Appendix A
Some Fundamentals of Random Variables and Stochastic Processes

In the following, we give a brief overview of random variables and stochastic processes. It should be understood as a summary of the most important findings, which are needed in digital communications and signal processing. For some lemmas, the derivations and proofs are outlined. Beyond that, the reader is referred to dedicated textbooks such as [1, 2].

A.1 Continuous Random Variables

We start with some basics on random variables. Let X be a real-valued random variable and x an event.

A.1.1 Probability Density Function and Probability

$p(x)$ is denoted as probability density function (in short density function or density) with the property $p(x) \geq 0$ and $\int_{-\infty}^{\infty} p(x)dx = 1$. We call $F(x) = \int_{-\infty}^{x} p(u)du$ the (cumulative) distribution function. $F(b)$ is the probability of the event that the random variable is located in the interval $-\infty < X < b$ and we write with the probability operator P,

$$\mathrm{P}[X < b] = \mathrm{P}[-\infty < X < b] = \int_{-\infty}^{b} p(x)dx \tag{A.1}$$

from which we conclude

$$\mathrm{P}[a < X < b] = \int_{a}^{b} p(x)dx = F(b) - F(a) \tag{A.2}$$

Please note, if the density function $p(x)$ contains no Dirac impulses at the boarders a and b, then also the equal sign holds for the interval. With $b \to \infty$ we obtain

J. Speidel, *Introduction to Digital Communications*, Signals and Communication Technology, https://doi.org/10.1007/978-3-030-00548-1

$$\mathrm{P}[X > a] = \mathrm{P}[a < X < \infty] = \int_a^\infty p(x)dx \tag{A.3}$$

A.1.2 Two Random Variables

Joint Probability Density Function

For the two random variables X_1 and X_2 we define the (two-dimensional) joint probability density function $p_{12}(x_1, x_2)$. The density functions of the individual random variables X_i are $p_i(x_i)$, $i = 1, 2$ and called marginal probability density functions. They are calculated as

$$p_1(x_1) = \int_{-\infty}^{\infty} p_{12}(x_1, x_2)dx_2 \;;\; p_2(x_2) = \int_{-\infty}^{\infty} p_{12}(x_1, x_2)dx_1 \tag{A.4}$$

Conditional Probability Density Function

We define $p_{1/2}\,(x_1 \mid X_2 = x_2)$ or with short hand notation $p_{1/2}\,(x_1 \mid x_2)$ as the conditional probability density function of X_1 under the condition $X_2 = x_2$.

$$p_{1/2}\,(x_1 \mid x_2) = \frac{p_{12}(x_1, x_2)}{p_2(x_2)} \tag{A.5}$$

Conditional Probabilities

With the conditional densities we can calculate conditional probabilities

$$\mathrm{P}[X_1 < x_1 \mid X_2 = x_2] = \int_{-\infty}^{x_1} p_{1/2}\,(u_1 \mid x_2)\,du_1 \tag{A.6}$$

Bayes Theorem

The Bayes theorem relates the two conditional densities as

$$p_{1/2}\,(x_1 \mid x_2) = \frac{p_{2/1}\,(x_2 \mid x_1)\,p_1(x_1)}{p_2(x_2)}, \;\; p_2(x_2) \neq 0 \tag{A.7}$$

Statistical Independence of Random Variables

We call two random variables statistically independent, if and only if $p_{1/2}\,(x_1 \mid x_2) = p_1(x_1)$ is independent of x_2. Then follows from (A.5)

$$p_{12}(x_1, x_2) = p_1(x_1)p_2(x_2) \tag{A.8}$$

A.2 Statistical Parameters for Random Variables

A.2.1 Expected Value

The expected value (or in short expectation) of a real-valued random variable X with density function $p(x)$ is defined as

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x p(x) dx = m_x \tag{A.9}$$

$\mathbf{E}[X]$ is also called first moment or mean value of X. We see that $\mathbf{E}[...]$ is a linear operator.

A.2.2 Function of a Random Variable, nth Moments

Let $g(...)$ be a function of the random variable X yielding the new random variable $Y = g(X)$. Then

$$\mathbf{E}[Y] = \mathbf{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\, p(x) dx \tag{A.10}$$

holds. On that basis we can define the nth *moment* of X as

$$\mathbf{E}[X^n] = \int_{-\infty}^{\infty} x^n p(x) dx \tag{A.11}$$

and in particular for $n = 2$ we obtain $\mathbf{E}[X^2]$, which is the quadratic mean and physically the *mean power* of the random variable X.

The nth *central moment* is defined as

$$\mathbf{E}\left[(X - m_x)^n\right] = \int_{-\infty}^{\infty} (x - m_x)^n\, p(x) dx \tag{A.12}$$

which yields for $n = 2$ the *variance* of X

$$\mathrm{var}[X] = \sigma_x^2 = \mathbf{E}\left[(X - m_x)^2\right] \tag{A.13}$$

σ_x is known as *standard deviation*. It is straightforward to show that

$$\sigma_x^2 = \mathbf{E}\left[X^2\right] - m_x^2 \tag{A.14}$$

is true. If the random variable has zero mean, then the variance equals the mean power.

A.2.3 Covariance and Correlation of Two Random Variables

For two real-valued random variables X_i with $\mathbf{E}[X_i] = m_i$, the marginal density functions $p_i(x_i)$; $i = 1, 2$, and the joint density function $p_{12}(x_1, x_2)$ we define the *joint central moment* of order (k, n) as

$$\mathbf{E}\left[(X_1 - m_1)^k (X_2 - m_2)^n\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1 - m_1)^k (x_2 - m_2)^n \, p_{12}(x_1, x_2) dx_1 dx_2 \tag{A.15}$$

from which follows for $k = n = 1$ the *covariance* between X_1 and X_2 as

$$\mu_{12} = \mathbf{E}[(X_1 - m_1)(X_2 - m_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1 - m_1)(x_2 - m_2) \, p_{12}(x_1, x_2) dx_1 dx_2 \tag{A.16}$$

It is straightforward to show that

$$\mu_{12} = \mathbf{E}[X_1 X_2] - m_1 m_2 \tag{A.17}$$

holds.

The *correlation coefficient* is defined as

$$\rho_{12} = \frac{\mu_{12}}{\sigma_1 \sigma_2} = \frac{\mathbf{E}[X_1 X_2] - m_1 m_2}{\sigma_1 \sigma_2} \tag{A.18}$$

Two random variables X_1 and X_2 are called *uncorrelated,* if and only if the covariance is zero

$$\mu_{12} = 0 \tag{A.19}$$

which yields $\rho_{12} = 0$[1] and finally

$$\mathbf{E}[X_1 X_2] = m_1 m_2 \tag{A.20}$$

Consequently, if we would like to see whether two random variables are uncorrelated, we have to check, if their covariance is zero.

The *Correlation* between X_1 and X_2 is defined as

$$c_{12} = \mathbf{E}[X_1 X_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_{12}(x_1, x_2) dx_1 dx_2 \tag{A.21}$$

We call two random variables *orthogonal*, if

$$\mathbf{E}[X_1 X_2] = 0 \tag{A.22}$$

[1] This is the reason why ρ_{12} is called correlation coefficient.

Let X_1 and X_2 be two random variables each with zero mean $m_1 = m_2 = 0$. If they are uncorrelated, then their covariance (A.19) is zero and from (A.20) follows for the correlation $c_{12} = \mathbf{E}[X_1 X_2] = 0$.

Let X_1 and X_2 be *statistically independent* random variables. Then (A.8) holds and from (A.21) follows

$$\mathbf{E}[X_1 X_2] = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x_1 p_1(x_1) dx_1 \right] x_2 p_2(x_2) dx_2 = \mathbf{E}[X_1]\,\mathbf{E}[X_2] = m_1 m_2 \tag{A.23}$$

which yields $\rho_{12} = 0$. Hence, we conclude that statistically independent random variables are also uncorrelated. In general, the reverse is not true.

A.3 Stochastic Processes

A.3.1 Definition of a Stochastic Process

For engineers a stochastic process is best explained with the help of a physical experiment. Consider a large number $N \to \infty$ of identical resistors, each resistor i generating a random noise voltage $X_i(t)$ as a function of time t, where $i = 1, 2, \ldots, N$. The stochastic process $X(t) = \{X_1(t), X_2(t), \ldots, X_N(t)\}$ represents the family also called ensemble of all voltages and $X_i(t)$ is the ith sample function or ith realization of the process. All sample functions ($i = 1, 2, \ldots, N$) belonging to the process have the same statistical parameters, such as probability density function, autocorrelation etc. To characterize the stochastic process statistical parameters can be defined in two ways, namely along the time axis of a dedicated sample function $X_i(t)$ or over all sample functions of $X(t)$ at a fixed time instant t_ν. Then $X(t_\nu)$ is a continuous random variable. In our measuring campaign we can further look at the stochastic process $X(t)$ at different time instants $t_1 < t_2 < t_3 < \cdots < t_M$ yielding a sequence of random variables

$$X(t_1), X(t_2), .., X(t_M) \tag{A.24}$$

and the stochastic process $X(t)$ can be regarded for each fixed time instant t as a random variable. Consequently, the corresponding definitions for random variables can be applied to describe the statistical parameters of the stochastic process. The set of random variables in (A.24) is characterized by its joint probability density function

$$p_{1M}\left(x_{t_1}, x_{t_2}, \ldots, x_{t_M}\right) \tag{A.25}$$

where $x_{t_\nu} = x(t_\nu)$ is the shorthand notation of an event of the random variable $X(t_\nu)$; $\nu = 1, 2, \ldots, M$. In general, the probability density function depends on the time instances $t_1, t_2, \ldots, t_M$, if the process is nonstationary, see Sect. A.3.4.

