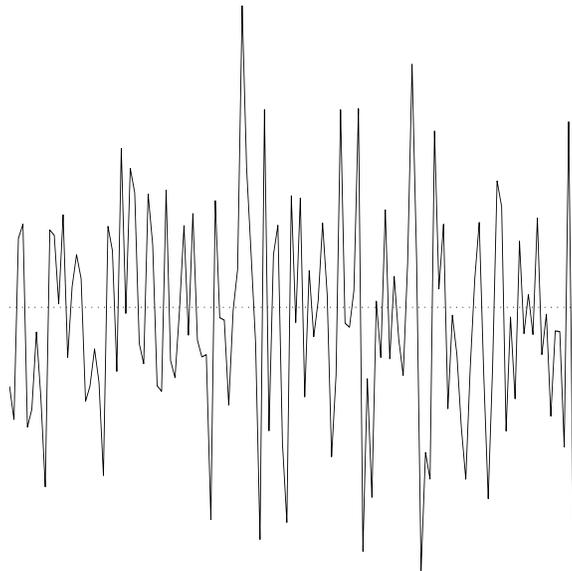

An Introduction to Stochastic Processes



Andreas Jakobsson
Lund University

Version 080515

An Introduction to Stochastic Processes

A first version of these notes were written as a part of a graduate level course on adaptive signal processing at Karlstad University during 2004. The material is aimed to be an introduction to stochastic processes, but also contains some brief notes on optimal and constrained filtering. For a more complete description of the material, the reader is referred to [1–5].

Andreas Jakobsson
andreas.jakobsson@ieee.org

CONTENTS

1	STOCHASTIC PROCESSES	1
1.1	What is a stochastic process?	1
1.2	Properties and peculiarities	2
1.3	Power spectral density	6
1.4	Linear filtering of a WSS process	9
1.5	Auto-regressive processes	10
2	THE WIENER FILTER	12
2.1	Optimal filtering	12
2.2	Constrained filtering	16
	BIBLIOGRAPHY	20
	SUBJECT INDEX	22

STOCHASTIC PROCESSES

*One must learn by doing the thing;
for though you think you know it
You have no certainty, until you try.
Sophocles*

1.1 What is a stochastic process?

When working with a measured data set, we often view it as a *realization* of a *stochastic process*; for this reason, it is important to spend a moment dwelling on the notion of such a creature, making the reader familiar with its properties and peculiarities. The material presented here is meant to be a short primer on the topic; it is in no way the complete story and these notes should mainly be seen as a brief summary. For a more complete treatment of the topic, the reader is referred to [1–3].

We begin by asking ourself why there is a need to introduce a statistical model for the measured data set; why not just work on the data set itself? Why make things more complicated than they are? This is a good question, and it deserves an answer: here, we will present two attempts in justifying it. First, the data sets we are interested in are mainly the result of some form of measurement procedure, and as such inherently sensitive to disturbances due to the measuring device as well as its surroundings. Unfortunately, it is very hard to produce an accurate description of such disturbances, typically being the result of a vast number of different sources; however, to develop different algorithms, we do need to make some assumptions on the properties of the data. In general, the more knowledge we have on the kind of data we are observing, the better we can exploit the particular properties of it. As a way to approach this problem, it is often very convenient to treat the measured signal as a particular *realization* of a stochastic process (if we have two data sets, we simply treat these as two different realizations of the *same* process), i.e., we assume that the stochastic process has some particular properties, and that the measured signal is nothing but an example (realization) of the possible outcomes. Imagine a (rather large) bucket containing (infinitely many) data vectors, all sharing

some certain properties. We then view the observed data set as a randomly chosen vector from the bucket, naming the bucket a *stochastic process*, and the selected data vector a *realization* of the process. This implies that if we were to perform the measurements again, we would simply select a random data vector from the bucket again; this vector would (very likely) be different from the first vector, but would share the properties that all the vectors in the bucket have. By working on these properties, not the actual data vector itself, we can allow for measurements that are made in presence of disturbances, as long as we can say something of the characteristics of them. Secondly, even in cases when we have a fairly good feeling for the signal, it might be preferable to treat it as a realization of a process well described by a few characteristics; then, we end up having to estimate a small set of parameters, well describing the data (if viewed as a realization of a stochastic process), instead of a large number of deterministic unknowns (if viewed as a sequence of samples). Typically, such an approximation will yield a non-optimal result, but might well be better than attempting to estimate all the unknowns from a limited data set (which often is infeasible).

1.2 Properties and peculiarities

Herein, we are primarily interested in discrete-time signals, and will thus limit our attention to discrete-time stochastic processes. Such a process can be viewed as a sequence of stochastic variables, $x(n)$, where each variable has an underlying *probability distribution function*

$$F_{x(n)}(\alpha) = \Pr \{x(n) \leq \alpha\} \quad (1.2.1)$$

and a *probability density function* (pdf)

$$f_{x(n)}(\alpha) = \frac{d}{d\alpha} F_{x(n)}(\alpha) \quad (1.2.2)$$

However, to fully characterize the process we must also specify how the different stochastic variables depends on each other; Kolmogorov's theorem tells us that:

The joint probability distribution function

$$F_{x(n_1), \dots, x(n_k)}(\alpha_1, \dots, \alpha_k) = \Pr \{x(n_1) \leq \alpha_1, \dots, x(n_k) \leq \alpha_k\}$$

forms a complete statistical characterization of a stochastic process, for any collection of random variables $x(n_i)$.

For example, the process might be formed by a sequence of uncorrelated¹ Gaussian

¹Recall that two stochastic variables are statistically *independent* if $f_{x,y}(\alpha, \beta) = f_x(\alpha)f_y(\beta)$. If $E\{xy\} = E\{x\}E\{y\}$, they are said to be *uncorrelated*.

variables. This particular process is called *Gaussian² white noise*, and is a very common model for describing a measurement disturbance such as thermal noise. The *mean* of a process is defined as

$$m_x(n) = E\{x(n)\}, \quad (1.2.3)$$

where $E\{\cdot\}$ denotes the statistical expectation³. Further, the *autocovariance* of the process is given by

$$c_x(k, l) = E\{[x(k) - m_x(k)][x(l) - m_x(l)]^*\}, \quad (1.2.4)$$

where $(\cdot)^*$ denotes the conjugate⁴. Defining the *autocorrelation function* (acf),

$$r_x(k, l) = E\{x(k)x^*(l)\}, \quad (1.2.5)$$

allows us to write

$$c_x(k, l) = r_x(k, l) - m_x(k)m_x^*(l). \quad (1.2.6)$$

Example 1.1: Let $x(n) = Ae^{in\omega_0+i\phi}$, where ϕ is a uniformly distributed random variable between $[-\pi, \pi]$. Then,

$$m_x(n) = E\{Ae^{in\omega_0+i\phi}\} = \int_{-\pi}^{\pi} Ae^{in\omega_0+i\phi} \frac{1}{2\pi} d\phi = 0 \quad (1.2.7)$$

and

$$r_x(k, l) = E\{|A|^2 e^{ik\omega_0+i\phi} e^{-il\omega_0-i\phi}\} = |A|^2 e^{i\omega_0(k-l)} \quad (1.2.8)$$

Note that the acf of a sinusoid is itself a sinusoid; furthermore, it only depends on the difference $(k - l)$, not the actual values of k and l themselves.

