begin{aligned} \mathrm{E}(\mathbf{e}_{k+1}) &= \mathrm{E}\Big(\mathbf{\Phi}\mathbf{x}_k + \mathbf{G}\mathbf{w}_k - \mathbf{\Phi}\hat{\mathbf{x}}_k - \hat{\mathbf{K}}_k(\mathbf{C}\mathbf{x}_k + \mathbf{v}_k - \mathbf{C}\hat{\mathbf{x}}_k)\Big) \\ &= \mathrm{E}\Big(\Big(\mathbf{\Phi} - \hat{\mathbf{K}}_k\mathbf{C}\Big)(\mathbf{x}_k - \hat{\mathbf{x}}_k) + \mathbf{G}\mathbf{w}_k - \hat{\mathbf{K}}_k\mathbf{v}_k\Big) \\ &= \Big(\mathbf{\Phi} - \hat{\mathbf{K}}_k\mathbf{C}\Big)\,\mathrm{E}(\mathbf{e}_k)\ . \end{aligned} \tag{2.92}$$

If $\hat{\mathbf{x}}_0 = \mathrm{E}(\mathbf{x}_0) = \mathbf{m}_0$ is set, then $\mathrm{E}(\mathbf{e}_k) = \mathbf{0}$ for all $k \geq 0$. Furthermore, it can be seen from (2.90) and (2.91) that, starting from the initial value $\mathbf{P}_0$, the error covariance matrix $\mathbf{P}_k$ and thus also $\hat{\mathbf{K}}_k$ can be pre-calculated and stored in the computer without knowledge of the measurements $\mathbf{y}_k$ for all $k \geq 0$.

If no previous measurement values about the process are available, one typically sets $\hat{\mathbf{x}}_0 = \mathbf{0}$ and $\mathbf{P}_0 = \alpha\mathbf{E}$ for $\alpha \gg 1$. If the observer runs for a very long time, the problem can be treated mathematically as if it were in operation indefinitely. It turns out that for infinitely long time, the covariance matrix of the estimation error converges to a stationary value $\mathbf{P}_\infty$. In this case, the observer gain matrix $\hat{\mathbf{K}}_\infty$ is also constant and is calculated as

$$\hat{\mathbf{K}}_\infty = \mathbf{\Phi}\mathbf{P}_\infty\mathbf{C}^{\mathrm{T}}\Big(\mathbf{C}\mathbf{P}_\infty\mathbf{C}^{\mathrm{T}} + \mathbf{R}\Big)^{-1} \tag{2.93}$$

with $\mathbf{P}_\infty$ as the solution of the so-called *discrete algebraic Riccati equation*

$$\mathbf{P}_\infty = \mathbf{\Phi}\mathbf{P}_\infty\mathbf{\Phi}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}} - \mathbf{\Phi}\mathbf{P}_\infty\mathbf{C}^{\mathrm{T}}\Big(\mathbf{C}\mathbf{P}_\infty\mathbf{C}^{\mathrm{T}} + \mathbf{R}\Big)^{-1}\mathbf{C}\mathbf{P}_\infty\mathbf{\Phi}^{\mathrm{T}}\ . \tag{2.94}$$

Equation (2.94) has a unique symmetric solution $\mathbf{P}_\infty$ with the property that all eigenvalues of $\Big(\mathbf{\Phi} - \hat{\mathbf{K}}_\infty\mathbf{C}\Big)$ lie in the open interior of the unit circle, if the following conditions are met:

(1) The pair $(\mathbf{C}, \mathbf{\Phi})$ is *detectable*, i.e., all eigenvalues outside the unit circle are observable.

(2) The pair $\Big(\mathbf{\Phi}, \mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}}\Big)$ is *stabilizable*, i.e., all eigenvalues outside the unit circle are controllable via the input $\mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}}$.

(3) The matrix $\mathbf{R}$ is positive definite.

Such a solution of the discrete algebraic Riccati equation (2.94) is also called a *stabilizing solution*. Since for this stabilizing solution all eigenvalues of $\Big(\mathbf{\Phi} - \hat{\mathbf{K}}_\infty\mathbf{C}\Big)$ lie in the open interior of the unit circle, the expected value of the observation error decreases according to (2.92), and $\lim_{k\to\infty}\mathrm{E}(\mathbf{e}_k) = \mathbf{0}$ holds. The solution $\mathbf{P}_\infty$ of the discrete algebraic Riccati equation (2.94) can be easily obtained by iterating the discrete Riccati equation (2.91) starting from the initial value $\mathbf{P}_0$ until $\mathbf{P}_k$ changes only insignificantly in terms of a norm. Although the iteration rule generally converges very quickly to a stationary value, in practice, as is also the case in MATLAB, the algebraic Riccati equation (2.94) is solved more efficiently numerically via an eigenvector decomposition, see the MATLAB commands `care` or `dare`.

Exercise 2.10. The motion of a satellite about an axis is modeled in the form

$$I\frac{\mathrm{d}^2}{\mathrm{d}t^2}\varphi = M_c - M_d$$

with the moment of inertia I, the torque M_c as the control variable, the torque M_d as the disturbance, and the angle φ. Determine the corresponding discrete-time system of the form (2.72) for the sampling time $T_a = 1\mathrm{s}$, $I = 1$, and the output variable φ. Assume that the measurement of the angle φ is superimposed with the measurement noise v and that the disturbance torque M_d corresponds to the process disturbance w. Let

$$\mathrm{E}(w) = 0 \qquad \mathrm{E}\left(w^2\right) = q$$
$$\mathrm{E}(v) = 0 \qquad \mathrm{E}\left(v^2\right) = 0.1\ .$$

Calculate and plot the elements of $\mathbf{P}_k$ of the Kalman filter according to the iteration rule (2.91) for the initial value $\mathbf{P}_0 = \mathbf{E}$ and $q = \{0.1, 0.01, 0.001\}$. Implement the Kalman filter in MATLAB/SIMULINK.

Remark: The relationship (2.93) with the constant observer gain matrix $\hat{\mathbf{K}}_\infty$ specifies an *optimal full observer* that is usable *for both single-variable and multivariable systems*. In contrast to observer design using the pole placement method (note the Ackermann formula according to Chapter 8 of the Automation kecture notes), no poles of the error system need to be chosen for the Kalman filter, which can be very difficult, especially in the multivariable case. The behavior of the error system is instead influenced by specifying the covariance matrices $\mathbf{Q}$ of the disturbance $\mathbf{w}$ and $\mathbf{R}$ of the measurement noise $\mathbf{v}$.