A.3.2 Autocovariance, Auto-, and Cross-Correlation Function

Single Stochastic Process $X(t)$

The *autocovariance function* $\mu_{xx}(t_1, t_2)$ of a stochastic process $X(t)$ is defined similar to the covariance (A.16) of two random variables $X(t_1)$ and $X(t_2)$

$$\begin{aligned}\mu_{xx}(t_1, t_2) = & \quad \mathbf{E}\left[\left(X(t_1) - m_x(t_1)\right)\left(X(t_2) - m_x(t_2)\right)\right] = \\ & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x_{t_1} - m_x(t_1)\right)\left(x_{t_2} - m_x(t_2)\right) p_{12}\left(x_{t_1}, x_{t_2}\right) dx_{t_1} dx_{t_2}\end{aligned} \tag{A.26}$$

where $p_{12}\left(x_{t_1}, x_{t_2}\right)$ is the joint probability density function of $X(t_1)$ and $X(t_2)$.

The *expected values* are $m_x(t_i) = \mathbf{E}\left[X(t_i)\right]\ ;\ i = 1, 2$. It is straightforward to show that

$$\mu_{xx}(t_1, t_2) = \mathbf{E}\left[X(t_1)X(t_2)\right] - m_x(t_1)m_x(t_2) \tag{A.27}$$

The *autocorrelation function* of the process $X(t)$ is defined similar to (A.21) as

$$R_{xx}(t_1, t_2) = \mathbf{E}\left[X(t_1)X(t_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_{t_1} x_{t_2} p_{12}\left(x_{t_1}, x_{t_2}\right) dx_{t_1} dx_{t_2} \tag{A.28}$$

Two Stochastic Processes $X(t)$ and $Y(t)$

We consider two stochastic processes $X(t)$ and $Y(t)$ with the corresponding random variables

$$X(t_i)\ ;\ i = 1, 2, \ldots, M_x\ ;\ t_1 < t_2 < t_3 \ldots \tag{A.29}$$

$$Y(\tilde{t_j})\ ;\ j = 1, 2, \ldots, M_y\ ;\ \tilde{t_1} < \tilde{t_2} < \tilde{t_3} \ldots \tag{A.30}$$

The joint set of random variables is characterized by the joint probability density function

$$p_{xy}\left(x_{t_1}, x_{t_2}, \ldots, x_{t_{M_x}}; y_{\tilde{t_1}}, y_{\tilde{t_2}}, \ldots, y_{\tilde{t}_{M_y}}\right) \tag{A.31}$$

where $x_{t_\nu} = x(t_\nu)$ and $y_{\tilde{t_\nu}} = y(\tilde{t_\nu})$. Then the *cross-correlation function* can be defined as

$$R_{xy}(t_1, t_2) = \mathbf{E}\left[X(t_1)Y(t_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_{t_1} y_{t_2} p_{xy}\left(x_{t_1}, y_{t_2}\right) dx_{t_1} dy_{t_2} \tag{A.32}$$

where we have renamed $\tilde{t_2}$ as t_2.

A.3.3 Time-Domain Parameters and Ergodicity

As already alluded, we can define parameters for a stochastic process $X(t)$ along the time axis of a dedicated sample function $X_i(t)$ or over the ensemble at a fixed time instant t_ν, yielding $X(t_\nu)$. As a consequence, we differentiate between time-domain averages (or moments) on one hand and ensemble values also called expected values on the other hand. On the basis of a sample function $X_i(t)$, in general complex, we get the following time-domain parameters:

Mean Value

$$\bar{x} = \lim_{T_0 \to \infty} \frac{1}{2T_0} \int_{-T_0}^{T_0} x_i(t) dt \tag{A.33}$$

Autocovariance Function

$$c_{xx}(\tau) = \lim_{T_0 \to \infty} \frac{1}{2T_0} \int_{-T_0}^{T_0} (x_i(t) - \bar{x})^* \, (x_i(t+\tau) - \bar{x}) \, dt \tag{A.34}$$

Autocorrelation Function

$$R_{xx}(\tau) = \lim_{T_0 \to \infty} \frac{1}{2T_0} \int_{-T_0}^{T_0} x_i^*(t) x_i(t+\tau) dt \tag{A.35}$$

For $\tau = 0$ follows the **mean power**

$$R_{xx}(0) = \lim_{T_0 \to \infty} \frac{1}{2T_0} \int_{-T_0}^{T_0} |x_i(t)|^2 \, dt \tag{A.36}$$

Ergodicity

A wide sense stationary stochastic process (see Sect. A.3.4) is called ergodic, if all statistical parameters calculated on the basis of the ensemble and with respect to time of any sample function $X_j(t)$ are identical. Thus, an ergodic process can be statistically described by just one realization. In engineering, ergodicity is often assumed as a hypothesis, because the experimental proof in many cases is difficult, although important. In the following, we exclusively consider ergodic processes and focus on the ensemble values.

A.3.4 Stationary Stochastic Process

Strict Sense Stationary (SSS) Stochastic Process

- A stochastic process $X(t)$ is "strict sense stationary" (SSS), if $X(t)$ and $X(t+a)$ $\forall a$ have the same statistical parameters. In other words, their statistics do not depend on time. This holds for all probability density functions such as (A.25) and all M,

$$p_{1M}\left(x(t_1), x(t_2), .., x(t_M)\right) = p_{1M}\left(x(t_1+a), x(t_2+a), .., x(t_M+a)\right) \quad \text{(A.37)}$$

- Two stochastic processes $X(t)$ and $Y(t)$ are jointly strict sense stationary, if the joint statistics of $X(t)$ and $Y(t)$ are equal to the joint statistics of $X(t+a)$ and $Y(t+a)$, $\forall a$, respectively.
- A complex-valued stochastic process $Z(t) = X(t) + \mathrm{j}Y(t)$ is strict sense stationary, if this condition holds jointly for the real and imaginary part.

Wide Sense Stationary (WSS) Stochastic Process

The conditions for a "wide sense stationary" (WSS) process are much weaker than for a strict sense stationary process, as they just impose conditions on the first and second order moments. Higher moments are not touched. From (A.37) follows that $p_1\left(x_{t_1}\right) = p_1\left(x_{t_1+a}\right)$ $\forall a$ with $x_{t_\nu} = x(t_\nu)$ as before. Consequently, the expected value is constant. Furthermore (A.37) results in $p_{12}\left(x_{t_1}, x_{t_2}\right) = p_{12}\left(x_{t_1+a}, x_{t_2+a}\right)$ $\forall a$ and thus the density function, the second moments and the autocorrelation function depend only on a time difference $t_2 - t_1 = \tau$.

Definition:

- A stochastic process $X(t)$ is wide sense stationary, if its expected value is constant

$$\mathbf{E}[X] = m_x = const. \quad \text{(A.38)}$$

- and if its autocorrelation function just depends on a time difference $\tau = t_2 - t_1$. Then follows from (A.28) by using $t_2 - t_1 = \tau$ and replacing t_1 by the fixed time instant t

$$R_{xx}(t_1, t_2) = R_{xx}(t, t+\tau) = \mathbf{E}\left[X(t)X(t+\tau)\right] = R_{xx}(\tau) \quad \text{(A.39)}$$

  where $R_{xx}(\tau)$ is a shorthand notation.

For a complex stochastic process, we define

$$R_{xx}(\tau) = \mathbf{E}\left[X^*(t)X(t+\tau)\right] \quad \text{(A.40)}$$

In general, the autocorrelation function exhibits the following properties:

$$R_{xx}(-\tau) = R_{xx}^{*}(\tau)\,;\; R_{xx}(0) \geq |R_{xx}(\tau)| \tag{A.41}$$

$R_{xx}(0) = \mathbf{E}[|X(t)|^2]$ is always a real value and is called the mean power of $X(t)$. We also see that a SSS stochastic process is also WSS.

A.3.5 Uncorrelated WSS Stochastic Processes

A Single Process $X(t)$

To check whether a WSS stochastic process $X(t)$ is uncorrelated, we have to extend the definition of the covariance of a random variable in (A.16) to a process.

The *autocovariance function* of the WSS process $X(t)$ is thus given by

$$C_{xx}(t, t+\tau) = C_{xx}(\tau) = \mathbf{E}\left[\left(X^{*}(t) - m_x^{*}\right)\left(X(t+\tau) - m_x\right)\right] = R_{xx}(\tau) - |m_x|^2 \tag{A.42}$$

where $m_x = \mathbf{E}[X(t)]$ is the expected value of $X(t)$.

The autocovariance function $C_{xx}(\tau)$ specifies the expected value of the product of the two random variables $X^{*}(t) - m_x^{*}$ and $X(t+\tau) - m_x$ for any given time shift $\tau \neq 0$. Similar to (A.19) we can say that the two random variables are uncorrelated, if $C_{xx}(\tau) = 0$. However, we have to exclude $\tau = 0$, because in this case both random variables just differ in the sign of the imaginary part and of course are strongly correlated. Consequently, we can formulate the following meaningful definition:

A WSS stochastic process $X(t)$ is *uncorrelated,* if its autocovariance function meets the condition

$$C_{xx}(\tau) \begin{cases} \neq 0\,;\; \tau = 0 \\ = 0\,;\; \tau \neq 0 \end{cases} \tag{A.43}$$

For the *autocorrelation function* of an *uncorrelated* process then follows with (A.43) and $R_{xx}(0) = \mathbf{E}\left[|X(t)|^2\right]$

$$R_{xx}(\tau) = \begin{cases} \mathbf{E}\left[|X(t)|^2\right]\,;\; \tau = 0 \\ m_x^2 \qquad\quad ;\; \tau \neq 0 \end{cases} \tag{A.44}$$

Now, consider a WSS process with *zero mean*, $m_x = 0$. Then we find from (A.44) that this process is *uncorrelated,* if

$$R_{xx}(\tau) = 0 \;\forall\, \tau \neq 0 \tag{A.45}$$

Two Processes $X(t)$ and $Y(t)$

The statistical interrelation between two WSS processes $X(t)$ and $Y(t)$ is defined by the *cross-covariance function* similar to (A.16)

$$C_{xy}(\tau) = \mathbf{E}\left[(X(t+\tau) - m_x)\left(Y^*(t) - m_y^*\right)\right] =$$

$$= R_{xy}(\tau) - m_x m_y^* \tag{A.46}$$

with $m_x = \mathbf{E}[X(t)]$, $m_y = \mathbf{E}[Y(t)]$, and the *cross-correlation function*

$$R_{xy}(\tau) = \mathbf{E}\left[X(t+\tau)Y^*(t)\right] \tag{A.47}$$

Using similar arguments as before with $C_{xx}(\tau)$, but no exception for $\tau = 0$ is required here, we define:

Two (WSS) processes $X(t)$ and $Y(t)$ are *uncorrelated* if

$$C_{xy}(\tau) = 0 \;\forall\, \tau \tag{A.48}$$

Then follows from (A.46)

$$R_{xy}(\tau) = m_x m_y^* \;\forall\, \tau \tag{A.49}$$

If at least one stochastic process has zero mean, then the processes are referred to as *orthogonal*

$$R_{xy}(\tau) = 0 \;\forall\, \tau \tag{A.50}$$

It is straightforward to show that the cross-correlation function has the following symmetry property:

$$R_{xy}(\tau) = R_{yx}^*(-\tau) \tag{A.51}$$

A.3.6 Statistically Independent Processes

Two stochastic processes $X(t)$ and $Y(t)$ are statistically independent, if and only if for any choice of t_i and $\tilde{t}_j$ as well as M_x and M_y

$$p_{xy}\left(x_{t_1}, x_{t_2}, \ldots, x_{t_{M_x}}; y_{\tilde{t}_1}, y_{\tilde{t}_2}, \ldots, y_{\tilde{t}_{M_y}}\right) = p_x\left(x_{t_1}, x_{t_2}, \ldots, x_{t_{M_x}}\right) p_y\left(y_{\tilde{t}_1}, y_{\tilde{t}_2}, \ldots, y_{\tilde{t}_{M_y}}\right) \tag{A.52}$$

holds, where $p_x\left(x_{t_1}, x_{t_2}, \ldots, x_{t_{M_x}}\right)$ and $p_y\left(y_{\tilde{t}_1}, y_{\tilde{t}_2}, \ldots, y_{\tilde{t}_{M_y}}\right)$ are the joint density functions of $x_{t_1}, x_{t_2}, \ldots, x_{t_{M_x}}$ and $y_{\tilde{t}_1}, y_{\tilde{t}_2}, \ldots, y_{\tilde{t}_{M_y}}$, respectively. From the statistical independence follows that the two processes are uncorrelated, but not reversely.

Two WSS stochastic processes $X(t)$ and $Y(t)$ with joint probability density function $p_{xy}(x, y)$ are *statistically independent*, if and only if

$$p_{xy}(x, y) = p_x(x) p_y(y) \tag{A.53}$$

where $p_x(x)$ and $p_y(y)$ are the marginal probability density functions of the stochastic processes $X(t)$ and $Y(t)$, respectively.

A.4 Stochastic Processes and Linear Time-Invariant Systems

A.4.1 Input–Output Relation of Linear System in Time Domain

Let $h(t)$ be the (deterministic) impulse response of a linear time-invariant system. At its input the sample function $x(t)$ of a WSS stochastic process $X(t)$ is active. Throughout the following we always consider stationary processes. Then the output process $Y(t)$ with sample function $y(t)$ is also stationary [1] and given by the convolution

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(u)h(t-u)du \tag{A.54}$$

However, as the stochastic signals cannot be expressed by a mathematical formula, we are not in a position to explore this equation further and have to find a statistical description using autocorrelation functions and power spectral densities.

A.4.2 Wiener-Lee Theorem for Input–Output Autocorrelation functions

Let $R_{xx}(\tau)$ and $R_{yy}(\tau)$ be the autocorrelation functions of the input and output stochastic process $x(t)$ and $y(t)$, respectively. We calculate the autocorrelation function $R_{hh}(\tau)$ of the deterministic (and thus also ergodic) impulse response $h(t)$ according to (A.35)

$$R_{hh}(\tau) = \mathbf{E}\left[h^*(t)h(t+\tau)\right] = \int_{-\infty}^{\infty} h^*(t)h(t+\tau)dt \tag{A.55}$$

where we have dropped $\lim_{T_0\to\infty} \frac{1}{2T_0}$, because $h(t)$ is a deterministic signal with finite energy $R_{hh}(0) = \int_{-\infty}^{\infty} |h(t)|^2\, dt$. The integral for $R_{hh}(\tau)$ can be considered as the convolution between $h(\tau)$ and $h^*(-\tau)$

$$R_{hh}(\tau) = h(\tau) * h^*(-\tau) \tag{A.56}$$

The Wiener-Lee theorem describes the relation between the input and the output autocorrelation function of a linear time-invariant system as follows:

$$R_{yy}(\tau) = R_{hh}(\tau) * R_{xx}(\tau) = h(\tau) * h^*(-\tau) * R_{xx}(\tau) \tag{A.57}$$

A.4.3 Wiener–Khintchine Theorem for Power Spectral Density

For communications engineers spectra of signals are important to get an idea about the required bandwidth. For deterministic signals with finite energy the Fourier spectrum exists according to the sufficient Dirichlet condition. However, a random signal $x(t)$ has infinite energy, because in general

$$\lim_{T_0 \to \infty} \int_{-T_0}^{T_0} |x(t)|^2 \, dt \to \infty \tag{A.58}$$

On the other hand an ergodic stochastic process $X(t)$ exhibits finite mean power which is

$$\mathbf{E}\left[|X|^2\right] = R_{xx}(0) = \lim_{T_0 \to \infty} \frac{1}{2T_0} \int_{-T_0}^{T_0} |x(t)|^2 \, dt < \infty \tag{A.59}$$

The Wiener–Khintchine theorem provides the power spectral density $S_{xx}(f)$ of $X(t)$ by means of the Fourier transform of the autocorrelation function of $X(t)$,

$$R_{xx}(\tau) \rightarrowtail S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) \mathrm{e}^{-\mathrm{j}2\pi f \tau} d\tau \tag{A.60}$$

From the symmetry $R_{xx}(-\tau) = R_{xx}^*(\tau)$ follows the property that $S_{xx}(f)$ is real and moreover

$$S_{xx}(f) \geq 0 \tag{A.61}$$

holds. With $h(t) \rightarrowtail H(f)$ and $h^*(-t) \rightarrowtail H^*(f)$ follows from (A.56) with the Fourier transform

$$R_{hh}(\tau) \rightarrowtail S_{hh}(f) = |H(f)|^2 \tag{A.62}$$

and with $R_{yy}(\tau) \rightarrowtail S_{yy}(f)$ we obtain the power spectral density of the output process $Y(t)$ with (A.57)

$$S_{yy}(f) = |H(f)|^2 \, S_{xx}(f) \tag{A.63}$$

With the inverse Fourier transform we obtain the from (A.60)

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} S_{xx}(f) \mathrm{e}^{\mathrm{j}2\pi f \tau} df \tag{A.64}$$

and for $\tau = 0$ the mean power of $X(t)$

$$R_{xx}(0) = \int_{-\infty}^{\infty} S_{xx}(f) df \tag{A.65}$$

Example: White noise $X(t)$ is defines by its constant power spectral density,

$$S_{xx}(f) = a = const. \; \forall \, f \tag{A.66}$$

Consequently the autocorrelation function is

$$R_{xx}(\tau) = a\delta(\tau) \tag{A.67}$$

and we see that $R_{xx}(\tau) = 0 \; \forall \, \tau \neq 0$. Thus, all pairs of random variables $X(t)$ and $X(t+\tau)$ are uncorrelated for $\tau \neq 0$. We observe that a is also the mean power of $X(t)$.

A.5 Modulation and Demodulation of a Stationary Stochastic Process

A.5.1 Modulation

We consider a WSS stationary process $X(t)$, which shall be modulated with the carrier $\mathrm{e}^{\mathrm{j}2\pi f_0 t}$. Then we obtain

$$Y(t) = X(t)\mathrm{e}^{\mathrm{j}2\pi f_0 t} \tag{A.68}$$

$X(t)$ shall have the autocorrelation function $R_{xx}(\tau) = \mathbf{E}\left[X^*(t)X(t+\tau)\right]$. For the autocorrelation function of the process $Y(t)$ we obtain

$$R_{yy}(\tau) = \mathbf{E}\left[Y^*(t)Y(t+\tau)\right] = \mathbf{E}\left[X(t+\tau)X^*(t)\mathrm{e}^{\mathrm{j}2\pi f_0 \tau}\right] \tag{A.69}$$

Noting that $\mathrm{e}^{\mathrm{j}2\pi f_0 \tau}$ is deterministic yields the final result

$$R_{yy}(\tau) = R_{xx}(\tau)\mathrm{e}^{\mathrm{j}2\pi f_0 \tau} \tag{A.70}$$

This shows that the modulation of a stationary stochastic process translates into the modulation of its autocorrelation function. With the frequency shifting property of the Fourier transform we obtain

$$R_{yy}(\tau) \rightarrowtail S_{yy}(f) = S_{xx}(f - f_0) \tag{A.71}$$

outlining that the modulation of $X(t)$ results in a frequency shift of its power spectral density $S_{xx}(f)$ by the carrier frequency f_0.

A.5.2 Demodulation

We consider the modulated stationary process $Y(t)$ and apply the synchronous demodulation with the carrier $\mathrm{e}^{-\mathrm{j}2\pi f_0 t}$ resulting in the demodulated stochastic process $Z(t)$,

$$Z(t) = Y(t)\mathrm{e}^{-\mathrm{j}2\pi f_0 t} = X(t) \tag{A.72}$$

Consequently, we obtain

$$R_{zz}(\tau) = R_{yy}(\tau)\mathrm{e}^{-\mathrm{j}2\pi f_0 \tau} = R_{xx}(\tau) \tag{A.73}$$

and

$$S_{zz}(f) = S_{yy}(f + f_0) = S_{xx}(f) \tag{A.74}$$

We also see that modulation and demodulation does not change the mean power of the processes, because with $\mathbf{E}[|X|^2] = R_{xx}(0)$, $\mathbf{E}[|Y|^2] = R_{yy}(0)$, and $\mathbf{E}[|Z|^2] = R_{zz}(0)$ follows

$$\mathbf{E}\left[|X|^2\right] = \mathbf{E}\left[|Y|^2\right] = \mathbf{E}\left[|Z|^2\right]. \tag{A.75}$$

A.6 Stationary, Real-Valued Bandpass Process

As is well known [3] any real-valued bandpass signal can be written in general as

$$n(t) = x(t)\cos(2\pi f_0 t) - y(t)\sin(2\pi f_0 t) \tag{A.76}$$

where $x(t)$ and $y(t)$ are real-valued lowpass signals with cut-off frequency f_c. This model shall be adopted to a stochastic bandpass process $N(t)$ with power spectral density $S_{nn}(f)$ and with the passband in the range of $f_0 - f_c \leq |f| \leq f_0 + f_c$, where $f_0 > f_c$ is the center frequency. $X(t)$ and $Y(t)$ shall be WSS lowpass processes with the power spectral densities

$$S_{xx}(f)\ ;\ S_{yy}(f)\ \begin{cases} \neq 0\ ;\ |f| \leq f_c \\ = 0\ ; \quad \text{else} \end{cases} \tag{A.77}$$

A.6.1 Condition for Stationarity

We would like to know under which conditions this bandpass process $N(t)$ is WSS stationary. Therefore we have to check whether $\mathbf{E}[N(t)] = const.$ holds and whether the autocorrelation function

$$R_{nn}(\tau) = \mathbf{E}\left[N(t)N(t+\tau)\right] \tag{A.78}$$

is independent of t.