The process in Example 1.1 is a typical example of the kind of processes we are interested in, namely processes that are wide-sense stationary. The definition is:

A stochastic process is *wide-sense stationary* (WSS) if

- (i) The mean is constant, $m_x(k) = m_x$.
- (ii) The auto-correlation, $r_x(k, l)$, depends only on the difference $(k - l)$; for this reason we normally just write $r_x(k - l)$.
- (iii) The *variance* is finite, $\sigma_x^2 = c_x(0) < \infty$.

²Johann Carl Friedrich Gauss (1777-1855) was born in Brunswick, the Duchy of Brunswick (now Germany). In 1801, he received his doctorate in mathematics, proving the fundamental theorem of algebra [6]. He made remarkable contributions to all fields of mathematics, and is generally considered to be the greatest mathematician of all times.

³Recall that $E\{g(x)\} = \int_{-\infty}^{\infty} g(x)f(x) dx$, where $f(x)$ is the pdf of x .

⁴Note that we will here also allow for processes that take on complex values.

Note that the process in Example 1.1 satisfy all three conditions. It might seem to be a rather strong restriction to assume that a process is WSS; this is indeed the case⁵. However, a surprising number of real-life data sets can be very well modelled as being WSS, and here we will limit our attention to processes having this property. Furthermore, without loss of generality, we will normally assume that the process is *zero-mean*, i.e., $m_x = 0$. Note that this implies that $c_x(k) = r_x(k)$. The definition of WSS implies some very useful properties of the acf:

- (i) $r_x(k) = r_x^*(-k)$.
- (ii) $r_x(0) = E\{|x(n)|^2\} \geq 0$.
- (iii) $r_x(0) \geq |r_x(k)|, \forall k$.

Most data sets contains a finite number of data samples, say N . A convenient way to represent the samples is in vector form,

$$\mathbf{x} = [x(0) \quad \cdots \quad x(N-1)]^T, \quad (1.2.9)$$

where $(\cdot)^T$ denotes the transpose. Assuming that $x(n)$ is WSS, we form the $L \times L$ *autocorrelation* matrix as (with $L \leq N$)

$$\mathbf{R}_x = \begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(L-1) \\ r_x^*(1) & r_x(0) & \cdots & r_x(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x^*(L-1) & r_x^*(L-2) & \cdots & r_x(0) \end{bmatrix} \quad (1.2.10)$$

which has the following properties:

The autocorrelation matrix satisfy

- (i) \mathbf{R}_x is a *Hermitian Toeplitz* matrix.
- (ii) \mathbf{R}_x is non-negative definite, $\mathbf{R}_x \geq 0$.

Recall that a Hermitian⁶ matrix satisfies

$$\mathbf{A} = \mathbf{A}^H, \quad (1.2.11)$$

⁵We note that wide-sense stationarity is in itself a weak form of stationarity. A stronger form of stationarity is *strict-sense stationarity* (SSS); a process is said to be SSS if its statistical properties are invariant to a shift of the origin, i.e., $f_{x(n)}(\alpha) = f_{x(n)}(\alpha+T)$, for any T , for a first-order density. This is a very strong assumption that rarely can be imposed.

⁶Charles Hermite (1822-1901) was born in Dieuze, France, with a defect in his right foot, making it hard for him to move around. After studying one year at the Ecole Polytechnique, Hermite was refused the right to continue his studies because of this disability. However, this did not prevent his work, and in 1856 he was elected to the Académie des Sciences. After making important contributions to number theory and algebra, orthogonal polynomials, and elliptic functions, he was in 1869 appointed professor of analysis both at the Ecole Polytechnique and at the Sorbonne.

where $(\cdot)^H$ denotes the Hermitian, or conjugate transpose, operator. Further, if a matrix \mathbf{A} is non-negative definite, it implies that the eigenvalues of \mathbf{A} are real-valued and non-negative. A matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ is said to be Toeplitz⁷ if it has the same element along each diagonal⁸. These properties imply a very strong structure, and enables the development of very efficient algorithms.

Example 1.2: Consider the process in Example 1.1. The $L \times L$ autocorrelation matrix is then given by

$$\mathbf{R}_x = |A|^2 \mathbf{a}_L(\omega_0) \mathbf{a}_L^H(\omega_0) \quad (1.2.12)$$

where $\mathbf{a}_L(\omega)$ is a so-called Fourier⁹ vector,

$$\mathbf{a}_L(\omega) = [1 \quad e^{i\omega} \quad \dots \quad e^{i\omega(L-1)}]^T \quad (1.2.13)$$

Thus, \mathbf{R}_x is a rank-one non-negative definite Toeplitz matrix.

Here, we will further make the assumption that all stochastic processes are *ergodic*; this means that we can use the different samples in the realization to estimate the properties of the process. For example, if we wish to estimate the mean of a process at a given time, say $m_x(p)$, we should average over all realizations, i.e.,

$$m_x(p) = \lim_{M \rightarrow \infty} \left\{ \frac{1}{M} \sum_{l=1}^M x_l(p) \right\} \quad (1.2.14)$$

where $x_l(p)$ denotes the p th sample of the l th realization. Obviously, this is problematic if we only have a limited number of realizations. However, if the process is *ergodic in the mean*, we can estimate the mean from a single realization; *asymptotically*, the average of any realization will approach the mean of the process (which then must be the same for all time, $m_x(p) = m_x$), i.e.,

$$m_x = \lim_{N \rightarrow \infty} \left\{ \frac{1}{N} \sum_{l=1}^N x(l) \right\} \quad (1.2.15)$$

Similarly, we assume that the processes are *ergodic in the autocorrelation*, so that we may estimate the acf from a single realization. Hereafter, we will for a finite

⁷Otto Toeplitz (1881-1940) was born in Breslau, Germany (now Wroclaw, Poland). He was awarded his doctorate in 1905 from the University of Breslau. He worked on infinite linear and quadratic forms, and developed a general theory of infinite dimensional spaces, criticizing Banach's work as being too abstract. Being Jewish, he was forced to retire from his chair at the University of Bonn in 1935, and emigrated to Palestine in 1939 to help building up Jerusalem University.