For the choice of the covariance matrix $\mathbf{R}$ of the measurement noise $\mathbf{v}$, a very often interpretable approach based on the (noise) characteristics of the sensor can be found. Furthermore, a distinction can be made between reliable and less reliable measurements via the weighting of the entries of the covariance matrix $\mathbf{R}$. If a *measurement is less reliable*, the *corresponding entry* on the main diagonal *of the covariance matrix is chosen to be very large*. This assumed large variance of the measurement causes the observer to weight this measurement less in the state estimation compared to the other measurements. In the practical application of the Kalman filter, it is even common to switch the covariance matrix during operation if one or more sensors deliver implausible measurements or if an operating range is reached where it is known in advance that certain sensors no longer provide reliable information.

These assumptions generally do not apply to the process disturbance $\mathbf{w}$ and thus to the covariance matrix $\mathbf{Q}$. The assumption that $\mathbf{w}$ is white noise is usually not true. One might now get the idea to choose the generally unknown matrix $\mathbf{Q}$ very small or even zero. However, the choice $\mathbf{Q} = \mathbf{0}$ corresponds to the scenario of a system without disturbance $\mathbf{w}$. As can be seen from condition (2) for the solution of the discrete algebraic Riccati equation, the choice $\mathbf{Q} = \mathbf{0}$ (or generally also for $\mathbf{Q} \ll 1$)

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

does not lead to a stabilizing solution. The choice of $\mathbf{Q}$ (and also $\mathbf{R}$) in practical application is usually done by trial and error.

Remark: In the MATLAB CONTROL SYSTEMS TOOLBOX, a slightly more general system of the form

$$\mathbf{x}_{k+1} = \mathbf{\Phi}\mathbf{x}_k + \mathbf{\Gamma}\mathbf{u}_k + \mathbf{G}\mathbf{w}_k, \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.95a}$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{H}\mathbf{w}_k + \mathbf{v}_k, \tag{2.95b}$$

with the additional term $\mathbf{H}\mathbf{w}_k$ in the measured output, is considered. Furthermore, instead of $\mathrm{E}\left(\mathbf{w}_k\mathbf{v}_k^{\mathrm{T}}\right) = \mathbf{0}$, it is allowed that

$$\mathrm{E}\left(\mathbf{w}_k\mathbf{v}_k^{\mathrm{T}}\right) = \mathbf{N}\delta_{kj} \neq \mathbf{0} \tag{2.96}$$

holds. For this case, the optimal state estimation is calculated as

$$\hat{\mathbf{x}}_{k+1} = \mathbf{\Phi}\hat{\mathbf{x}}_k + \mathbf{\Gamma}\mathbf{u}_k + \hat{\mathbf{K}}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k - \mathbf{D}\mathbf{u}_k) \tag{2.97}$$

with

$$\hat{\mathbf{K}}_k = \left(\mathbf{\Phi}\mathbf{P}_k\mathbf{C}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}\mathbf{H}^{\mathrm{T}} + \mathbf{G}\mathbf{N}\right)\left(\mathbf{C}\mathbf{P}_k\mathbf{C}^{\mathrm{T}} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}} + \mathbf{R} + \mathbf{H}\mathbf{N} + \mathbf{N}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\right)^{-1} \tag{2.98}$$

and

$$\mathbf{P}_{k+1} = \mathbf{\Phi}\mathbf{P}_k\mathbf{\Phi}^{\mathrm{T}} + \mathbf{G}\mathbf{Q}\mathbf{G}^{\mathrm{T}} - \hat{\mathbf{K}}_k\left(\mathbf{C}\mathbf{P}_k\mathbf{\Phi}^{\mathrm{T}} + \mathbf{H}\mathbf{Q}\mathbf{G}^{\mathrm{T}} + \mathbf{N}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\right). \tag{2.99}$$

The proof can be found in the literature, in particular [2.2].

2.3.2 Frequency-domain properties of the stationary Kalman filter

Assume that the measurement noise $\mathbf{v}_k$ from (2.72), i.e.,

$$\mathbf{x}_{k+1} = \mathbf{\Phi}\mathbf{x}_k + \mathbf{\Gamma}\mathbf{u}_k + \mathbf{G}\mathbf{w}_k \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.100a}$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{v}_k \tag{2.100b}$$

is generated by a dynamic system of the form

$$\mathbf{z}_{k+1} = \mathbf{\Phi}_{\mathbf{z}}\mathbf{z}_k + \mathbf{n}_k \tag{2.101a}$$
$$\mathbf{v}_k = \mathbf{C}_{\mathbf{z}}\mathbf{z}_k + \mathbf{m}_k \tag{2.101b}$$

with the stochastic disturbances $\mathbf{n}_k$ and $\mathbf{m}_k$. Substituting $\mathbf{v}_k$ from (2.101) into (2.100) yields the extended system in the form

$$\underbrace{\begin{bmatrix}\mathbf{x}_{k+1}\\ \mathbf{z}_{k+1}\end{bmatrix}}_{\tilde{\mathbf{x}}_{k+1}} = \underbrace{\begin{bmatrix}\boldsymbol{\Phi} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}_{\mathbf{z}}\end{bmatrix}}_{\tilde{\boldsymbol{\Phi}}} \underbrace{\begin{bmatrix}\mathbf{x}_k\\ \mathbf{z}_k\end{bmatrix}}_{\tilde{\mathbf{x}}_k} + \underbrace{\begin{bmatrix}\boldsymbol{\Gamma}\\ \mathbf{0}\end{bmatrix}}_{\tilde{\boldsymbol{\Gamma}}} \mathbf{u}_k + \underbrace{\begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{E}\end{bmatrix}}_{\tilde{\mathbf{G}}} \underbrace{\begin{bmatrix}\mathbf{w}_k\\ \mathbf{n}_k\end{bmatrix}}_{\tilde{\mathbf{w}}_k} \qquad \begin{matrix}\mathbf{x}(0) = \mathbf{x}_0\\ \mathbf{z}(0) = \mathbf{z}_0\end{matrix} \tag{2.102a}$$

$$\mathbf{y}_k = \underbrace{\begin{bmatrix}\mathbf{C} & \mathbf{C}_{\mathbf{z}}\end{bmatrix}}_{\tilde{\mathbf{C}}} \underbrace{\begin{bmatrix}\mathbf{x}_k\\ \mathbf{z}_k\end{bmatrix}}_{\tilde{\mathbf{x}}_k} + \mathbf{D}\mathbf{u}_k + \mathbf{m}_k \ . \tag{2.102b}$$