Expected Value

With(A.76) follows

$$\mathbf{E}[N(t)] = \mathbf{E}[X(t)]\cos(2\pi f_0 t) - \mathbf{E}[Y(t)]\sin(2\pi f_0 t) \tag{A.79}$$

$\mathbf{E}[N(t)] = const.\ \forall t$ holds, if

$$\mathbf{E}[X(t)] = \mathbf{E}[Y(t)] = 0 \tag{A.80}$$

Consequently from (A.79) also follows

$$\mathbf{E}[N(t)] = 0 \tag{A.81}$$

Autocorrelation Function

Next, please consider (A.78). By applying basic trigonometric formulas and using the fact, that terms with sin(), cos(), and the arguments $2\pi f_0 t$ and $2\pi f_0(t+\tau)$ are nonrandom and therefore can be taken out from the expectation operator, we obtain finally with (A.76)

$$\begin{aligned} R_{nn}(\tau) = \quad & \tfrac{1}{2}\left[R_{xx}(\tau) + R_{yy}(\tau)\right]\cos(2\pi f_0\tau) + \\ & + \tfrac{1}{2}\left[R_{xx}(\tau) - R_{yy}(\tau)\right]\cos(4\pi f_0 t + 2\pi f_0\tau) - \\ & - \tfrac{1}{2}\left[R_{xy}(\tau) - R_{yx}(\tau)\right]\sin(2\pi f_0\tau) - \\ & - \tfrac{1}{2}\left[R_{xy}(\tau) + R_{yx}(\tau)\right]\sin(4\pi f_0 t + 2\pi f_0\tau) \end{aligned} \tag{A.82}$$

To get $R_{nn}(\tau)$ independent of t the following conditions must hold:

$$R_{xx}(\tau) = R_{yy}(\tau) \tag{A.83}$$

and

$$R_{xy}(\tau) = -R_{yx}(\tau) \tag{A.84}$$

Then we obtain from (A.82)

$$R_{nn}(\tau) = R_{xx}(\tau)\cos(2\pi f_0\tau) - R_{xy}(\tau)\sin(2\pi f_0\tau) \tag{A.85}$$

Knowing that an autocorrelation function provides the mean power of the process for $\tau = 0$ we conclude from (A.85) and (A.83)

$$\mathbf{E}\left[|N(t)|^2\right] = \mathbf{E}\left[|X(t)|^2\right] = \mathbf{E}\left[|Y(t)|^2\right] \tag{A.86}$$

Furthermore, we can find another property by applying (A.51) on (A.84) yielding

$$R_{yx}(-\tau) = -R_{yx}(\tau) \tag{A.87}$$

which indicates that $R_{yx}(\tau)$ is an odd function. Consequently, $R_{yx}(0) = 0$ must be true and the property (A.84) yields

$$R_{xy}(0) = 0 \tag{A.88}$$

This means that the random variables $X(t)$ and $Y(t)$ for any given t are not correlated. Please note that $R_{xy}(0) = 0$ does not require $X(t+\tau)$ and $Y(t)$ to be uncorrelated for any τ. However, if the zero mean processes $X(t)$ and $Y(t)$ are assumed to be uncorrelated for any τ, $R_{xy}(\tau) = 0\ \forall\,\tau$ holds and from (A.85) follows

$$R_{nn}(\tau) = R_{xx}(\tau)\cos(2\pi f_0\tau) \tag{A.89}$$

Using the Wiener–Khintchine theorem (A.60) and the frequency shifting property of the Fourier transform we obtain from (A.89) the power spectral density of the process $N(t)$

$$S_{nn}(f) = \frac{1}{2}\left[S_{xx}(f - f_0) + S_{xx}(f + f_0)\right] \tag{A.90}$$

which clearly exhibits a bandpass shape.

A.6.2 Summary on Stationary Bandpass Process

A WSS bandpass process $N(t)$ exhibits zero mean and is composed of the in-phase component $X(t)$ and the quadrature component $Y(t)$, which are zero mean WSS lowpass processes. Moreover, $N(t)$, $X(t)$, and $Y(t)$ have the same mean power. For the cross-correlation holds $R_{xy}(0) = \mathbf{E}[X(t)Y(t)] = 0$, which means that the random variables $X(t)$ and $Y(t)$ are assumed to be uncorrelated for any fixed t.

A.6.3 Complex Envelope of a Bandpass Process

It is straightforward to show that (A.76) can be written as

$$N(t) = \mathrm{Re}\left[Z(t)\mathrm{e}^{\mathrm{j}2\pi f_0 t}\right] \tag{A.91}$$

where

$$Z(t) = X(t) + \mathrm{j}Y(t) \tag{A.92}$$

is called the complex envelope. If $X(t)$ and $Y(t)$ are WSS lowpass processes, then $Z(t)$ is a WSS complex lowpass process. It is easy to show that for the autocorrelation function of $Z(t)$ follows with (A.83) and (A.84)

$$R_{zz}(\tau) = \mathbf{E}\left[Z^*(t)Z(t+\tau)\right] = 2\left[R_{xx}(\tau) - \mathrm{j}R_{xy}(\tau)\right]. \tag{A.93}$$

A.7 Two-Dimensional Gaussian Process

A.7.1 Joint Gaussian Probability Density Function

We consider now two real-valued SSS Gaussian processes $X(t)$ and $Y(t)$. For any fixed t they represent random variables [1] with the Gaussian joint probability density function

$$p_{xy}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{(x-m_x)^2\sigma_y^2 - 2(x-m_x)(y-m_y)\rho\sigma_x\sigma_y + (y-m_y)^2\sigma_x^2}{2\sigma_x^2\sigma_y^2(1-\rho^2)}} \tag{A.94}$$

with the mean values m_x and m_y,
the variances

$$\sigma_x^2 = \mathbf{E}[X^2] - m_x^2 \;;\; \sigma_y^2 = \mathbf{E}[Y^2] - m_y^2 \tag{A.95}$$

the normalized cross-covariance (correlation coefficient)

$$\rho = \frac{R_{xy}(0) - m_x m_y}{\sigma_x\sigma_y} \tag{A.96}$$

and the marginal probability density function of $X(t)$

$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \mathrm{e}^{-\frac{(x-m_x)^2}{2\sigma_x^2}} \tag{A.97}$$

as well as of $Y(t)$

$$p_y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \mathrm{e}^{-\frac{(y-m_y)^2}{2\sigma_y^2}} \tag{A.98}$$

A.7.2 Uncorrelated Gaussian Random Processes

Let $X(t)$ and $Y(t)$ be two WSS, real and uncorrelated Gaussian processes. Then $R_{xy}(\tau) = m_x m_y$ holds according to (A.49). This is valid for any τ, including $\tau = 0$. With $R_{xy}(0) = m_x m_y$ follows from (A.96)

$$\rho = 0 \tag{A.99}$$

and consequently from (A.94) with (A.97) and (A.98) follows

$$p_{xy}(x, y) = p_x(x)p_y(y) \tag{A.100}$$

Hence, we conclude that uncorrelated Gaussian processes $X(t)$ and $Y(t)$ are even statistically independent.

A.7.3 Complex Gaussian Random Process

Let $X(t)$ and $Y(t)$ be WSS real-valued Gaussian lowpass processes with properties given in Sect. A.7.1. Then they constitute a complex Gaussian random lowpass process

$$Z(t) = X(t) + \mathrm{j}Y(t) \tag{A.101}$$

A.7.4 Gaussian Bandpass Process

Any real-valued bandpass process in general is given by (A.76)

$$n(t) = x(t)\cos(2\pi f_0 t) - y(t)\sin(2\pi f_0 t)$$

If $X(t)$ and $Y(t)$ are stationary Gaussian lowpass processes with properties given in Sect. A.7.1, we denote $N(t)$ as a Gaussian bandpass process. We know from Sect. A.6 that $N(t)$, $X(t)$, and $Y(t)$ have zero mean and the same mean power. Furthermore, if the two Gaussian lowpass processes $X(t)$ and $Y(t)$ are uncorrelated then they are even statistically independent.

A.8 Sampling of a Stochastic Process

The received signal in a digital communication system is a random process. Before detection, the signal is sampled. In this section, the main basics for sampling of a stochastic process are summarized.

A.8.1 Prerequisites

$X(t)$ shall be a WSS stochastic process. Then $X(t)$ and $X(t + a)$ $\forall a$ have the same density function, because $p_x(x)$ is independent of t. Consequently, after sampling

with a sampling frequency $f_S = \frac{1}{T}$ the resulting samples $X(kT)$ exhibit the same probability density function $p_x(x)$ as $X(t)$. Furthermore, $X(kT) = X_S(k)$ can be considered as a sequence of equidistant random variables of the process $X(t)$, which constitute a stationary discrete-time stochastic process, where $k \epsilon \mathbb{Z}$ is the discrete-time.