⁸Another common matrix structure is a *Hankel* matrix. Such a matrix has identical elements on each of the anti-diagonals.

⁹Jean Baptiste Joseph Fourier (1768-1830) was born in Auxerre, France. He was closely involved in the revolution, for which he was twice imprisoned. In 1798, Fourier joined Napoleon's army in its invasion of Egypt as scientific adviser, and later held various administrative positions under Napoleon. His work on the propagation of heat in solid bodies caused controversy due to Fourier's expansions of functions as trigonometrical series.

length data sequence generally estimate the mean and the acf as

$$\hat{m}_x = \frac{1}{N} \sum_{l=0}^{N-1} x(l) \quad (1.2.16)$$

$$\hat{r}_x(k) = \frac{1}{N} \sum_{l=0}^{N-k-1} x(l)x^*(l+k), \quad k = 0, \dots, N-1 \quad (1.2.17)$$

Note that the acf estimate in (1.2.17) is *biased* as

$$E\{\hat{r}_x(k)\} = \frac{N-|k|}{N} r_x(k) \neq r_x(k) \quad (1.2.18)$$

Note that it is generally undesirable with biased estimators¹⁰. By simply replacing the factor N in (1.2.17) with $N-k$, we would obtain an *unbiased* estimate of the acf; however, this estimator is not often used! The reasons for this are:

- (i) For most stationary signals, the acf decays rather rapidly, so that $r_x(k)$ is quite small for large lags k . This will be true for the biased estimator in (1.2.17), but not necessarily for the unbiased estimator which may take large and erratic values for large k (as it is obtained by only a few products in such a case).
- (ii) The sequence $\{\hat{r}_x(k)\}$ obtained with (1.2.17) is positive semi-definite (as it should be), while this is not the case for the unbiased estimate. This is especially important for spectral estimation (see below).

1.3 Power spectral density

An often convenient way to characterize a stochastic process is via its *power spectral density* (psd), defined as the (discrete-time) Fourier transform of the acf, i.e.,

$$\phi_x(\omega) = \sum_{k=-\infty}^{\infty} r_x(k) e^{-i\omega k} \quad (1.3.1)$$

The inverse transform recovers $r_x(k)$,

$$r_x(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_x(\omega) e^{i\omega k} d\omega \quad (1.3.2)$$

from which we note that

$$r_x(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_x(\omega) d\omega \quad (1.3.3)$$

¹⁰An estimator is biased if the mean of the estimator is different from the true mean, i.e., if $E\{\hat{q}\} \neq q$, where \hat{q} is an estimate of q . Typically, it is desirable that an estimator is *consistent*, meaning that the estimation error, $\hat{q} - q$, is small for large data sets. If the estimator is biased, it is not consistent.

Since $r_x(0) = E\{|x(n)|^2\}$ measures the power of $x(n)$, the equality in (1.3.3) shows that $\phi_x(\omega)$ is indeed correctly named a power spectral density as it is representing the distribution of the signal power over frequencies. Under weak assumptions, it can be shown that (1.3.1) is equivalent to

$$\phi_x(\omega) = \lim_{N \rightarrow \infty} \left\{ \left| \frac{1}{N} \sum_{k=0}^{N-1} x(k) e^{-i\omega k} \right|^2 \right\} \quad (1.3.4)$$

Using the discrete-time Fourier transform (DFT),

$$X(\omega) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) e^{-i\omega k}, \quad (1.3.5)$$

the psd in (1.3.4) can be expressed as

$$\phi_x(\omega) = \lim_{N \rightarrow \infty} \left\{ |X(\omega)|^2 \right\} \quad (1.3.6)$$

which also suggests the most natural way to estimate the psd, i.e., as the magnitude square of the DFT of the data vector. This estimator, termed the *Periodogram*, was introduced in 1898 by Sir Arthur Schuster¹¹, who derived it to determine “hidden periodicities” (non-obvious periodic signals) in time series [8].

Example 1.3: Let $e(n)$ be a zero-mean white Gaussian process with variance σ_e^2 . Then,

$$r_e(k) = \sigma_e^2 \delta_K(k), \quad (1.3.7)$$

where $\delta_K(k)$ is the Kronecker¹² delta function,

$$\delta_K(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \quad (1.3.8)$$

and the $L \times L$ autocorrelation matrix is given by

$$\mathbf{R}_e = \sigma_e^2 \mathbf{I}_L \quad (1.3.9)$$

where \mathbf{I}_L is the $L \times L$ identity matrix. Further,

$$\phi_e(\omega) = \sigma_e^2. \quad (1.3.10)$$

¹¹Schuster applied the Periodogram to find hidden periodicities in the monthly sunspot numbers for the years 1749 to 1894, yielding the classical estimate of 11.125 years for the sunspot cycle [7].

¹²Leopold Kronecker (1823-1891) was born in Liegnitz, Prussia (now Poland). He received his doctorate in 1845 on his work on algebraic number theory. Kronecker was a wealthy man, and pursued mathematics primarily as a hobby. In 1883, he was appointed full professor at Berlin University, and in 1884 he was elected a foreign member of the Royal Society of London.

Since $\phi_x(\omega)$ is a power density, it should be *real-valued* and *non-negative*. That is indeed the case which can readily be seen from (1.3.4). Hence,

$$\phi_x(\omega) \geq 0, \quad \forall \omega. \quad (1.3.11)$$

Further, the power spectral density is periodic, such that

$$\phi(\omega) = \phi(\omega + 2\pi k), \quad (1.3.12)$$

for any integer k . Also; the psd of a real-valued process is *symmetric*; if the process is complex-valued, this is not necessarily the case.