The stationary Kalman filter according to (2.89) and (2.93) for the extended system (2.102) is then

$$\begin{bmatrix}\hat{\mathbf{x}}_{k+1}\\ \hat{\mathbf{z}}_{k+1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Phi} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}_{\mathbf{z}}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}_k\\ \hat{\mathbf{z}}_k\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Gamma}\\ \mathbf{0}\end{bmatrix}\mathbf{u}_k + \underbrace{\begin{bmatrix}\hat{\mathbf{K}}_{\mathbf{x}}\\ \hat{\mathbf{K}}_{\mathbf{z}}\end{bmatrix}}_{\hat{\mathbf{K}}_\infty}(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k - \mathbf{C}_{\mathbf{z}}\hat{\mathbf{z}}_k - \mathbf{D}\mathbf{u}_k) \tag{2.103}$$

or

$$\begin{bmatrix}\hat{\mathbf{x}}_{k+1}\\ \hat{\mathbf{z}}_{k+1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Phi} - \hat{\mathbf{K}}_{\mathbf{x}}\mathbf{C} & -\hat{\mathbf{K}}_{\mathbf{x}}\mathbf{C}_{\mathbf{z}}\\ -\hat{\mathbf{K}}_{\mathbf{z}}\mathbf{C} & \boldsymbol{\Phi}_{\mathbf{z}} - \hat{\mathbf{K}}_{\mathbf{z}}\mathbf{C}_{\mathbf{z}}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}_k\\ \hat{\mathbf{z}}_k\end{bmatrix} + \begin{bmatrix}\hat{\mathbf{K}}_{\mathbf{x}}\\ \hat{\mathbf{K}}_{\mathbf{z}}\end{bmatrix}\mathbf{y}_k + \begin{bmatrix}\boldsymbol{\Gamma} - \hat{\mathbf{K}}_{\mathbf{x}}\mathbf{D}\\ -\hat{\mathbf{K}}_{\mathbf{z}}\mathbf{D}\end{bmatrix}\mathbf{u}_k \ . \tag{2.104}$$

Since the system is linear and thus the superposition principle applies, the input $\mathbf{u}_k = \mathbf{0}$ is set in the following. The z-transfer matrix from the input $\mathbf{y}_k$ to the output $\hat{\mathbf{x}}_k$ is

$$\mathbf{G}(z) = \frac{\hat{\mathbf{x}}_z(z)}{\mathbf{y}_z(z)} = \begin{bmatrix}\mathbf{E} & \mathbf{0}\end{bmatrix}\begin{bmatrix}z\mathbf{E} - \boldsymbol{\Phi} + \hat{\mathbf{K}}_{\mathbf{x}}\mathbf{C} & \hat{\mathbf{K}}_{\mathbf{x}}\mathbf{C}_{\mathbf{z}}\\ \hat{\mathbf{K}}_{\mathbf{z}}\mathbf{C} & z\mathbf{E} - \boldsymbol{\Phi}_{\mathbf{z}} + \hat{\mathbf{K}}_{\mathbf{z}}\mathbf{C}_{\mathbf{z}}\end{bmatrix}^{-1}\begin{bmatrix}\hat{\mathbf{K}}_{\mathbf{x}}\\ \hat{\mathbf{K}}_{\mathbf{z}}\end{bmatrix} \ . \tag{2.105}$$

The following theorem now applies:

Theorem 2.8 (Transmission zeros of the stationary Kalman filter)**.** *The transmission zeros of* (2.105) *are given by the relationship*

$$\det(z\mathbf{E} - \boldsymbol{\Phi}_{\mathbf{z}}) = 0 \tag{2.106}$$

Proof. Before Theorem 2.8 is proven, the concept of a transmission zero must be briefly discussed. The zeros of the transfer function $G(z)$ of a single-input single-output system are usually characterized as the roots of the numerator polynomial of $G(z)$. In the case of a multivariable system with a transfer matrix $\mathbf{G}(z)$, this is no longer as simple. A strict definition is given as follows: The zeros of the transfer matrix $\mathbf{G}(z)$ are the roots of the numerator polynomials of the Smith-McMillan form of $\mathbf{G}(z)$. For the definition of the Smith-McMillan form, refer to the literature cited at the end. A physical interpretation of a transmission zero is now given. Consider

the linear, time-invariant, discrete-time system of the form

$$\mathbf{x}_{k+1} = \mathbf{\Phi}\mathbf{x}_k + \mathbf{\Gamma}\mathbf{u}_k \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.107a}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k \tag{2.107b}$$

with the n-dimensional state $\mathbf{x} \in \mathbb{R}^n$, the p-dimensional input $\mathbf{u} \in \mathbb{R}^p$, the q-dimensional output $\mathbf{y} \in \mathbb{R}^q$, and the matrices $\mathbf{\Phi} \in \mathbb{R}^{n\times n}$, $\mathbf{\Gamma} \in \mathbb{R}^{n\times p}$, and $\mathbf{C} \in \mathbb{R}^{q\times n}$. A complex number z_j is a transmission zero if, for an input of the form $(\mathbf{u}_k) = \mathbf{u}_0\left(z_j^k\right)$, $\mathbf{u}_0 \neq \mathbf{0}$, there exists an initial state $\mathbf{x}_0 \neq \mathbf{0}$ such that the output $(\mathbf{y}_k)$ vanishes identically for all times $k \geq 0$. This property is also known in the literature as *transmission-blocking*. In the z-domain, the output $\mathbf{y}_z(z) = \mathcal{Z}\{(\mathbf{y}_k)\}$ of (2.107) in response to the input

$$\mathbf{u}_z(z) = \mathcal{Z}\{(\mathbf{u}_k)\} = \mathcal{Z}\left\{\mathbf{u}_0\left(z_j^k\right)\right\} = \mathbf{u}_0\frac{z}{z - z_j} \tag{2.108}$$

with the initial value $\mathbf{x}(0) = \mathbf{x}_0$ is calculated as

$$\mathbf{y}_z(z) = \mathbf{C}(z\mathbf{E} - \mathbf{\Phi})^{-1}\left(z\mathbf{x}_0 + \mathbf{\Gamma}\mathbf{u}_0\frac{z}{z - z_j}\right). \tag{2.109}$$

Using the resolvent identity

$$(z\mathbf{E} - \mathbf{\Phi})^{-1} - (z_j\mathbf{E} - \mathbf{\Phi})^{-1} = (z\mathbf{E} - \mathbf{\Phi})^{-1}(z_j - z)(z_j\mathbf{E} - \mathbf{\Phi})^{-1} \tag{2.110}$$