A.8.2 Auto- and Cross-Correlation Function of a Discrete-Time Stochastic Process

$X(kT)$ for every fixed $k \epsilon \mathbb{Z}$ can be considered as a random variable. Consequently, we apply (A.40) to get the autocorrelation function as $R_{xx}(\tau) = \mathbf{E}[X^*(kT)X(kT + \tau)]$. Obviously, $X(kT) = X_S(k)$ is only defined for discrete-time arguments. Therefore $R_{xx}(\tau)$ can take on defined values only for $\tau = lT$ with $l \epsilon \mathbb{Z}$ and the autocorrelation function (autocorrelation sequence) of $X_S(k)$ will become a function of a discrete variable to be be written as

$$R_{x_S x_S}(l) = \mathbf{E}[X_S^*(k)X_S(k + l)] \tag{A.102}$$

Interestingly

$$R_{x_S x_S}(l) = R_{xx}(lT) \tag{A.103}$$

can also be considered as the sampled version of the continuous-time autocorrelation function $R_{xx}(\tau)$. It is straightforward to show that the properties of (A.41) hold similarly as

$$R_{x_S x_S}(-l) = R^*_{x_S x_S}(l) \; ; \; R_{x_S x_S}(0) \geq \left| R_{x_S x_S}(l) \right| \tag{A.104}$$

With the same arguments as for (A.47) we can define the *cross-correlation function* of two discrete-time WSS stochastic processes $X_S(k)$ and $Y_S(k)$ as

$$R_{x_S y_S}(l) = \mathbf{E}[X_S(k + l)Y_S^*(k)] \tag{A.105}$$

Symmetry properties and conditions for uncorrelated processes are similar as for the continuous-time processes.

A.8.3 Power Spectral Density

According to the Wiener–Khintchine theorem we find the power density spectrum $S_{x_S x_S}(f)$ of the WSS stochastic process $X_S(k)$ by the Fourier transform of the sampled autocorrelation function. Applying ideal sampling on $R_{xx}(\tau)$ yields[2]

[2]We multiply the Dirac impulses by T to ensure that the autocorrelation functions on both sides of the equation have the same physical dimension.

$$R_{xx,S}(\tau) = R_{xx}(\tau) \sum_{l=-\infty}^{\infty} T\delta(\tau - lT) \tag{A.106}$$

With the Fourier correspondence

$$\sum_{l=-\infty}^{\infty} T\delta(\tau - lT) \;\rightarrowtail\; \sum_{m=-\infty}^{\infty} \delta\left(f - m\frac{1}{T}\right) \tag{A.107}$$

and with $R_{xx}(\tau) \;\rightarrowtail\; S_{xx}(f)$ we obtain from (A.106)

$$R_{xx,S}(\tau) \;\rightarrowtail\; S_{x_S x_S}(f) = S_{xx}(f) * \sum_{m=-\infty}^{\infty} \delta\left(f - m\frac{1}{T}\right)$$

which results after executing the convolution integral in

$$S_{x_S x_S}(f) = \sum_{m=-\infty}^{\infty} S_{xx}\left(f - m\frac{1}{T}\right) \tag{A.108}$$

We see that the spectrum is a periodic repetition of the baseband power spectral density $S_{xx}(f)$, where the period is given by the sampling frequency $\frac{1}{T}$.

Appendix B
Some Fundamentals of Linear Algebra

B.1 Eigenvalue Decomposition

In this Section, we review some properties of the eigenvalue decomposition of a matrix $\mathbf{A}$, assuming for the moment that such a decomposition shall exist for the given matrix. The eigenvalue–eigenvector problem of linear algebra can be stated as follows: Given a NxN matrix $\mathbf{A} \epsilon \mathbb{C}^{N\text{x}N}$ with in general complex entries a_{ij}, a column vector $\mathbf{v}_i \epsilon \mathbb{C}^{N\text{x}1}$, and a scalar factor λ_i. We are looking for the vector $\mathbf{A}\mathbf{v}_i$, which is equal to the vector $\lambda_i \mathbf{v}_i$

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i \ ; \ i = 1, \ldots, N \tag{B.1}$$

with $\mathbf{v}_i \neq \mathbf{0}$, otherwise we would have the trivial solution, which is of no interest. A matrix can change the length (by its determinant) and the direction of the vector after multiplication. Thus, we are looking for the vector $\mathbf{A}\mathbf{v}_i$ with the same direction as $\mathbf{v}_i$ but with the length changed by λ_i. The nontrivial solutions λ_i and $\mathbf{v}_i$ are called eigenvalues and eigenvectors of the matrix, respectively. The set of all eigenvalues is denoted as spectrum and the absolute value of the largest eigenvalue is referred to as spectral radius. We can rewrite (B.1) with matrix notation as

$$\mathbf{A}\mathbf{V} = \mathbf{V}\Lambda \tag{B.2}$$

where

$$\mathbf{V} = \left(\mathbf{v}_1 \ \mathbf{v}_2 \ \ldots \ \mathbf{v}_N \right) \epsilon \mathbb{C}^{N\text{x}N} \tag{B.3}$$

is the matrix of eigenvectors and

$$\mathbf{\Lambda} = \text{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_N\right) \tag{B.4}$$

is a diagonal matrix composed of the eigenvalues of $\mathbf{A}$. To solve the eigenvalue–eigenvector problem (B.1) can be written as

J. Speidel, *Introduction to Digital Communications*, Signals and Communication Technology, https://doi.org/10.1007/978-3-030-00548-1

$$\mathbf{A}\mathbf{v}_i - \lambda_i \mathbf{v}_i = \mathbf{0} \iff (\mathbf{A} - \lambda_i \mathbf{I}_N)\,\mathbf{v}_i = \mathbf{0}\ ;\ i = 1, \ldots, N \tag{B.5}$$

This system of homogeneous equations has a nontrivial solution only if

$$\det(\mathbf{A} - \lambda_i \mathbf{I}_N) = 0\ ;\ i = 1, \ldots, N \tag{B.6}$$

with the NxN identity matrix $\mathbf{I}_N$. Equation (B.6) is called characteristic equation for the matrix $\mathbf{A}$ and the left-hand side is the characteristic polynomial with degree N as a function of λ_i. We can conclude that the eigenvalues of the matrix $\mathbf{A}$ are the roots of the characteristic polynomial. After all λ_i are calculated from (B.6) we can insert each into (B.5) and find the corresponding eigenvectors $\mathbf{v}_i$. Note that the solution for each $\mathbf{v}_i$ contains at least one free parameter, because (B.5) is a homogeneous system of equation and thus the rank of the matrix $\mathbf{A} - \lambda_i \mathbf{I}_N$ is

$$\operatorname{rank}(\mathbf{A} - \lambda_i \mathbf{I}_N) \leq N - 1 \tag{B.7}$$

The free parameters have to be used to normalize all eigenvectors such that

$$\mathbf{v}_i^H \mathbf{v}_i = \|\mathbf{v}_i\|^2 = 1\ ;\ i = 1, \ldots, N \tag{B.8}$$

Please note that the resulting eigenvectors associated to different eigenvalues are non-orthogonal in general, i.e.,

$$\mathbf{v}_i^H \mathbf{v}_j = 0\ ;\ i = 1, \ldots, N\ ;\ i \neq j \tag{B.9}$$

does not hold. Equation (B.9) is true for Hermiteian and symmetric matrices with real entries, see Sect. B.3. On the other hand we will see later that the singular value decomposition of a matrix yields pairwise orthogonal eigenvectors. The components of the eigenvectors are in general complex, where $\mathbf{v}_i^H = \left(\mathbf{v}_i^*\right)^T = \left(\mathbf{v}_i^T\right)^*$ is the conjugate transpose vector (Hermiteian vector) to $\mathbf{v}_i$. The inverse matrix $\mathbf{V}^{-1}$ is obtained from the relation

$$\mathbf{V}^{-1}\mathbf{V} = \mathbf{V}\mathbf{V}^{-1} = \mathbf{I}_N \tag{B.10}$$

Taking $\mathbf{V}$ we can now transform $\mathbf{A}$ into diagonal form $\mathbf{\Lambda}$. The procedure is also called principal axis transformation or eigenvalue decomposition of $\mathbf{A}$. For that purpose we multiply (B.2) from the left by $\mathbf{V}^{-1}$and obtain with (B.10)

$$\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{\Lambda} \tag{B.11}$$

$\mathbf{V}$ is therefore called transform matrix. We see that there is no need for orthogonal eigenvectors. If we multiply (B.11) in a first step from the left with $\mathbf{V}$ and secondly from the right with $\mathbf{V}^{-1}$ we get

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \tag{B.12}$$

which is another form of the eigenvalue decomposition or diagonalization of $\mathbf{A}$. Such an eigenvalue decomposition will not exist for all square matrices, in other words not all square matrices are diagonalizable. In any case the inverse matrix $\mathbf{V}^{-1}$ must exist. Alternative formulation are: $\mathbf{V}$ must be a non-singular matrix or all eigenvectors $\mathbf{v}_i$ are linearly independent of each other. In the following, let us pinpoint some special NxN matrices, which are diagonalizable, and which are of importance for MIMO systems.

B.2 Normal Matrices

By definition $\mathbf{A}$ is a normal matrix, if and only if

$$\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H \tag{B.13}$$

$\mathbf{A}^H$ is called the Hermiteian[3] or conjugate transpose matrix with respect to $\mathbf{A}$. The Hermiteian operator $(...)^H$ is defined as

$$\mathbf{A}^H = \left(\mathbf{A}^*\right)^T = (\mathbf{A}^T)^* \tag{B.14}$$

It can be shown [4] that every normal matrix is unitarily diagonalizable, i.e., an eigenvalue decomposition exists, where the transform matrix $\mathbf{V}$ is a unitary matrix defined by

$$\mathbf{V}^H\mathbf{V} = \mathbf{V}\mathbf{V}^H = \mathbf{I}_N \iff \mathbf{V}^H = \mathbf{V}^{-1} \tag{B.15}$$

Then from (B.11) follows

$$\mathbf{\Lambda} = \mathbf{V}^H\mathbf{A}\mathbf{V} \tag{B.16}$$

It should be noted that (B.13) is only a sufficient and not a necessary condition for diagonalizable matrices. That means there are diagonalizable matrices which are not normal. An example is the matrix $\begin{pmatrix} 0 & 1 \\ 4 & 0 \end{pmatrix}$. Please note that the Hermiteian operator can also be applied to non-square matrices $\mathbf{A}\epsilon\,\mathbb{C}^{M\text{x}N}$ and $\mathbf{B}\epsilon\,\mathbb{C}^{N\text{x}M}$ with the property similar to the transposition operation

$$(\mathbf{A}\mathbf{B})^H = \mathbf{B}^H\mathbf{A}^H \tag{B.17}$$

[3] Charles Hermite, French mathematician.