Example 1.4: Let $y(t) = x(t) + e(t)$, $t = 1, \dots, N$, where $e(t)$ is assumed to be a zero-mean stationary white signal, with variance σ_e^2 , independent of $x(t)$, and

$$x(t) = \sum_{l=1}^n \alpha_l e^{i\omega_l t + i\varphi_l} \quad (1.3.13)$$

Then,

$$\mathbf{R}_y = \sum_{l=1}^n \alpha_l^2 \mathbf{a}_L(\omega_l) \mathbf{a}_L^H(\omega_l) + \sigma_e^2 \mathbf{I} \quad (1.3.14)$$

and

$$\phi_y(\omega) = \sum_{l=1}^n \alpha_l \delta_D(\omega - \omega_l) + \sigma_e^2 \quad (1.3.15)$$

where $\delta_D(\omega)$ is the Dirac delta function¹³, satisfying

$$f(a) = \int_{-\infty}^{\infty} f(x) \delta_D(x - a) dx \quad (1.3.16)$$

Another useful property of the psd is the so-called *eigenvalue extremal property*, stating that the eigenvalues of the $L \times L$ autocorrelation matrix of zero-mean WSS process are upper and lower bounded by the maximum and minimum value, respectively, of the power spectrum, i.e.,

$$\min_{\omega} \phi_x(\omega) \leq \lambda_k \leq \max_{\omega} \phi_x(\omega), \quad \text{for } k = 1, \dots, L \quad (1.3.17)$$

where λ_k denotes the k th eigenvalue¹⁴ of \mathbf{R}_x .

¹³Strictly speaking, $\delta_D(\omega)$ is not a function, it is a distribution or “generalized function”.

¹⁴Consider a square matrix \mathbf{A} . By definition, \mathbf{x} is an *eigenvector* of \mathbf{A} if it satisfies $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, which implies that $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0$, requiring \mathbf{x} to be in $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I})$, i.e., the *nullspace* of $\mathbf{A} - \lambda\mathbf{I}$. The *eigenvector* \mathbf{x} must (by definition) be non-zero, and as a result, $\mathbf{A} - \lambda\mathbf{I}$ must be singular.

1.4 Linear filtering of a WSS process

It is very common to apply a linear filter to a measured data set, for instance to remove the influence of high-frequency noise. This naturally raises the question of how the properties of a stochastic process is affected by filtering. We will in the following consider $h(n)$ to be a *linear, time-invariant* and *stable* filter, and form the convolution sum

$$y(n) = \sum_{l=-\infty}^{\infty} h(n)x(l-n) \quad (1.4.1)$$

Recalling

$$r_x(k) = E\{x(n)x^*(n-k)\} \quad (1.4.2)$$

we find that

$$\begin{aligned} r_y(k) &= E\{y(n)y^*(n-k)\} \\ &= \sum_{l=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} h(l)h^*(p)r_x(k-l+p) \end{aligned} \quad (1.4.3)$$

or, if viewed in the frequency domain,

$$\phi_y(\omega) = |H(\omega)|^2 \phi_x(\omega), \quad (1.4.4)$$

where $H(\omega)$ is the DFT of $h(n)$. This is a very powerful result, and the reader is advised to learn it by heart. Further, we note that the variance of the filter output can be written as

$$\sigma_y^2 \triangleq r_y(0) = \sum_{l=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} h(l)h^*(p)r_x(p-l) \quad (1.4.5)$$

Typically, we are interested in finite length impulse responses. Hereafter, we assume that $h(n)$ is finite in length and zero outside the interval $[0, L-1]$. Then, (1.4.5) can be written as

$$\sigma_y^2 = \mathbf{h}^H \mathbf{R}_x \mathbf{h} \quad (1.4.6)$$

where

$$\mathbf{h} = [h(0) \quad \cdots \quad h(L-1)]^T \quad (1.4.7)$$

Example 1.5: If $x(n)$ is a zero-mean WSS process, and $y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k)$, is it true that

$$\sigma_y^2 \stackrel{?}{=} \sigma_x^2 \sum_{n=-\infty}^{\infty} |h(n)|^2 \quad (1.4.8)$$

Answer: As $\sigma_x^2 = r_x(0)$, the result in (1.4.8) only holds if $x(n)$ is a white noise sequence. Thus, (1.4.8) is not true in general.

Note that if we assume that $x(n)$ was a white sequence, such that $r_x(k) = \delta_K(k)$, then (1.4.4) yields the *spectral factorization theorem*.

1.5 Auto-regressive processes

A wide range of stochastic processes may be modeled as being auto-regressive moving-average processes (ARMA), having a rational power spectral density (psd)

$$\phi_x(\omega) = \left| \frac{A(\omega)}{B(\omega)} \right|^2 \sigma_w^2 \quad (1.5.1)$$

where σ_w^2 is a positive scalar, and $A(\omega)$ and $B(\omega)$ are polynomials,

$$A(\omega) = 1 + a_1 e^{-i\omega} + \dots + a_p e^{-ip\omega} \quad (1.5.2)$$

$$B(\omega) = 1 + b_1 e^{-i\omega} + \dots + b_q e^{-iq\omega} \quad (1.5.3)$$

We note that according to Weierstrass theorem, any continuous psd can be approximated arbitrarily closely by a rational psd of the form (1.5.1), provided that the degrees p and q are sufficiently large. This fact makes the ARMA processes particularly interesting; herein, we will however mainly be concerned with the *auto-regressive* (AR) process, being a special case of (1.5.1) having $q = 0$. The AR process is very commonly used in applications, and as such deserves an additional comment. We say that the time-series $x(n), x(n-1), \dots, x(n-M)$ represents the realization of an AR process of order M if it satisfy

$$x(n) + a_1^* x(n-1) + \dots + a_M^* x(n-M) = w(n), \quad (1.5.4)$$

where a_1, a_2, \dots, a_M are constants called the AR parameters and $w(n)$ is an additive white noise. Multiplying (1.5.4) with $x^*(n-k)$ and taking expectation yields

$$E \left\{ x^*(n-k) \sum_{l=0}^M a_l^* x(n-l) \right\} = \sum_{l=0}^M a_l^* r_x(k-l) = \sigma_w^2 \delta_K(k), \quad (1.5.5)$$

with $a_0 = 1$, which recalling that $r_x(k) = r_x^*(-k)$, can be written in matrix form as

$$\begin{bmatrix} r_x(0) & r_x(1) & \dots & r_x(M) \\ r_x^*(1) & r_x(0) & \dots & r_x(M-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_x^*(M) & r_x^*(M-1) & \dots & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1^* \\ \vdots \\ a_M^* \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (1.5.6)$$

or, alternatively,

$$\begin{bmatrix} r_x(0) & r_x(1) & \dots & r_x(M-1) \\ r_x^*(1) & r_x(0) & \dots & r_x(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x^*(M-1) & r_x^*(M-2) & \dots & r_x(0) \end{bmatrix} \begin{bmatrix} a_1^* \\ \vdots \\ a_M^* \end{bmatrix} = \begin{bmatrix} r_x^*(1) \\ r_x^*(2) \\ \vdots \\ r_x^*(M) \end{bmatrix} \quad (1.5.7)$$

The above equations are called the *Yule-Walker equations* or *normal equations*, and form the basis for many AR estimation techniques.