(2.109) can be rewritten as

$$\begin{aligned}\mathbf{y}_z(z) = \mathbf{C}\Big((z_j\mathbf{E} - \mathbf{\Phi})^{-1} + (z\mathbf{E} - \mathbf{\Phi})^{-1}(z_j - z)(z_j\mathbf{E} - \mathbf{\Phi})^{-1}\Big)\frac{\mathbf{\Gamma}\mathbf{u}_0 z}{z - z_j} + \\ \mathbf{C}(z\mathbf{E} - \mathbf{\Phi})^{-1}z\mathbf{x}_0 \\ = \mathbf{C}(z\mathbf{E} - \mathbf{\Phi})^{-1}z\Big(\mathbf{x}_0 - (z_j\mathbf{E} - \mathbf{\Phi})^{-1}\mathbf{\Gamma}\mathbf{u}_0\Big) + \mathbf{C}(z_j\mathbf{E} - \mathbf{\Phi})^{-1}\mathbf{\Gamma}\mathbf{u}_0\frac{z}{z - z_j}\end{aligned} \tag{2.111}$$

Exercise 2.11. Prove the identity of (2.110).

From (2.111), it can be seen that $\mathbf{y}_z(z)$ vanishes identically only if the equations

$$\mathbf{x}_0 - (z_j\mathbf{E} - \mathbf{\Phi})^{-1}\mathbf{\Gamma}\mathbf{u}_0 = \mathbf{0} \tag{2.112a}$$

$$\mathbf{C}(z_j\mathbf{E} - \mathbf{\Phi})^{-1}\mathbf{\Gamma}\mathbf{u}_0 = \mathbf{0} \tag{2.112b}$$

are satisfied. In summary, it can be stated that for a transmission zero z_j, non-trivial vectors $\mathbf{u}_0$ and $\mathbf{x}_0$ exist that satisfy the system of equations

$$\begin{bmatrix} (z_j\mathbf{E} - \mathbf{\Phi}) & -\mathbf{\Gamma} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{u}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{2.113}$$

and have the property that the output $(\mathbf{y}_k)$ of (2.107) for the input $(\mathbf{u}_k) = \mathbf{u}_0\left(z_j^k\right)$ and the initial value $\mathbf{x}_0$ vanishes identically for all times $k \geq 0$. Applying this result to the transfer matrix $\mathbf{G}(z)$ of (2.105) with the input $\mathbf{y}_k$ and the output $\hat{\mathbf{x}}_k$, the conditions (2.113) in this case are

$$\begin{bmatrix} z_j\mathbf{E} - \mathbf{\Phi} + \hat{\mathbf{K}}_\mathbf{x}\mathbf{C} & \hat{\mathbf{K}}_\mathbf{x}\mathbf{C}_\mathbf{z} \\ \hat{\mathbf{K}}_\mathbf{z}\mathbf{C} & z_j\mathbf{E} - \mathbf{\Phi}_\mathbf{z} + \hat{\mathbf{K}}_\mathbf{z}\mathbf{C}_\mathbf{z} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_0 \\ \hat{\mathbf{z}}_0 \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{K}}_\mathbf{x} \\ \hat{\mathbf{K}}_\mathbf{z} \end{bmatrix} \mathbf{y}_0 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{2.114a}$$

$$\begin{bmatrix} \mathbf{E} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_0 \\ \hat{\mathbf{z}}_0 \end{bmatrix} = \mathbf{0} \ . \tag{2.114b}$$

Exercise 2.12. Derive the relationships (2.114).

From (2.114b), we obtain $\hat{\mathbf{x}}_0 = \mathbf{0}$, and thus

$$\begin{aligned} \hat{\mathbf{K}}_\mathbf{x}(\mathbf{C}_\mathbf{z}\hat{\mathbf{z}}_0 - \mathbf{y}_0) &= \mathbf{0} \\ \left(z_j\mathbf{E} - \mathbf{\Phi}_\mathbf{z} + \hat{\mathbf{K}}_\mathbf{z}\mathbf{C}_\mathbf{z}\right)\hat{\mathbf{z}}_0 - \hat{\mathbf{K}}_\mathbf{z}\mathbf{y}_0 &= \mathbf{0} \ . \end{aligned} \tag{2.115}$$

Assuming the full column rank of $\hat{\mathbf{K}}_\mathbf{x}$, $\mathbf{C}_\mathbf{z}\hat{\mathbf{z}}_0 - \mathbf{y}_0 = \mathbf{0}$ follows, and thus $(z_j\mathbf{E} - \mathbf{\Phi}_\mathbf{z})\hat{\mathbf{z}}_0 = \mathbf{0}$. It is now seen that (2.115) has non-trivial solutions for $\mathbf{y}_0$ and $\hat{\mathbf{z}}_0$ if and only if $\det(z_j\mathbf{E} - \mathbf{\Phi}_\mathbf{z}) = 0$ holds, which proves Theorem 2.8. □

Remark: This result shows that the stationary Kalman filter (2.103) has zeros at the poles of the disturbance model (2.101). To now design a Kalman filter that blocks certain frequencies, simply choose a disturbance model with poles at these frequencies. In general, it can be said that the greater the energy of the disturbance at the respective frequencies, the more the Kalman filter suppresses these frequencies.

Exercise 2.13. Design a Kalman filter to estimate the rotational speed ω of a permanent magnet DC motor based on a measurement of the rotation angle φ. For a certain choice of parameters and neglecting the dynamics of the electrical subsystem, the

model of the DC motor is obtained in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\varphi\\ \omega\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & -1\end{bmatrix}\begin{bmatrix}\varphi\\ \omega\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix}u$$
$$y = \varphi = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}\varphi\\ \omega\end{bmatrix}.$$

It is now desired that resonance frequencies with an angular frequency ω_0, resulting from the mechanical load, are suppressed by the Kalman filter. The corresponding disturbance model can be modeled in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}z_1\\ z_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ -\omega_0^2 & -2\xi\omega_0\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix} + \begin{bmatrix}0\\ \omega_0^2\end{bmatrix}n$$
$$v = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix}$$

with white noise n as the input signal. Design a Kalman filter for these specifications for a sampling time $T_a = 0.05$ s and investigate the influence of different values for ξ in the range $0.01 < \xi < 0.7$. Choose $\omega_0 = 3$, $\mathbf{G} = \mathbf{E}$, $\mathbf{Q} = \mathbf{E}$, and $\mathbf{R} = 1$. Plot the Bode diagram of the Kalman filter and test your Kalman filter via simulation.