B.3 Hermiteian Matrices

Definition of a Hermiteian Matrix

By definition, $\mathbf{A} \epsilon \mathbb{C}^{N\mathrm{x}N}$ is called a Hermiteian matrix, if and only if

$$\mathbf{A}^H = \mathbf{A} \tag{B.18}$$

It is easy to show that a Hermiteian matrix is also a normal matrix. For the proof, we insert (B.18) into (B.13) and obtain

$$\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A} = \mathbf{A}\mathbf{A}^H \tag{B.19}$$

As a consequence, Hermiteian matrices are also unitarily diagonalizable given by (B.16).

Quadratic Form

Let $\mathbf{A} \epsilon \mathbb{C}^{N\mathrm{x}N}$ be a Hermiteian matrix with eigenvalues $\lambda_i \ ; \ i = 1, \ldots, N$. Then this matrix can be defined as

$$\mathbf{A} = \mathbf{a}\mathbf{a}^H \tag{B.20}$$

with the column vector $\mathbf{a} \epsilon \mathbb{C}^{N\mathrm{x}1}$. With the column vector $\mathbf{z} \epsilon \mathbb{C}^{N\mathrm{x}1}$ we define the quadratic form

$$\mathbf{z}^H\mathbf{A}\mathbf{z} \tag{B.21}$$

which has the property

$$\mathbf{z}^H\mathbf{A}\mathbf{z} \geq 0 \ \forall \, \mathbf{z} \neq 0 \tag{B.22}$$

For the proof we insert (B.20) into (B.22) and find

$$\mathbf{z}^H\mathbf{A}\mathbf{z} = \mathbf{z}^H\mathbf{a}\mathbf{a}^H\mathbf{z} = \mathbf{z}^H\mathbf{a}\left(\mathbf{z}^H\mathbf{a}\right)^H = \left|\mathbf{z}^H\mathbf{a}\right|^2 \geq 0 \ \forall \, \mathbf{z} \neq 0 \tag{B.23}$$

Eigenvalues of a Hermiteian Matrix

All eigenvalues λ_i of a Hermiteian matrix are positive, i.e.,

$$\lambda_i \geq 0 \ ; \ i = 1, \ldots, N \tag{B.24}$$

This implies that all eigenvalues of a Hermiteian matrix are real. For the proof let $\mathbf{v}_i$ be the eigenvector associated with the eigenvalue λ_i. Then

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{B.25}$$

The corresponding positive semi-definite quadratic form according to (B.23) then is

$$\mathbf{v}_i^H \mathbf{A}\mathbf{v}_i = \mathbf{v}_i^H \lambda_i \mathbf{v}_i = \lambda_i \left\|\mathbf{v}_i\right\|^2 \geq 0 \tag{B.26}$$

from which we conclude the proposition (B.24).

Eigenvectors of a Hermiteian Matrix

Lemma

The eigenvectors $\mathbf{v}_i \in \mathbb{C}^{N\text{x}1}$ and $\mathbf{v}_j \in \mathbb{C}^{N\text{x}1}$ of a Hermiteian matrix $\mathbf{A}$ associated with two different nonzero eigenvalues $\lambda_i \neq \lambda_j \neq 0$ are (pairwise) orthogonal, i.e.,

$$\mathbf{v}_i^H \mathbf{v}_j = \mathbf{0} \; ; \; i = 1, \ldots, N \; ; \; i \neq j \tag{B.27}$$

For the proof we use the definition of the eigenvectors $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ and $\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}_j$. Then we calculate

$$\mathbf{v}_i^H \mathbf{v}_j \lambda_i = (\lambda_i \mathbf{v}_i)^H \mathbf{v}_j = (\mathbf{A}\mathbf{v}_i)^H \mathbf{v}_j = \mathbf{v}_i^H \mathbf{A}^H \mathbf{v}_j = \mathbf{v}_i^H \mathbf{A}\mathbf{v}_j = \mathbf{v}_i^H \mathbf{v}_j \lambda_j \tag{B.28}$$

and the result is $\mathbf{v}_i^H \mathbf{v}_j \lambda_i = \mathbf{v}_i^H \mathbf{v}_j \lambda_j$. As the eigenvalues are unequal and unequal to zero, proposition (B.27) follows and the proof is finished.

Please note that (B.27) also holds, if the eigenvectors are not normalized, which can be easily proven by checking (B.27) with the vectors $\alpha_i \mathbf{v}_i$ and $\alpha_j \mathbf{v}_j$.

Orthogonal and normalized vectors are called orthonormal. For the matrix $\mathbf{V}$ of eigenvectors in (B.3) then follows

$$\mathbf{V}^H \mathbf{V} = \begin{pmatrix} \mathbf{v}_1^H \\ \\ \mathbf{v}_2^H \\ \vdots \\ \\ \mathbf{v}_N^H \end{pmatrix} \left(\mathbf{v}_1 \; \mathbf{v}_2 \; \ldots \; \mathbf{v}_N \right) = \mathbf{I}_N \tag{B.29}$$

and with (B.10) we conclude that $\mathbf{V}$ is unitary, $\mathbf{V}^H = \mathbf{V}^{-1}$.

B.4 Unitary Matrices

Definition

$\mathbf{V} \in \mathbb{C}^{N\text{x}N}$ is called a unitary matrix, if and only if

$$\mathbf{V}^{-1} = \mathbf{V}^H \tag{B.30}$$

The inverse matrix $\mathbf{V}^{-1}$ is the solution of

$$\mathbf{V}^{-1}\mathbf{V} = \mathbf{V}\mathbf{V}^{-1} = \mathbf{I}_N \tag{B.31}$$

Consequently, with (B.30) follows

$$\mathbf{V}^H\mathbf{V} = \mathbf{V}\mathbf{V}^H = \mathbf{I}_N \tag{B.32}$$

$\mathbf{V}$ is composed of orthonormal column vectors.

$$\mathbf{V} = \left[\, \mathbf{v}_1 \; \mathbf{v}_2 \; \ldots \; \mathbf{v}_N \,\right] \tag{B.33}$$

satisfying (B.8) and (B.9), i.e.,

$$\mathbf{v}_i^H\mathbf{v}_j = \begin{cases} 1 \; ; \; i = j = 1, \ldots, N \\ 0 \; ; \;\; i \; , \;\; j = 1, \ldots, N \;\; ; \; i \neq j \end{cases} \tag{B.34}$$

Properties

- An inverse matrix is defined only for a square matrix. Therefore, all unitary matrices are square matrices.
- Equation (B.32) is also the property of a normal matrix. Consequently, unitary matrices are a subset of normal matrices and thus unitarily diagonalizable. With a unitary transform matrix $\mathbf{V}$ the eigenvalue decomposition of (B.12) can be written as

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H \tag{B.35}$$

- All eigenvalues ϱ_i of a unitary matrix $\mathbf{V} \epsilon \, \mathbb{C}^{N\mathrm{x}N}$ have absolute values equal to 1

$$|\varrho_i| = 1 \; ; \;\; i = 1, \ldots, N \tag{B.36}$$

Proof: We calculate the scalar product of the two vectors and obtain with (B.32)

$$(\mathbf{V}\mathbf{v}_i)^H \, \mathbf{V}\mathbf{v}_i = \mathbf{v}_i^H\mathbf{V}^H\mathbf{V}\mathbf{v}_i = \|\mathbf{v}_i\|^2 \tag{B.37}$$

On the other hand the eigenvalue–eigenvector condition

$$\mathbf{V}\mathbf{v}_i = \varrho_i\mathbf{v}_i \tag{B.38}$$

holds. The left-hand side of (B.37) yields with (B.38)

$$(\mathbf{V}\mathbf{v}_i)^H \, \mathbf{V}\mathbf{v}_i = \varrho_i^* \varrho_i \mathbf{v}_i^H\mathbf{v}_i = |\varrho_i|^2 \, \|\mathbf{v}_i\|^2 \tag{B.39}$$

The left-hand sides of (B.37) and (B.39) are identical. Consequently, this must also hold for the right-hand sides

$$\|\mathbf{v}_i\|^2 = |\varrho_i|^2 \, \|\mathbf{v}_i\|^2 \tag{B.40}$$

from which the proposition (B.36) directly follows and the proof is finalized.

- The input signal $\mathbf{s}$ and output signal $\mathbf{y} = \mathbf{V}\mathbf{s}$ of a system described by a unitary matrix $\mathbf{V}$ have the same mean power

$$\mathbf{E}\left[\|\mathbf{y}\|^2\right] = \mathbf{E}\left[\|\mathbf{s}\|^2\right] \tag{B.41}$$

Proof:

$$\mathbf{E}\left[\|\mathbf{y}\|^2\right] = \mathbf{E}\left[(\mathbf{V}\mathbf{s})^H \mathbf{V}\mathbf{s}\right] = \mathbf{E}\left[\mathbf{s}^H \mathbf{V}^H \mathbf{V}\mathbf{s}\right] = \mathbf{E}\left[\mathbf{s}^H \mathbf{I}_N \mathbf{s}\right] = \mathbf{E}\left[\|\mathbf{s}\|^2\right]. \tag{B.42}$$

B.5 Norm of a Vector, Norm of a Matrix

The squared norm of a vector $\mathbf{v} = \begin{pmatrix} v_1 & v_2 & \cdots & v_N \end{pmatrix}^T$ with complex components is given by the sum of the squared absolute values of the components

$$\|\mathbf{v}\|^2 = \sum_{i=1}^{N} |v_i|^2 \tag{B.43}$$

The squared norm (Frobenius norm) of a matrix $\boldsymbol{A} \in \mathbb{C}^{M \times N}$ with complex entries a_{ij} is given by the sum of the squared absolute values of the entries

$$\|\mathbf{A}\|_{\mathrm{F}}^2 = \sum_{i=1}^{M} \sum_{j=1}^{N} \left|a_{ij}\right|^2 = \sum_{j=1}^{N} \left\|\mathbf{a}_j\right\|^2 \tag{B.44}$$

Alternatively, the Frobenius norm can be calculated as the sum of the squared norms of the column vectors $\mathbf{a}_j$ or row vectors of $\mathbf{A}$, respectively.