Example 1.6: Let $x(n)$ be a first-order AR-process, with $\sigma_w^2 = 1$. Then, the Yule-Walker equations can be written as

$$\begin{cases} r_x(0) + a_1^* r_x(1) = 1 \\ r_x(1) + a_1^* r_x(0) = 0 \end{cases} \quad (1.5.8)$$

and, thus,

$$a_1 = -\frac{r_x^*(1)}{r_x(0)} \quad (1.5.9)$$

An interesting consequence of the Yule-Walker equations is the so-called *Gohberg-Semencul formula* which states that the inverse of a Hermitian Toeplitz matrix, \mathbf{R}_x formed as in (1.5.6), may be found in *closed form* as [5, 9]

$$\sigma_w^2 \mathbf{R}_x^{-1} = \begin{bmatrix} 1 & & & \mathbf{0} \\ a_1^* & \ddots & & \\ \vdots & \ddots & \ddots & \\ a_M^* & \dots & a_1^* & 1 \end{bmatrix} \begin{bmatrix} 1 & a_1 & \dots & a_M \\ & \ddots & \ddots & \vdots \\ & & \ddots & a_1 \\ \mathbf{0} & & & 1 \end{bmatrix} - \begin{bmatrix} 0 & & & \mathbf{0} \\ a_M & \ddots & & \\ \vdots & \ddots & \ddots & \\ a_1 & \dots & a_M & 0 \end{bmatrix} \begin{bmatrix} 0 & a_M^* & \dots & a_1^* \\ & \ddots & \ddots & \vdots \\ & & \ddots & a_M^* \\ \mathbf{0} & & & 0 \end{bmatrix} \quad (1.5.10)$$

We will make nice use of this formula in the next chapter; here, we merely conclude this chapter by commenting that the above formula is closely related to the fascinating topic of displacement theory [10].

THE WIENER FILTER

*A thing of beauty is a joy for ever:
Its loveliness increases; it will never
Pass into nothingness; but still will keep
A bower quiet for us, and a sleep
Full of sweet dreams, and health, and
quiet breathing.*

John Keats

2.1 Optimal filtering

We will now briefly discuss optimal (in a least squares sense) filtering. The problem is as follows; consider a signal of interest, say $d(n)$, that for some reason is corrupted by some form of additive noise, here denoted $v(n)$. Thus, the measured signal, $x(n)$, can be written as

$$x(n) = d(n) + v(n). \quad (2.1.1)$$

This is depicted in Figure 2.1 below. We now ask ourselves how to design a filter, $W(z)$, such that the filter removes as much of the influence of the noise process as is possible, i.e., we wish to design a filter such that the filter output, $\hat{d}(n)$, is as close as possible to $d(n)$. Forming the estimation error,

$$e(n) = d(n) - \hat{d}(n), \quad (2.1.2)$$

the problem is that of minimizing the power of the error signal, i.e., we wish to find

$$W(z) = \arg \min_{W(z)} E \{|e(n)|^2\} \triangleq \arg \min_{W(z)} J_W, \quad (2.1.3)$$

where J_W is the cost-function to be minimized. Let

$$\hat{d}(n) = \sum_{k=0}^{\infty} w^*(k)x(n-k) \quad (2.1.4)$$

where

$$w(k) = a(k) + ib(k), \quad (2.1.5)$$

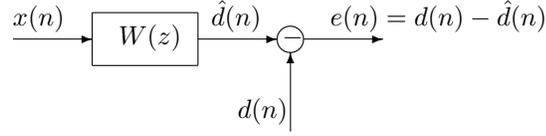


Figure 2.1. Optimal filter structure

with $a(k), b(k) \in \mathbb{R}$. We proceed to find the optimal filter by minimizing the cost-function in (2.1.3) with respect to $W(z)$, i.e., we set

$$\nabla_k J_W = 0, \quad k = 0, 1, \dots \quad (2.1.6)$$

where ∇_k is the complex gradient operator,

$$\nabla_k = \frac{\partial}{\partial a(k)} + i \frac{\partial}{\partial b(k)} \quad (2.1.7)$$

Thus,

$$\nabla_k J_W = E \left\{ \frac{\partial e(n)}{\partial a(k)} e^*(n) + i \frac{\partial e(n)}{\partial b(k)} e^*(n) + e(n) \frac{\partial e^*(n)}{\partial a(k)} + i e(n) \frac{\partial e^*(n)}{\partial b(k)} \right\} \quad (2.1.8)$$

where

$$\begin{aligned} \frac{\partial e(n)}{\partial a(k)} &= \frac{\partial}{\partial a(k)} [d(n) - \hat{d}(n)] \\ &= -\frac{\partial}{\partial a(k)} \sum_{k=0}^{\infty} w^*(k) x(n-k) \\ &= -x(n-k) \end{aligned} \quad (2.1.9)$$

and, similarly,

$$\frac{\partial e^*(n)}{\partial a(k)} = -x^*(n-k) \quad (2.1.10)$$

$$\frac{\partial e(n)}{\partial b(k)} = ix(n-k) \quad (2.1.11)$$

$$\frac{\partial e^*(n)}{\partial b(k)} = -ix^*(n-k) \quad (2.1.12)$$

yielding

$$\nabla_k J_W = -2E \{x(n-k)e^*(n)\} = 0 \quad k = 0, 1, \dots \quad (2.1.13)$$

Note that the measured signal is orthogonal to the estimation error; using (2.1.13), it is easy to show that the estimated desired signal is orthogonal to the estimation error, i.e.,

$$E \{ \hat{d}(n) e^*(n) \} = 0 \quad (2.1.14)$$

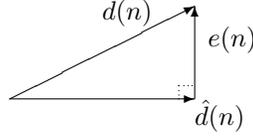


Figure 2.2. The principle of orthogonality

These equations form the so-called *orthogonality principle*, which is a central part of linear prediction, geometrically depicted in Figure 2.2. From (2.1.4) and (2.1.13), we conclude that