2.4 The Extended Kalman Filter

Before discussing the Extended Kalman Filter as an observer for nonlinear systems, the Kalman Filter from Section 2.3 for linear time-variant sampled systems of the form

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_k\mathbf{x}_k + \mathbf{\Gamma}_k\mathbf{u}_k + \mathbf{G}_k\mathbf{w}_k \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.116a}$$
$$\mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{D}_k\mathbf{u}_k + \mathbf{v}_k \tag{2.116b}$$

with the n-dimensional state $\mathbf{x} \in \mathbb{R}^n$, the p-dimensional deterministic input $\mathbf{u} \in \mathbb{R}^p$, the q-dimensional output $\mathbf{y} \in \mathbb{R}^q$, the r-dimensional disturbance $\mathbf{w} \in \mathbb{R}^r$, the measurement noise $\mathbf{v}$, and the time-variant matrices $\mathbf{\Phi}_k \in \mathbb{R}^{n\times n}$, $\mathbf{\Gamma}_k \in \mathbb{R}^{n\times p}$, $\mathbf{G}_k \in \mathbb{R}^{n\times r}$, $\mathbf{C}_k \in \mathbb{R}^{q\times n}$, and $\mathbf{D}_k \in \mathbb{R}^{q\times p}$ will be written down. Analogous to Section 2.3, the following assumptions are made:

(1) For the disturbance $\mathbf{w}$ and the measurement noise $\mathbf{v}$, it is assumed that

$$\mathrm{E}(\mathbf{v}_k) = \mathbf{0} \qquad \mathrm{E}\left(\mathbf{v}_k\mathbf{v}_j^{\mathrm{T}}\right) = \mathbf{R}_k\delta_{kj} \tag{2.117a}$$
$$\mathrm{E}(\mathbf{w}_k) = \mathbf{0} \qquad \mathrm{E}\left(\mathbf{w}_k\mathbf{w}_j^{\mathrm{T}}\right) = \mathbf{Q}_k\delta_{kj} \tag{2.117b}$$
$$\mathrm{E}\left(\mathbf{w}_k\mathbf{v}_j^{\mathrm{T}}\right) = \mathbf{0} \tag{2.117c}$$

with $\mathbf{Q}_k \geq 0$ and $\mathbf{R}_k > 0$ and the Kronecker delta $\delta_{kj} = 1$ for $k = j$ and $\delta_{kj} = 0$ otherwise.

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

(2) The expected value of the initial value and the covariance matrix of the initial error are given by

$$\mathrm{E}(\mathbf{x}_0) = \mathbf{m}_0 \qquad \mathrm{E}\Big([\mathbf{x}_0 - \hat{\mathbf{x}}_0][\mathbf{x}_0 - \hat{\mathbf{x}}_0]^{\mathrm{T}}\Big) = \mathbf{P}_0 \geq 0 \tag{2.118}$$

with the estimate $\hat{\mathbf{x}}_0$ of the initial value $\mathbf{x}_0$.

(3) The disturbance $\mathbf{w}_k$, $k \geq 0$, and the measurement noise $\mathbf{v}_l$, $l \geq 0$, are uncorrelated with the initial value $\mathbf{x}_0$, i.e., it holds that

$$\mathrm{E}\Big(\mathbf{w}_k \mathbf{x}_0^{\mathrm{T}}\Big) = \mathbf{0} \tag{2.119a}$$

$$\mathrm{E}\Big(\mathbf{v}_l \mathbf{x}_0^{\mathrm{T}}\Big) = \mathbf{0}\ . \tag{2.119b}$$

The derivation of the Kalman filter for the system (2.116) proceeds in a completely analogous manner as in Section 2.3 and is given for $k \geq 0$, compare (2.89)–(2.91),

$$\hat{\mathbf{K}}_k = \mathbf{\Phi}_k \mathbf{P}_k \mathbf{C}_k^{\mathrm{T}} \Big(\mathbf{C}_k \mathbf{P}_k \mathbf{C}_k^{\mathrm{T}} + \mathbf{R}_k\Big)^{-1} \tag{2.120a}$$

$$\hat{\mathbf{x}}_{k+1} = \mathbf{\Phi}_k \hat{\mathbf{x}}_k + \mathbf{\Gamma}_k \mathbf{u}_k + \hat{\mathbf{K}}_k (\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k - \mathbf{D}_k \mathbf{u}_k) \tag{2.120b}$$

$$\mathbf{P}_{k+1} = \mathbf{\Phi}_k \mathbf{P}_k \mathbf{\Phi}_k^{\mathrm{T}} + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\mathrm{T}} - \mathbf{\Phi}_k \mathbf{P}_k \mathbf{C}_k^{\mathrm{T}} \Big(\mathbf{C}_k \mathbf{P}_k \mathbf{C}_k^{\mathrm{T}} + \mathbf{R}_k\Big)^{-1} \mathbf{C}_k \mathbf{P}_k \mathbf{\Phi}_k^{\mathrm{T}}\ . \tag{2.120c}$$

If the initial value $\mathbf{x}_0$ is known, then we set $\hat{\mathbf{x}}_0 = \mathbf{x}_0$ and $\mathbf{P}_0 = \mathbf{0}$, and in the case that no information about the initial value is available, $\hat{\mathbf{x}}_0 = \mathbf{0}$ and $\mathbf{P}_0 = \alpha \mathbf{E}$ with $\alpha \gg 1$ is typically chosen. It should be noted that this representation can also be generalized for $\mathbf{H} \neq \mathbf{0}$ and $\mathrm{E}\Big(\mathbf{w}_k \mathbf{v}_j^{\mathrm{T}}\Big) \neq \mathbf{0}$ analogously to the discussions in the last section.