B.6 Singular Value Decomposition

The Procedure

The Singular Value Decomposition (SVD) of a matrix $\mathbf{H} \in \mathbb{C}^{N \times M}$ is given by

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H \tag{B.45}$$

with

$$\mathbf{D} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ & & & \ddots & & & & \\ 0 & 0 & 0 & \cdots & \sqrt{\lambda_P} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ & & \vdots & \cdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}_P^{\frac{1}{2}} & 0 \cdots 0 \\ 0 & 0 \cdots 0 \\ \vdots & \vdots \cdots \vdots \\ 0 & 0 \cdots 0 \end{pmatrix} \epsilon \mathbb{R}^{N \mathrm{x} M} \tag{B.46}$$

$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_P > 0$, and $\lambda_{P+1} = \lambda_{P+2} = \cdots = \lambda_N = 0$ are the N eigenvalues of the Hermiteian matrix

$$\mathbf{Q}_N = \mathbf{H}\mathbf{H}^H \epsilon \mathbb{C}^{N \mathrm{x} N} \tag{B.47}$$

and

$$P = \text{rank}\,(\mathbf{Q}_N) \tag{B.48}$$

is the rank of the matrix $\mathbf{Q}_N$. In general the rank of a matrix $\mathbf{H} \epsilon \mathbb{C}^{N \mathrm{x} M}$ is defined as the number of linearly independent rows or columns of the matrix, thus

$$\text{rank}\,(\mathbf{H}) \leq \min\{M, N\} \tag{B.49}$$

From this definition follows for (B.48)

$$P \leq N \tag{B.50}$$

$\sqrt{\lambda_i}$; $i = 1, \ldots, P$ are called the singular values of the matrix $\mathbf{H}$.
$\mathbf{U} \epsilon \mathbb{C}^{N \mathrm{x} N}$ and $\mathbf{V} \epsilon \mathbb{C}^{M \mathrm{x} M}$ are unitary matrices, thus

$$\mathbf{U}^{-1} = \mathbf{U}^H \; ; \; \mathbf{V}^{-1} = \mathbf{V}^H \tag{B.51}$$

hold. Furthermore, $\mathbf{U}$ is the matrix of the normalized eigenvectors with respect to the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_N$. Let

$$\mathbf{\Lambda}_N = \text{diag}\;(\lambda_1, \lambda_2, \ldots, \lambda_P, 0, \ldots, 0) \;\epsilon \mathbb{R}^{N \mathrm{x} N} \tag{B.52}$$

be a diagonal matrix composed of the eigenvalues of $\mathbf{Q}_N$. Then the eigenvalue decomposition of $\mathbf{Q}_N$ is

$$\mathbf{U}^H \mathbf{Q}_N \mathbf{U} = \mathbf{\Lambda}_N \tag{B.53}$$

One method to find the matrix $\mathbf{V}$ in (B.45) is the eigenvalue decomposition of the matrix

$$\mathbf{Q}_M = \mathbf{H}^H \mathbf{H} \epsilon \mathbb{C}^{M \mathrm{x} M} \tag{B.54}$$

which is

$$\mathbf{V}^H \mathbf{Q}_M \mathbf{V} = \mathbf{\Lambda}_M \tag{B.55}$$

with the diagonal matrix

$$\mathbf{\Lambda}_M = \text{diag } (\lambda_1, \lambda_2, \ldots, \lambda_P, 0, \ldots, 0) \ \epsilon \mathbb{R}^{M \text{x} M} \tag{B.56}$$

$\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{Q}_M$ with respect to the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_M$. Note that the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_P$ are the same as for the matrix $\mathbf{Q}_N$. Furthermore

$$\text{rank } (\mathbf{Q}_M) = \text{rank } (\mathbf{Q}_N) = P \tag{B.57}$$

holds and $\mathbf{\Lambda}_M$ as well as Λ_N contain the same diagonal matrix

$$\mathbf{\Lambda}_P = \text{diag } (\lambda_1, \lambda_2, \ldots, \lambda_P) \ \epsilon \mathbb{R}^{P \text{x} P} \tag{B.58}$$

of the P eigenvalues, which are unequal to zero. Note that in (B.46)

$$\mathbf{\Lambda}_P^{\frac{1}{2}} = \text{diag } \left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_P} \right) \tag{B.59}$$

holds.

Notes

- In contrast to the eigenvalue decomposition, which is only feasible for square matrices ($M = N$), the SVD in (B.45) can be done for any matrix $\mathbf{H} \epsilon \mathbb{C}^{N \text{x} M}$ with arbitrary M and N.
- Exercise: Consider the SVD of a square matrix $\mathbf{H} \epsilon \mathbb{C}^{M \text{x} M}$.
- The matrix $\mathbf{D}$ in (B.46) contains the square matrix $\mathbf{\Lambda}_P^{\frac{1}{2}} \epsilon \mathbb{R}^{P \text{x} P}$ defined in (B.59). As the remaining elements in $\mathbf{D}$ are zero, the SVD can also be formulated with $\mathbf{D} = \mathbf{\Lambda}_P^{\frac{1}{2}}$, a non-square matrices $\mathbf{U} \epsilon \mathbb{C}^{N \text{x} P}$, and $\mathbf{V} \epsilon \mathbb{C}^{M \text{x} P}$.

Proof of SVD Lemma

Proof of (B.45) and (B.46)

We prove that with the eigenvalue decomposition (B.53) of $\mathbf{Q}_N$ and with a unitary matrix $\mathbf{V} \epsilon \mathbb{C}^{M \text{x} M}$ the proposition (B.45) with (B.46) follows. First, we easily see that $\mathbf{Q}_N$ in (B.47) is a Hermiteian matrix, i.e., $\mathbf{Q}_N = \mathbf{Q}_N^H$ holds. We know from sections (B.2) and (B.3) that for any Hermiteian matrix an eigenvalue decomposition exists according to (B.53). By inserting (B.47) we obtain

$$\mathbf{U}^H \mathbf{H} \mathbf{H}^H \mathbf{U} = \Lambda_N \tag{B.60}$$

Now we introduce the following identity matrix $\mathbf{I} = \mathbf{V}\mathbf{V}^H$ making use of the prerequisite (B.51) that $\mathbf{V}$ is a unitary matrix, which yields

$$\mathbf{U}^H\mathbf{H}\mathbf{V}\mathbf{V}^H\mathbf{H}^H\mathbf{U} = \mathbf{U}^H\mathbf{H}\mathbf{V}\left(\mathbf{U}^H\mathbf{H}\mathbf{V}\right)^H = \mathbf{\Lambda}_N \tag{B.61}$$

Next we decompose the diagonal matrix on the right-hand side into the product of two matrices

$$\mathbf{\Lambda}_N = \mathbf{D}\mathbf{D}^H \tag{B.62}$$

Inserting (B.62) into (B.61) results in

$$\mathbf{U}^H\mathbf{H}\mathbf{V}\left(\mathbf{U}^H\mathbf{H}\mathbf{V}\right)^H = \mathbf{D}\mathbf{D}^H \tag{B.63}$$

and by comparison of the left and right-hand part we obtain

$$\mathbf{U}^H\mathbf{H}\mathbf{V} = \mathbf{D} \tag{B.64}$$

from which we conclude the proposition $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H$ and the proof ends. The only condition we have imposed so far on $\mathbf{V} \epsilon \mathbb{C}^{M\text{x}M}$ is the requirement that $\mathbf{V}$ is a unitary matrix. Moreover, we see that the derived SVD holds for arbitrary matrices $\mathbf{H} \epsilon \mathbb{C}^{N\text{x}M}$.

Proof of (B.55)

We now prove that $\mathbf{V} \epsilon \mathbb{C}^{M\text{x}M}$ can be obtained by the eigenvalue decomposition of $\mathbf{Q}_M = \mathbf{H}^H\mathbf{H}$. Assume that the singular value decomposition of $\mathbf{H}$ is given by (B.45). From this equation follows by applying the Hermiteian operation on both sides

$$\mathbf{H}^H = \mathbf{V}\mathbf{D}^H\mathbf{U}^H \tag{B.65}$$

Multiplying (B.65) from the right-hand side with (B.45) and knowing that $\mathbf{U}^H\mathbf{U} = \mathbf{I}_N$ results in

$$\mathbf{H}^H\mathbf{H} = \mathbf{V}\mathbf{D}^H\mathbf{U}^H\mathbf{U}\mathbf{D}\mathbf{V}^H = \mathbf{V}\mathbf{D}^H\mathbf{D}\mathbf{V}^H \tag{B.66}$$

From (B.46) follows

$$\mathbf{D}^H\mathbf{D} = \mathbf{\Lambda}_M \tag{B.67}$$

and we obtain from (B.66)

$$\mathbf{Q}_M = \mathbf{H}^H\mathbf{H} = \mathbf{V}\mathbf{\Lambda}_M\mathbf{V}^H \tag{B.68}$$

From (B.68) follows by multiplication with $\mathbf{V}^H$ and $\mathbf{V}$ directly the eigenvalue decomposition $\mathbf{V}^H\mathbf{Q}_M\mathbf{V} = \mathbf{\Lambda}_M$ of $\mathbf{Q}_M$ in (B.55). Consequently, $\mathbf{V}$ must be the matrix of eigenvectors associated to the eigenvalues given in $\mathbf{\Lambda}_M$. As $\mathbf{Q}_M$ is a Hermiteian matrix, we know from Sect. B.3 that $\mathbf{V}$ is unitary. Furthermore, we see from (B.68) and (B.53) together with (B.52) and (B.56) that $\mathbf{Q}_M$ and $\mathbf{Q}_N$ have the same positive eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_P$ and that their remaining eigenvalues are zero. This finalizes the proof.