$$\underbrace{E\{x(n-k)d^*(n)\}}_{r_{xd}(-k)} = E\{x(n-k)\hat{d}^*(n)\} = \sum_{l=0}^{\infty} w(l) \underbrace{E\{x(n-k)x^*(n-l)\}}_{r_x(l-k)}$$

or, in short,

$$r_{xd}(-k) = \sum_{l=0}^{\infty} w(l)r_x(l-k) \quad k = 0, 1, \dots \quad (2.1.15)$$

which are the so-called *Wiener-Hopf* equations. Often, we are interested in assuming $W(z)$ to be an M -tap finite impulse response (FIR) filter, yielding

$$r_{xd}(-k) = \sum_{l=0}^{M-1} w(l)r_x(l-k) \quad (2.1.16)$$

or, in matrix form,

$$\mathbf{R}_x \mathbf{w}_o = \mathbf{r}_{xd} \quad (2.1.17)$$

where

$$\mathbf{R}_x = E\{\mathbf{x}_n \mathbf{x}_n^H\} \quad (2.1.18)$$

$$\mathbf{r}_{xd} = E\{\mathbf{x}_n d^*(n)\} \quad (2.1.19)$$

$$\mathbf{x}_n = [x(n) \quad x(n-1) \quad \dots \quad x(n-M+1)]^T \quad (2.1.20)$$

$$\mathbf{w}_o = [w(0) \quad \dots \quad w(M-1)]^T \quad (2.1.21)$$

From (2.1.17), we find the optimal filter, the so-called *Wiener* filter, as

$$\mathbf{w}_o = \mathbf{R}_x^{-1} \mathbf{r}_{xd} \quad (2.1.22)$$

noting that \mathbf{R}_x will (typically) be invertible as it contains an additive noise term. The minimum error is found as the lowest value of the cost-function J_W , i.e.,

$$\begin{aligned}
J_{\min} &= E \left\{ \left[d(n) - \hat{d}(n) \right] e^*(n) \right\} \\
&= E \left\{ d(n) \left[d^*(n) - \sum_{k=0}^{M-1} w(k)x^*(n-k) \right] \right\} \\
&= \sigma_d^2 - \sum_{k=0}^{M-1} w(k) E \{ d(n)x^*(n-k) \} \\
&= \sigma_d^2 - \mathbf{r}_{\mathbf{x}d}^H \mathbf{w}_o \\
&= \sigma_d^2 - \mathbf{r}_{\mathbf{x}d}^H \mathbf{R}_x^{-1} \mathbf{r}_{\mathbf{x}d}
\end{aligned} \tag{2.1.23}$$

Similarly, it is easy to show that the cost function can for a general filter, \mathbf{w} , be written as

$$J_W = J_{\min} + (\mathbf{w} - \mathbf{w}_o)^H \mathbf{R} (\mathbf{w} - \mathbf{w}_o) = J_{\min} + \sum_{k=0}^{M-1} \lambda_k |u(k)|^2 \tag{2.1.24}$$

where $u(k)$ is the k th component of \mathbf{u} , with

$$\mathbf{u} = \mathbf{Q}^H (\mathbf{w} - \mathbf{w}_o). \tag{2.1.25}$$

In (2.1.24) and (2.1.25), we have made use of the eigenvalue decomposition of \mathbf{R}_x ,

$$\mathbf{R}_x = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H, \tag{2.1.26}$$

where $\mathbf{\Lambda}$ a diagonal matrix with the eigenvalues, $\{\lambda_k\}_{k=0}^{M-1}$, along its diagonal, and \mathbf{Q} is a unitary matrix¹ formed from the eigenvectors. In the particular case when the desired signal and the additive noise term are uncorrelated, the optimal filtering problem can be further simplified. Then,

$$r_{\mathbf{x}d}(k) = E \{ x(n)d^*(n-k) \} = E \{ d(n)d^*(n-k) \} = r_d(k) \tag{2.1.27}$$

and

$$r_x(k) = E \{ x(n)x^*(n-k) \} = r_d(k) + r_v(k) \tag{2.1.28}$$

yielding

$$\mathbf{R}_x = \mathbf{R}_d + \mathbf{R}_v \tag{2.1.29}$$

¹Recall that a *unitary* matrix is a matrix \mathbf{B} satisfying $\mathbf{B}\mathbf{B}^H = \mathbf{B}^H\mathbf{B} = \mathbf{I}$. In the case of real-valued data, this simplifies to $\mathbf{B}\mathbf{B}^T = \mathbf{B}^T\mathbf{B} = \mathbf{I}$. A matrix satisfying the latter expression is said to be an *orthogonal* matrix. Here, the term *orthonormal* matrix would be more appropriate, but this term is not commonly used.

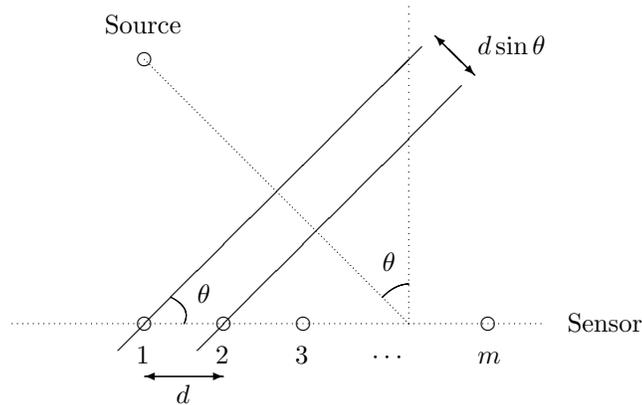


Figure 2.3. Array signal processing

Example 2.1: Let $x(n) = d(n) + v(n)$, where $r_d(k) = |\alpha|^k$, and $v(k)$ is an additive white Gaussian noise sequence with variance σ_v^2 . Design a first-order FIR filter removing as much as possible the influence of $v(n)$.