In the literature, it is often common to represent the Kalman filter in a slightly different form. The optimal estimation of the state $\mathbf{x}_k$ and the error covariance matrix $\mathbf{P}_k$ considering $0, \ldots, k-1$ measurements, compare Definition 2.3,

$$\hat{\mathbf{x}}_k^- = \hat{\mathbf{x}}(k|k-1) \tag{2.121a}$$

$$\mathbf{P}_k^- = \mathbf{P}(k|k-1) = \mathrm{E}\Big(\Big[\mathbf{x}_k - \hat{\mathbf{x}}_k^-\Big]\Big[\mathbf{x}_k - \hat{\mathbf{x}}_k^-\Big]^{\mathrm{T}}\Big) \tag{2.121b}$$

is referred to as the *a priori estimate*, and the optimal estimation of $\mathbf{x}_k$ and $\mathbf{P}_k$ considering $0, \ldots, k$ measurements

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}(k|k) \tag{2.122a}$$

$$\mathbf{P}_k^+ = \mathbf{P}(k|k) = \mathrm{E}\Big(\Big[\mathbf{x}_k - \hat{\mathbf{x}}_k^+\Big]\Big[\mathbf{x}_k - \hat{\mathbf{x}}_k^+\Big]^{\mathrm{T}}\Big) \tag{2.122b}$$

is referred to as the *a posteriori estimate*. (2.120) can thus be written in the equivalent form

Kalman Gain Matrix:	$\hat{\mathbf{L}}_k = \mathbf{P}_k^- \mathbf{C}_k^{\mathrm{T}} \left(\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^{\mathrm{T}} + \mathbf{R}_k \right)^{-1}$	(2.123a)
State Estimate Update:	$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \hat{\mathbf{L}}_k \left(\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^- - \mathbf{D}_k \mathbf{u}_k \right)$	(2.123b)
Error Covariance Update:	$\mathbf{P}_k^+ = \left(\mathbf{E} - \hat{\mathbf{L}}_k \mathbf{C}_k \right) \mathbf{P}_k^-$	(2.123c)
State Extrapolation (2.87):	$\hat{\mathbf{x}}_{k+1}^- = \mathbf{\Phi}_k \hat{\mathbf{x}}_k^+ + \mathbf{\Gamma}_k \mathbf{u}_k$	(2.123d)
Error Covariance Extrapolation (2.88):	$\mathbf{P}_{k+1}^- = \mathbf{\Phi}_k \mathbf{P}_k^+ \mathbf{\Phi}_k^{\mathrm{T}} + \mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\mathrm{T}}$	(2.123e)

for $k \geq 0$ and the initial values $\hat{\mathbf{x}}_0^- = \hat{\mathbf{x}}_0$ and $\mathbf{P}_0^- = \mathbf{P}_0$.

Exercise 2.14. Show the equivalence of the relations (2.120) and (2.123). For this, perform the following substitutions in (2.123): $\hat{\mathbf{x}}_{k+1}^- = \hat{\mathbf{x}}_{k+1}$, $\hat{\mathbf{x}}_k^- = \hat{\mathbf{x}}_k$, $\mathbf{P}_{k+1}^- = \mathbf{P}_{k+1}$, $\mathbf{P}_k^- = \mathbf{P}_k$, and $\mathbf{\Phi}_k \hat{\mathbf{L}}_k = \hat{\mathbf{K}}_k$.

The Extended Kalman Filter (EKF) design is generally based on a nonlinear, time-variant, continuous-time multivariable system of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{w}, t) \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.124a}$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \mathbf{v}, t) \tag{2.124b}$$

with the n-dimensional state $\mathbf{x} \in \mathbb{R}^n$, the p-dimensional deterministic input $\mathbf{u} \in \mathbb{R}^p$, the q-dimensional output $\mathbf{y} \in \mathbb{R}^q$, the r-dimensional disturbance $\mathbf{w} \in \mathbb{R}^r$, and the measurement noise $\mathbf{v}$. Since the Kalman filter is normally implemented in a digital computer, the control variables are applied to the process via a zero-order hold (D/A converter) with the sampling time T_a, and the measured variables are sampled with the sampling time T_a via an A/D converter, the corresponding sampled system to (2.124)

$$\mathbf{x}_{k+1} = \mathbf{F}_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.125a}$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k) \tag{2.125b}$$

must be calculated. From the lecture on Automation (Section 6.2.1), it is known that the exact solution of (2.124) is necessary to determine the sampled system.

Remark: The solution of a nonlinear system of differential equations of the form (2.124) is known to be only possible in special cases. Therefore, an approximate solution using an integration method is sought below. It is assumed that the control variable $\mathbf{u}(t)$ and the disturbance $\mathbf{w}(t)$ are constant for the sampling interval $kT_a \leq t < (k+1)T_a$, i.e., $\mathbf{u}(t) = \mathbf{u}(kT_a) = \mathbf{u}_k$ and $\mathbf{w}(t) = \mathbf{w}(kT_a) = \mathbf{w}_k$, and the differential equation (2.124a) is integrated over the sampling interval

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \int_{kT_a}^{(k+1)T_a} \mathbf{f}(\mathbf{x}(t), \mathbf{u}_k, \mathbf{w}_k, t)\, \mathrm{d}t \tag{2.126}$$

with $\mathbf{x}_{k+1} = \mathbf{x}((k+1)T_a)$ and $\mathbf{x}_k = \mathbf{x}(kT_a)$. The approximation of the integral in (2.126) can be done in different ways. In the following, only two possible solutions

will be given:

(1) *Euler method*

$$\int_{kT_a}^{(k+1)T_a} \mathbf{f}(\mathbf{x}(t), \mathbf{u}_k, \mathbf{w}_k, t)\,\mathrm{d}t = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k, kT_a)T_a \tag{2.127}$$

(2) *Fourth-order Runge-Kutta method*

$$\int_{kT_a}^{(k+1)T_a} \mathbf{f}(\mathbf{x}(t), \mathbf{u}_k, \mathbf{w}_k, t)\,\mathrm{d}t = \frac{\Delta\mathbf{x}_1 + 2\Delta\mathbf{x}_2 + 2\Delta\mathbf{x}_3 + \Delta\mathbf{x}_4}{6} \tag{2.128}$$

with

$$\begin{aligned}
\Delta\mathbf{x}_1 &= \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k, kT_a)T_a \\
\Delta\mathbf{x}_2 &= \mathbf{f}\Big(\mathbf{x}_k + \frac{\Delta\mathbf{x}_1}{2}, \mathbf{u}_k, \mathbf{w}_k, kT_a + \frac{T_a}{2}\Big)T_a \\
\Delta\mathbf{x}_3 &= \mathbf{f}\Big(\mathbf{x}_k + \frac{\Delta\mathbf{x}_2}{2}, \mathbf{u}_k, \mathbf{w}_k, kT_a + \frac{T_a}{2}\Big)T_a \\
\Delta\mathbf{x}_4 &= \mathbf{f}(\mathbf{x}_k + \Delta\mathbf{x}_3, \mathbf{u}_k, \mathbf{w}_k, (k+1)T_a)T_a\ .
\end{aligned} \tag{2.129}$$

Here, for $\Delta\mathbf{x}_4$ in (2.129), the left-hand limit $\lim_{t\to(k+1)T_a} \mathbf{u}(t) = \mathbf{u}_k$ and $\lim_{t\to(k+1)T_a} \mathbf{w}(t) = \mathbf{w}_k$ was used.