B.7 Some Lemmas of Determinants

The proofs can be found in [4].

In the following, we assume "compatible" matrices **A**, **B**, **C**, and **I**, which means that their dimensions allow multiplication and addition.

- Determinant of the product of two matrices

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}) \tag{B.69}$$

- Determinant of the sum of matrices

$$\det(\mathbf{A} + \mathbf{BC}) = \det(\mathbf{A} + \mathbf{CB}) \;\; ; \;\; \text{if} \;\; \mathbf{AB} = \mathbf{BA} \tag{B.70}$$

- Determinant of the sum of matrices with cyclic permutation

$$\det(\mathbf{I} + \mathbf{ABC}) = \det(\mathbf{I} + \mathbf{BCA}) = \det(\mathbf{I} + \mathbf{CAB}) \tag{B.71}$$

- Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the eigenvalues of the matrix $\mathbf{A} \in \mathbb{C}^{N\text{x}N}$. Then

$$\det(\mathbf{A}) = \lambda_1 \lambda_2, \ldots, \lambda_N. \tag{B.72}$$

B.8 Trace of a Matrix

Definition of Trace

Given the square matrix

$$\mathbf{A} = (a_{ik}) \in \mathbb{C}^{N\text{x}N} \tag{B.73}$$

The trace of **A** is defined as

$$\text{tr}(\mathbf{A}) = a_{11} + a_{22} + \ldots + a_{NN} \tag{B.74}$$

With a scalar factor α follows

$$\text{tr}(\alpha\mathbf{A}) = \alpha \,\text{tr}(\mathbf{A}) \tag{B.75}$$

Note, for a non-square matrix the trace does not exist, because there is no main diagonal. The proof of the following Lemmas is straightforward.

Cyclic Permutation

Let $\mathbf{A} \in \mathbb{C}^{N\text{x}M}$; $\mathbf{B} \in \mathbb{C}^{M\text{x}N}$ and $\mathbf{C} \in \mathbb{C}^{N\text{x}N}$. Consequently, the product **ABC** is an NxN square matrix. Then

$$\mathrm{tr}\,(\mathbf{ABC}) = \mathrm{tr}\,(\mathbf{BCA}) = \mathrm{tr}\,(\mathbf{CAB}) \neq \mathrm{tr}\,(\mathbf{ACB}) \tag{B.76}$$

In particular

$$\mathrm{tr}\,(\mathbf{AB}) = \mathrm{tr}\,(\mathbf{BA}) \tag{B.77}$$

This also holds for $M = N$.

Trace of the Sum of Matrices

Let $\mathbf{A} \in \mathbb{C}^{N\mathrm{x}N}$ and $\mathbf{B} \in \mathbb{C}^{N\mathrm{x}N}$ be square matrices of the same dimension. Then

$$\mathrm{tr}\,(\mathbf{A} + \mathbf{B}) = \mathrm{tr}\,(\mathbf{A}) + \mathrm{tr}\,(\mathbf{B}) \tag{B.78}$$

With (B.77) follows

$$\mathrm{tr}\,(\mathbf{AB} - \mathbf{BA}) = 0 \tag{B.79}$$

Trace and Eigenvalues

Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the eigenvalues of the matrix $\mathbf{A} \in \mathbb{C}^{N\mathrm{x}N}$. Then

$$\mathrm{tr}\,(\mathbf{A}) = \sum_{i=1}^{N} \lambda_i \tag{B.80}$$

and

$$\mathrm{tr}\,(\mathbf{A}^{-1}) = \sum_{i=1}^{N} \lambda_i^{-1} \tag{B.81}$$

For the latter, the eigenvalues must be unequal to zero.

B.9 Differentiation of a Scalar Function $f(\mathbf{Z})$ with Respect to a Matrix Z

For the proof see [5].

Definition

Differentiation of a scalar function with respect to a matrix is a shorthand notation meaning that the scalar function is partially differentiated with respect to all matrix elements of $\mathbf{Z}$ and arranged in a matrix. Example:

$$\mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \end{pmatrix} \;;\; f(\mathbf{Z}) = f(z_{11}, z_{12}, z_{13}, z_{21}, z_{22}, z_{23}) \;;\; z_{ik} \in \mathbb{C} \tag{B.82}$$

Obviously, $f(\mathbf{Z})$ is a multivariate scalar function of $z_{11}, \ldots, z_{23}$. Then we define

$$\frac{\partial f}{\partial \mathbf{Z}^*} = \begin{pmatrix} \frac{\partial f}{\partial z_{11}^*} & \frac{\partial f}{\partial z_{12}^*} & \frac{\partial f}{\partial z_{13}^*} \\ \\ \frac{\partial f}{\partial z_{21}^*} & \frac{\partial f}{\partial z_{22}^*} & \frac{\partial f}{\partial z_{23}^*} \end{pmatrix} \tag{B.83}$$

For complex variables z_{ik} we define

$$\frac{\partial f}{\partial z_{ik}^*} = \frac{1}{2}\left(\frac{\partial f}{\partial \mathrm{Re}\,(z_{ik})} + j\frac{\partial f}{\partial \mathrm{Im}\,(z_{ik})}\right) \tag{B.84}$$

Differentiation of the Trace of a Matrix with Respect to a Matrix

We start with the differentiation of a constant α

$$\frac{\partial \alpha}{\partial \mathbf{Z}^*} = 0 \;;\; \mathbf{Z} \epsilon \mathbb{C}^{N\mathrm{x}M} \;;\; \alpha \epsilon \mathbb{C} \tag{B.85}$$

Please note in the following that the argument of tr (...) has to be a square matrix.

$$\frac{\partial \mathrm{tr}\,(\mathbf{Z}^H)}{\partial \mathbf{Z}^*} = \mathbf{I}_N \;;\; \mathbf{Z} \epsilon C^{N\mathrm{x}N} \tag{B.86}$$

$$\frac{\partial \mathrm{tr}\,(\mathbf{Z})}{\partial \mathbf{Z}^*} = 0 \;;\; \mathbf{Z} \epsilon \mathbb{C}^{N\mathrm{x}N} \tag{B.87}$$

$$\frac{\partial \mathrm{tr}\,(\mathbf{A}\mathbf{Z}^H)}{\partial \mathbf{Z}^*} = \mathbf{A} \;;\; \mathbf{A} \epsilon \mathbb{C}^{N\mathrm{x}M} \;;\; \mathbf{Z} \epsilon \mathbb{C}^{N\mathrm{x}M} \;;\; \mathbf{A}\mathbf{Z}^H \epsilon \mathbb{C}^{N\mathrm{x}N} \tag{B.88}$$

$$\frac{\partial \mathrm{tr}\,(\mathbf{A}\mathbf{Z})}{\partial \mathbf{Z}^*} = 0 \;;\; \mathbf{A} \epsilon \mathbb{C}^{M\mathrm{x}N} \;;\; \mathbf{Z} \epsilon \mathbb{C}^{N\mathrm{x}M} \;;\; \mathbf{A}\mathbf{Z} \epsilon \mathbb{C}^{M\mathrm{x}M} \tag{B.89}$$

$$\frac{\partial \mathrm{tr}\,(\mathbf{Z}\mathbf{Z}^H)}{\partial \mathbf{Z}^*} = \mathbf{Z} \;;\; \mathbf{Z} \epsilon \mathbb{C}^{N\mathrm{x}M} \;;\; \mathbf{Z}\mathbf{Z}^H \epsilon \mathbb{C}^{N\mathrm{x}N} \tag{B.90}$$

$$\frac{\partial \mathrm{tr}\,(\mathbf{A}\mathbf{Z}\mathbf{Z}^H)}{\partial \mathbf{Z}^*} = \mathbf{A}\mathbf{Z} \;;\; \mathbf{A} \epsilon \mathbb{C}^{N\mathrm{x}N} \;;\; \mathbf{Z} \epsilon \mathbb{C}^{N\mathrm{x}M} \;;\; \mathbf{A}\mathbf{Z}\mathbf{Z}^H \epsilon \mathbb{C}^{N\mathrm{x}N} \tag{B.91}$$

With cyclic permutation Lemma (B.76) we obtain from (B.91)

$$\frac{\partial \mathrm{tr}\,(\mathbf{Z}\mathbf{Z}^H\mathbf{A})}{\partial \mathbf{Z}^*} = \frac{\partial \mathrm{tr}\,(\mathbf{A}\mathbf{Z}\mathbf{Z}^H)}{\partial \mathbf{Z}^*} = \mathbf{A}\mathbf{Z} \;;\; \mathbf{A} \epsilon \mathbb{C}^{N\mathrm{x}N} \;;\; \mathbf{Z} \epsilon \mathbb{C}^{N\mathrm{x}M} \;;\; \mathbf{Z}\mathbf{Z}^H\mathbf{A} \epsilon \mathbb{C}^{N\mathrm{x}N} \tag{B.92}$$

$$\frac{\partial \text{tr}\,(\mathbf{ZAZ}^H\mathbf{B})}{\partial \mathbf{Z}^*} = \mathbf{BZA}\ ;\ \mathbf{A} \epsilon\, \mathbb{C}^{M\text{x}M}\ ;\ \mathbf{Z} \epsilon\, \mathbb{C}^{N\text{x}M}\ ;\ \mathbf{B} \epsilon\, \mathbb{C}^{N\text{x}N}\ ;\ \mathbf{ZAZ}^H\mathbf{B} \epsilon\, \mathbb{C}^{N\text{x}N} \tag{B.93}$$

References

1. A. Papoulis, S.U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th edn. (McGraw-Hill, Boston, 2002)
2. A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, Boston, 1965)
3. J.G. Proakis, M. Salehi, *Digital Communications* (McGraw-Hill, Boston, 2007)
4. R.A. Horn, C.R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 2013)
5. A. Hjorungnes, *Complex-Valued Matrix Derivatives with Applications in Signal Processing and Communications* (Cambridge University Press, Cambridge, 2011)

Index

J. Speidel, *Introduction to Digital Communications*, Signals and Communication Technology, https://doi.org/10.1007/978-3-030-00548-1