Solution: Using the fact that \mathbf{R}_x can be written as in (2.1.29), and

$$r_x(k) = \alpha^{|k|} + \sigma_v^2 \delta_K(k) \quad (2.1.30)$$

$$r_{dx}(k) = r_d(k) \quad (2.1.31)$$

we write

$$\begin{bmatrix} 1 + \sigma_v^2 & \alpha \\ \alpha & 1 + \sigma_v^2 \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha \end{bmatrix} \quad (2.1.32)$$

yielding the optimal filter

$$\begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \frac{1}{(1 + \sigma_v^2)^2 - \alpha^2} \begin{bmatrix} 1 + \sigma_v^2 - \alpha^2 \\ \alpha \sigma_v^2 \end{bmatrix} \quad (2.1.33)$$

2.2 Constrained filtering

Wiener filters are designed to minimize the mean-squared estimation error; another common form of filtering is constrained filtering, where one wishes to minimize the mean-squared estimation error given some set of constraints. To illustrate this form of filtering, we will briefly discuss the topic of array signal processing and spatial filtering². Consider a (bandpass) signal source emitting a wavefront that is being

²The presentation given below is somewhat simplified given that we are here primarily interested in illustrating constrained filtering, not discussing array signal processing as such. We refer the interested reader to the nice introduction to the field in [5].

received by an array of m sensors as shown in Figure 2.1. Under the assumption that the source is in the *far-field*, i.e., the wavefronts impinging on the array can be assumed to be planar waves, we can write the time-delay for the wavefronts to imping on the k th sensor, relative to the first sensor, as

$$\tau_k = (k - 1) \frac{d \sin \theta}{c}, \quad (2.2.1)$$

where c is the speed of propagation, θ is the direction of arrival (DOA), and d is the inter-element spacing. Thus, we can write the received signal vector, \mathbf{y}_t , as

$$\mathbf{y}_t = \mathbf{a}_m(\theta)s(t) + \mathbf{e}_t, \quad t = 0, \dots, N - 1 \quad (2.2.2)$$

where \mathbf{e}_t is an additive noise process, and the *array steering vector*, $\mathbf{a}_m(\theta)$, is given as

$$\begin{aligned} \mathbf{a}_m(\theta) &= \left[1 \quad e^{i\omega_c d \sin \theta / c} \quad \dots \quad e^{i(m-1)\omega_c d \sin \theta / c} \right]^T \\ &= \left[1 \quad e^{i\omega_s} \quad \dots \quad e^{i(m-1)\omega_s} \right]^T \end{aligned} \quad (2.2.3)$$

with ω_c denoting the carrier frequency of the source, and

$$\omega_s = \omega_c \frac{d \sin \theta}{c} \quad (2.2.4)$$

is the so-called *spatial frequency*, the resemblance to the temporal case clearly seen by comparing (2.2.3) with (1.2.13). Similar to the temporal case, we can apply filtering on the measured signal; in this case, this *spatial filtering* will steer the receiver beam to a direction of interest, i.e., the spatial filter will enhance signals from a direction θ while suppressing signals from other directions. We write the spatially filtered signal as

$$y_F(t) = \mathbf{h}_\theta^H \mathbf{y}_t = \sum_{k=0}^{M-1} h_\theta^*(k) y(t - k) \quad (2.2.5)$$

where

$$\mathbf{h}_\theta = \left[h_\theta(0) \quad \dots \quad h_\theta(M - 1) \right]^T \quad (2.2.6)$$

Combining (2.2.2) with (2.2.5), we obtain

$$y_F(t) = \left[\mathbf{h}_\theta^H \mathbf{a}_m(\theta) \right] s(t) + \mathbf{h}_\theta^H \mathbf{e}_t \quad (2.2.7)$$

Thus, if we wish to form a spatial filter such that signals from DOA θ are passed undistorted, we require that

$$\mathbf{h}_\theta^H \mathbf{a}_m(\theta) = 1 \quad (2.2.8)$$

while attempting to reduce the influence of signals impinging from directions other than θ . Recalling (1.4.6), we write the power of the filter output as

$$\sigma_\theta^2 = E \{ |y_F(t)|^2 \} = \mathbf{h}_\theta^H \mathbf{R}_y \mathbf{h}_\theta \quad (2.2.9)$$

where

$$\mathbf{R}_y = E \{ \mathbf{y}_t^H \mathbf{y}_t \} \quad (2.2.10)$$

The *Capon*, or *minimum variance distortionless response* (MVDR), beamformer is obtained as the spatial filter such that

$$\mathbf{h}_\theta = \arg \min_{\mathbf{h}_\theta} \mathbf{h}_\theta^H \mathbf{R}_y \mathbf{h}_\theta \quad \text{subject to} \quad \mathbf{h}_\theta^H \mathbf{a}_m(\theta) = 1 \quad (2.2.11)$$

The *linearly constrained optimization problem* in (2.2.11) is typically solved using the method of *Lagrange multipliers*; as this approach is well covered in [11], we will here present a generalization of the optimization problem.

Theorem 2.1: *Let \mathbf{A} be an $(n \times n)$ Hermitian positive definite matrix, and let $\mathbf{X} \in \mathbb{C}^{n \times m}$, $\mathbf{B} \in \mathbb{C}^{n \times k}$ and $\mathbf{C} \in \mathbb{C}^{m \times k}$. Assume that \mathbf{B} has full column rank, k (hence $n \geq k$). Then, the unique solution minimizing*

$$\min_{\mathbf{X}} \mathbf{X}^H \mathbf{A} \mathbf{X} \quad \text{subject to} \quad \mathbf{X}^H \mathbf{B} = \mathbf{C} \quad (2.2.12)$$

is given by

$$\mathbf{X}_o = \mathbf{A}^{-1} \mathbf{B} (\mathbf{B}^H \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{C}^H \quad (2.2.13)$$

Proof: See [5]. □

Thus, the filter minimizing (2.2.11) is, using Theorem 2.1, given as

$$\mathbf{h}_\theta = \frac{\mathbf{R}_y^{-1} \mathbf{a}_m(\theta)}{\mathbf{a}_m^H(\theta) \mathbf{R}_y^{-1} \mathbf{a}_m(\theta)} \quad (2.2.14)$$

which inserted in (2.2.9) yields

$$\sigma_\theta^2 = \mathbf{h}_\theta^H \mathbf{R}_y \mathbf{h}_\theta = \frac{1}{\mathbf{a}_m^H(\theta) \mathbf{R}_y^{-1} \mathbf{a}_m(\theta)} \quad (2.2.15)$$