The output equation (2.125b) of the sampled system is obtained very easily for both cases by evaluating the output equation (2.124b) of the continuous-time multivariable system for $t = kT_a$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, kT_a) = \mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k)\ . \tag{2.130}$$

Assuming that a discrete-time system of the form (2.125) is given by the relations (2.126)–(2.130), the *idea of the Extended Kalman Filter* is based on the fact that a Taylor series expansion is performed for the right-hand side of (2.125a) around the point $\mathbf{x}_k = \hat{\mathbf{x}}_k^+$, $\mathbf{u}_k = \mathbf{u}_k$ and $\mathbf{w}_k = \mathbf{0}$ and truncated after the linear term, i.e.,

$$\mathbf{x}_{k+1} = \mathbf{F}_k\Big(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\Big) + \frac{\partial}{\partial\mathbf{x}_k}\mathbf{F}_k\Big(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\Big)\Big(\mathbf{x}_k - \hat{\mathbf{x}}_k^+\Big) + \frac{\partial}{\partial\mathbf{w}_k}\mathbf{F}_k\Big(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\Big)\mathbf{w}_k\ . \tag{2.131}$$

Analogously, the right-hand side of the output equation (2.125b) is developed into a Taylor series around the point $\mathbf{x}_k = \hat{\mathbf{x}}_k^-$, $\mathbf{u}_k = \mathbf{u}_k$ and $\mathbf{v}_k = \mathbf{0}$ and truncated after the linear term

$$\mathbf{y}_k = \mathbf{h}_k\Big(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\Big) + \frac{\partial}{\partial\mathbf{x}_k}\mathbf{h}_k\Big(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\Big)\Big(\mathbf{x}_k - \hat{\mathbf{x}}_k^-\Big) + \frac{\partial}{\partial\mathbf{v}_k}\mathbf{h}_k\Big(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\Big)\mathbf{v}_k\ . \tag{2.132}$$

Note that the following simplified notation is used consistently here:

$$\frac{\partial}{\partial\mathbf{x}_k}\mathbf{F}_k\Big(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\Big) = \frac{\partial}{\partial\mathbf{x}_k}\mathbf{F}_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k)\Big|_{\mathbf{x}_k=\hat{\mathbf{x}}_k^-, \mathbf{u}_k=\mathbf{u}_k, \mathbf{w}_k=\mathbf{0}} \tag{2.133}$$

The relations (2.131) and (2.132) can be written more compactly for further considerations in the form

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_k \mathbf{x}_k + \bar{\mathbf{u}}_k + \mathbf{G}_k \mathbf{w}_k \tag{2.134a}$$
$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \breve{\mathbf{u}}_k + \breve{\mathbf{v}}_k \tag{2.134b}$$

with

$$\begin{aligned} \mathbf{\Phi}_k &= \frac{\partial}{\partial \mathbf{x}_k} \mathbf{F}_k\left(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\right) & \bar{\mathbf{u}}_k &= \mathbf{F}_k\left(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\right) - \mathbf{\Phi}_k \hat{\mathbf{x}}_k^+ \\ \mathbf{G}_k &= \frac{\partial}{\partial \mathbf{w}_k} \mathbf{F}_k\left(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\right) & \mathbf{C}_k &= \frac{\partial}{\partial \mathbf{x}_k} \mathbf{h}_k\left(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\right) \\ \breve{\mathbf{u}}_k &= \mathbf{h}_k\left(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\right) - \mathbf{C}_k \hat{\mathbf{x}}_k^- & \breve{\mathbf{v}}_k &= \frac{\partial}{\partial \mathbf{v}_k} \mathbf{h}_k\left(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\right) \mathbf{v}_k \end{aligned} \tag{2.135}$$

It is now obvious that the structure of the system (2.134) directly allows the application of the Kalman filter according to (2.123). The necessary calculation steps for the implementation are summarized again below.

(1) For the nonlinear, time-variant, continuous-time multivariable system (2.124), calculate a discrete-time system of the form (2.125).

(2) The estimated state and the covariance matrix of the estimation error must be initialized for the initial time point with $\hat{\mathbf{x}}_0^-$ and $\mathbf{P}_0^-$.

(3) It is again assumed for the disturbance $\mathbf{w}_k$ and the measurement noise $\breve{\mathbf{v}}_k$ in (2.134) that

$$\mathrm{E}(\breve{\mathbf{v}}_k) = \mathbf{0} \qquad \mathrm{E}\left(\breve{\mathbf{v}}_k \breve{\mathbf{v}}_j^{\mathrm{T}}\right) = \mathbf{R}_k \delta_{kj} \tag{2.136a}$$
$$\mathrm{E}(\mathbf{w}_k) = \mathbf{0} \qquad \mathrm{E}\left(\mathbf{w}_k \mathbf{w}_j^{\mathrm{T}}\right) = \mathbf{Q}_k \delta_{kj} \tag{2.136b}$$
$$\mathrm{E}\left(\mathbf{w}_k \breve{\mathbf{v}}_j^{\mathrm{T}}\right) = \mathbf{0} \tag{2.136c}$$

with $\mathbf{Q}_k \geq 0$ and $\mathbf{R}_k > 0$ and the Kronecker delta $\delta_{kj} = 1$ for $k = j$ and $\delta_{kj} = 0$ otherwise.

(4) The iteration equations of the Extended Kalman Filter are then for $k \geq 0$

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

$$\mathbf{C}_k = \frac{\partial}{\partial \mathbf{x}_k}\mathbf{h}_k\Big(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\Big) \tag{2.137a}$$

$$\hat{\mathbf{L}}_k = \mathbf{P}_k^-\mathbf{C}_k^{\mathrm{T}}\Big(\mathbf{C}_k\mathbf{P}_k^-\mathbf{C}_k^{\mathrm{T}} + \mathbf{R}_k\Big)^{-1} \tag{2.137b}$$

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \hat{\mathbf{L}}_k\Big(\mathbf{y}_k - \mathbf{C}_k\hat{\mathbf{x}}_k^- - \breve{\mathbf{u}}_k\Big) = \hat{\mathbf{x}}_k^- + \hat{\mathbf{L}}_k\Big(\mathbf{y}_k - \mathbf{h}_k\Big(\hat{\mathbf{x}}_k^-, \mathbf{u}_k, \mathbf{0}\Big)\Big) \tag{2.137c}$$

$$\mathbf{P}_k^+ = \Big(\mathbf{E} - \hat{\mathbf{L}}_k\mathbf{C}_k\Big)\mathbf{P}_k^- \tag{2.137d}$$

$$\mathbf{\Phi}_k = \frac{\partial}{\partial \mathbf{x}_k}\mathbf{F}_k\Big(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\Big) \tag{2.137e}$$

$$\mathbf{G}_k = \frac{\partial}{\partial \mathbf{w}_k}\mathbf{F}_k\Big(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\Big) \tag{2.137f}$$

$$\hat{\mathbf{x}}_{k+1}^- = \mathbf{\Phi}_k\hat{\mathbf{x}}_k^+ + \underbrace{\mathbf{F}_k\Big(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\Big) - \mathbf{\Phi}_k\hat{\mathbf{x}}_k^+}_{\bar{\mathbf{u}}_k} = \mathbf{F}_k\Big(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}\Big) \tag{2.137g}$$

$$\mathbf{P}_{k+1}^- = \mathbf{\Phi}_k\mathbf{P}_k^+\mathbf{\Phi}_k^{\mathrm{T}} + \mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^{\mathrm{T}}\ . \tag{2.137h}$$

Remark: In the Extended Kalman Filter design, it is assumed that the linearized transformation of the mean and covariance corresponds with good accuracy to the mean and covariance of the nonlinear transformation. This assumption is generally not true, which is why the *Unscented Kalman Filter* considered in the next section is often used to improve the observer design for nonlinear systems.

Exercise 2.15. The mathematical model

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x_1 &= x_2 + w_1\\ \frac{\mathrm{d}}{\mathrm{d}t}x_2 &= \frac{1}{2}\rho_0\exp\Big(-\frac{x_1}{k}\Big)C_w\frac{A}{m}x_2^2 - g + w_2\end{aligned}$$

describes the free fall of a body of mass m and cross-sectional area A in the Earth's atmosphere with altitude x_1 and velocity x_2. The term $\rho_0\exp(-x_1/k)$ corresponds to the altitude-dependent density in the atmosphere (ρ_0 density at sea level), whereby the term with x_2^2 describes the deceleration due to air resistance with the drag coefficient C_w. Furthermore, g represents the acceleration due to gravity. The process noise is given by the stochastic variables w_1 and w_2. The altitude x_1 can be determined via the output equation

$$y = x_1 + v$$

with the measurement noise v. Design an Extended Kalman Filter for the parameters $\rho_0 = 1.2\,\mathrm{kg/m^3}$, $g = 9.81\,\mathrm{m/s^2}$, $k = 9100\,\mathrm{m}$, $A = 0.5\,\mathrm{m^2}$, $m = 100\,\mathrm{kg}$, and $C_w = 0.5$ that estimates, in addition to the altitude x_1 and velocity x_2, the constant drag coefficient C_w. Extend the system of differential equations to include the state $x_3 = C_w$ with

$$\frac{\mathrm{d}}{\mathrm{d}t}x_3 = 0 + w_3$$

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

and the process noise component w_3. Use the Euler method to determine the discrete-time system. Assume that the nominal values or initial conditions of the variables C_w, x_1, and x_2 are normally distributed. The corresponding values of the means and variances can be found in the following Table 2.1. For the simulation, assume the following values: $C_w = 0.6$, $x_1(0) = 39\,500\,\mathrm{m}$, and $x_2(0) = -10\,\mathrm{m/s}$.

Variable	Mean	Variance
C_w	0.5	1
$x_1(0)$	$39 \cdot 10^3\,\mathrm{m}$	$1 \cdot 10^4\,\mathrm{m}^2$
$x_2(0)$	$0\,\mathrm{m/s}$	$1\,\mathrm{m}^2/\mathrm{s}^2$

Table 2.1: Means and variances of the parameters and initial conditions.

Exercise 2.16. An Extended Kalman Filter is to be used for position determination of a vehicle in a two-dimensional space (o-axis: East coordinate, n-axis: North coordinate). Several measuring stations with coordinates (O_i, N_i), $i = 1, \ldots, M$ measure the distance to the vehicle. The acceleration of the vehicle in the North and East directions is modeled by white noise. The system of difference equations

$$\underbrace{\begin{bmatrix} o_{k+1} \\ n_{k+1} \\ o_{v,k+1} \\ n_{v,k+1} \end{bmatrix}}_{\mathbf{x}_{k+1}} = \underbrace{\begin{bmatrix} 1 & 0 & T_a & 0 \\ 0 & 1 & 0 & T_a \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{\Phi}} \underbrace{\begin{bmatrix} o_k \\ n_k \\ o_{v,k} \\ n_{v,k} \end{bmatrix}}_{\mathbf{x}_k} + \underbrace{\begin{bmatrix} w_{1,k} \\ w_{2,k} \\ w_{3,k} \\ w_{4,k} \end{bmatrix}}_{\mathbf{w}_k}$$

describes the vehicle behavior, where o_k and n_k and $o_{v,k}$ and $n_{v,k}$ denote the coordinates and velocities of the vehicle relative to the origin of a fixed coordinate system in the East and North directions at time kT_a with sampling time T_a, and $w_{j,k}$, $j = 1, \ldots, 4$ are the components of the process noise. Furthermore, the distance measurements of the vehicle from the stations are given by

$$y_{i,k} = \sqrt{(n_k - N_i)^2 + (o_k - O_i)^2} + v_{i,k}, \quad i = 1, \ldots, M$$

with the measurement noise $v_{i,k}$, $i = 1, \ldots, M$. Assume that all stochastic variables $(w_{j,k})$ and $(v_{i,k})$ are normally distributed, uncorrelated, and zero-mean. The sampling time is given as $T_a = 0.1\,\mathrm{s}$. For the covariance matrix of the process noise, it applies that

$$\mathrm{E}\left(\mathbf{w}_k \mathbf{w}_j^{\mathrm{T}}\right) = \mathbf{Q}\delta_{kj} \quad \text{with} \quad \mathbf{Q} = \mathrm{diag}(0, 0, 4, 4)$$

and the covariance of the measurement noise is

$$\mathrm{E}(v_{i,k} v_{i,j}) = R_i \delta_{kj} \quad \text{with} \quad R_i = 1, \quad i = 1, \ldots, M\ .$$