Note that the Capon beamformer is data-adaptive, placing nulls in the directions of power other than θ ; we can view (2.2.15) as a set of filters in a filterbank, with a separate filter focused at each given direction of interest. The output of this filterbank will form a spatial spectral estimate, with strong peaks in the source directions. By applying the spatial filters corresponding to the source directions, the signals from these may be emphasized while reducing the influence of the other, interfering, sources. We remark that a similar approach may be taken to obtain high-resolution (temporal) spectral estimates. This is done by forming an L -tap data-adaptive (temporal) filter, \mathbf{h}_ω , focused on each frequency of interest, say ω , forming each filter such that

$$\mathbf{h}_\omega = \arg \min_{\mathbf{h}_\omega} \mathbf{h}_\omega^H \mathbf{R}_x \mathbf{h}_\omega \quad \text{subject to} \quad \mathbf{h}_\omega^H \mathbf{a}_L(\omega) = 1 \quad (2.2.16)$$

where \mathbf{R}_x is the covariance matrix in (1.2.10), and $\mathbf{a}_L(\omega)$ is the Fourier vector defined in (1.2.13). Paralleling the spatial case, the resulting *Capon spectral estimator* is obtained as

$$\phi_x^C(\omega) = \frac{1}{\mathbf{a}_L^H(\omega)\mathbf{R}_x^{-1}\mathbf{a}_L(\omega)} \quad (2.2.17)$$

This estimator is far superior to the Periodogram, defined in (1.3.6), for signals peaky spectra. Typically, the Capon spectral estimates have very high resolution while suffering significantly less from the side-lobes distorting the Periodogram (see [5] for a further discussion on the topic). We conclude this chapter by recalling the Gohberg-Semencul formula in (1.5.10), noting that by exploiting the Toeplitz structure of the covariance matrix, the Capon spectral estimate in (2.2.17) can be written as [12]

$$\phi_x^C(\omega) = \frac{\sigma_w^2}{\sum_{s=-L}^L \mu(s)e^{i\omega s}} \quad (2.2.18)$$

where

$$\mu(s) = \sum_{k=0}^{L-s} (L+1-2k-s)a_k^*a_{k+s} = \mu^*(-s), \quad s = 0, \dots, L \quad (2.2.19)$$

The so-called *Musicus* algorithm in (2.2.18) and (2.2.19) enables us to compute the Capon spectral estimate using the Fast Fourier transform (FFT), costing about $\mathcal{O}(L^2 + P \log_2 P)$ operations, where P is the number of frequency grid points desired. This is dramatically less than computing the estimate as given in (2.2.17). Note that to evaluate the estimate in (2.2.18), one requires an estimate of the linear prediction coefficients, a_k . Such an estimate can be obtained in various ways; as typically the different choices of estimation method will result in spectral estimates of different quality, it is important that the selection of estimation procedure is done with some care. Finally, we remark that it is possible to extend the Musicus algorithm to higher dimensional data sets [13].

BIBLIOGRAPHY

- [1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill Book Co., 1984.
- [2] L. C. Ludeman, *Random Processes: Filtering, Estimation and Detection*. John Wiley and Sons, Inc., 2003.
- [3] P. Z. Peebles, Jr., *Probability, Random Variables and Random Signal Principles, 4th Ed.* McGraw-Hill, Inc., 2001.
- [4] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: John Wiley and Sons, Inc., 1996.
- [5] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, N.J.: Prentice Hall, 1997.
- [6] J. C. F. Gauss, *Demonstratio Nova Theorematis Omnem Functionem Algebraicam Rationalem Integram Unius Variabilis in Factores Reales Primi Vel Secundi Gradus Resolve Posse*. PhD thesis, University of Helmstedt, Germany, 1799.
- [7] A. Schuster, "On the Periodicities of Sunspots," *Philos. Trans. of the Royal Society of London*, vol. 206, pp. Series A:69–100, 1906.
- [8] A. Schuster, "On the Investigation of Hidden Periodicities with Application to a Supposed Twenty-Six-Day Period of Meteorological Phenomena," *Terr. Mag.*, vol. 3, pp. 13–41, March 1898.
- [9] I. C. Gohberg and A. A. Semencul, "On the Inversion of Finite Toeplitz Matrices and their Continuous Analogs," *Mat. Issled.*, vol. 2, pp. 201–233, 1972. (In Russian).
- [10] T. Kailath and A. H. Sayed, *Fast Reliable Algorithms for Matrices with Structure*. Philadelphia, USA: SIAM, 1999.
- [11] S. Haykin, *Adaptive Filter Theory (4th edition)*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 2002.

-
- [12] B. Musicus, "Fast MLM Power Spectrum Estimation from Uniformly Spaced Correlations," *IEEE Trans. ASSP*, vol. 33, pp. 1333–1335, October 1985.
- [13] A. Jakobsson, S. L. Marple, Jr., and P. Stoica, "Two-Dimensional Capon Spectrum Analysis," *IEEE Trans. on Signal Processing*, vol. 48, pp. 2651–2661, September 2000.

SUBJECT INDEX

- $\delta_D(\omega)$, 8
- $\delta_K(k)$, 7
- ∇_k , 13
- $\phi_x(\omega)$, 6
- σ_x^2 , 3
- $r_x(k)$, 3
- \mathbf{R}_x , 4
- $\mathbf{a}_L(\omega)$, 5
- $\mathbf{a}_m(\theta)$, 17

- AR process, 10
- ARMA process, 10
- Array steering vector, 17
- Auto-regressive process, 10
- Autocorrelation, 3
- Autocorrelation matrix, 4
- Autocovariance, 3

- Bias, 6

- Capon beamformer, 18
- Capon spectral estimator, 19
- Consistent, 6
- Constrained optimization, 18

- Dirac delta, 8
- DOA, 17

- Eigenvalue extremal property, 8
- Eigenvector, 8
- Endymion, 12
- Ergodic, 5

- Far-field assumption, 17

- FIR, 14
- Fourier vector, 5
- Fourier, J. B. J., 5

- Gauss, J. C. F., 3
- Gohberg-Semencul formula, 11, 19

- Hankel matrix, 5
- Hermite, C., 4
- Hermitian matrix, 4

- Identity matrix, 7
- Independence, 2

- Keats, J., 12
- Kolmogorov's theorem, 2
- Kronecker delta, 7
- Kronecker, L., 7

- Lagrange multiplier, 18

- Mean, 3
- Musicus algorithm, 19
- MVDR beamformer, 18

- Normal equations, 10

- Optimal filtering, 12
- Orthogonal matrix, 15
- Orthogonality principle, 14

- Periodogram, 7
- Power spectral density, 6
- Probability density, 2

Probability distribution, 2

Realization, 1

Schuster, A., 7

Sophocles, 1

Spatial filtering, 17

Spatial frequency, 17

Spectral factorization, 9

Stochastic process, 1

Strict-sense stationarity, 4

Toeplitz matrix, 4, 5

Toeplitz, O., 5

Unitary matrix, 15

Variance, 3

White noise, 3, 7

Wide-sense stationary, 3

Wiener filter, 14

Wiener-Hopf equations, 14

Yule-Walker equations, 10