Chapter 2
Information Measures and Spectral Analysis

The idea of a statistical message source is central to Shannon’s work. The study of random processes had entered into communication before his communication theory. There was a growing understanding of and ability to deal with problems of random noise


In this chapter we introduce the notations and background knowledge that will be used in this book. In particular, the key notions from information theory and stochastic processes, e.g., entropy, mutual information, channel capacity, “water-filling” power allocation, stationarity and asymptotic stationarity, are exposed concisely together with discussions on their properties. In addition, we propose a measure of how Gaussian and white an asymptotically stationary stochastic process is, which is therefore termed Gaussianity-whiteness. Towards this end, we first introduce the notion of negentropy rate, which provides a measure of non-Gaussianity for asymptotically stationary stochastic processes. Then, by combining negentropy rate with spectral flatness in a non-trivial way, the Gaussianity-whiteness is defined. Properties of the notions are also analyzed.

The chapter is organized as follows. Section 2.1 introduces the basic concepts from information theory. Section 2.2 provides the mathematical background in spectral analysis. Section 2.3 gives an overview on the channel capacity proposed by Claude Shannon to characterize the quality of communication channels. Section 2.4 discusses the negentropy and negentropy rate. Section 2.5 is devoted to the spectral flatness and Gaussianity-whiteness.
2.1 Basic Information Concepts

Throughout the book, we consider real-valued continuous random variables and random vectors, and discrete-time stochastic processes (except in Chap. 9). All the random variables, random vectors, and stochastic processes will be assumed zero-mean unless otherwise specified. All the random variables and random vectors will be represented in boldface letters. The covariance matrix of a random vector \( \mathbf{x} \in \mathbb{R}^m \) is denoted as \( \Sigma_x = \mathbb{E} \left[ \mathbf{x} \mathbf{x}^T \right] \). Given a vector process \( \{ \mathbf{x}_k \} \), we adopt \( \mathbf{x}_{0\ldots,k} \) to denote the sequence \( \mathbf{x}_0, \ldots, \mathbf{x}_k \). By a slight abuse of notation, \( \mathbf{x}_{0\ldots,k} \) is also identified with the random vector \( [\mathbf{x}_0^T, \ldots, \mathbf{x}_k^T]^T \).

The logarithm is defined with base 2. All the functions are assumed to be measurable, and all the integrals are defined over certain appropriate sets on which the variables are measurable. We say that a system \( F \) is causal if for any \( k \in \mathbb{N} \), \( y_k = F_k (\mathbf{x}_{0\ldots,k}) \), where \( \{ \mathbf{x}_k \} \) and \( \{ y_k \} \) are the input and output processes of the system, respectively. Moreover, a system \( F \) is said to be strictly causal if for any \( k \in \mathbb{N} \), \( y_k = F_k (\mathbf{x}_{0\ldots,k-1}) \).

Entropy and mutual information are the two most basic notions in information theory [34, 117, 142, 143].

**Definition 2.1** The differential entropy of a random vector \( \mathbf{x} \in \mathbb{R}^m \) with density \( p_x (x) \) is defined as

\[
h (\mathbf{x}) = - \int p_x (x) \log p_x (x) \, dx.
\]

**Definition 2.2** The conditional differential entropy of random vector \( \mathbf{x} \in \mathbb{R}^m_1 \) given random vector \( \mathbf{y} \in \mathbb{R}^m_2 \) with joint density \( p_{x,y} (x, y) \) and conditional density \( p_{x|y} (x, y) \) is defined as

\[
h (\mathbf{x} | \mathbf{y}) = - \int p_{x,y} (x, y) \log p_{x|y} (x, y) \, dx \, dy.
\]

**Definition 2.3** The mutual information between random vectors \( \mathbf{x} \in \mathbb{R}^{m_1}, \mathbf{y} \in \mathbb{R}^{m_2} \) with densities \( p_x (x), p_y (y) \) and joint density \( p_{x,y} (x, y) \) is defined as

\[
I (\mathbf{x}; \mathbf{y}) = \int p_{x,y} (x, y) \log \frac{p_{x,y} (x, y)}{p_x (x) p_y (y)} \, dx \, dy.
\]

**Definition 2.4** The entropy rate of a vector stochastic process \( \{ \mathbf{x}_k \}, \mathbf{x}_k \in \mathbb{R}^m \) is defined as

\[
h_\infty (\mathbf{x}) = \lim sup_{k \to \infty} \frac{h (\mathbf{x}_{0\ldots,k})}{k + 1}.
\]
Definition 2.5  The (mutual) information rate between two stochastic processes \( x \in \mathbb{R}^{m_1}, y \in \mathbb{R}^{m_2} \) is defined as

\[
I_\infty (x; y) = \lim_{k \to \infty} \sup \frac{I(x_0, \ldots, k; y_0, \ldots, k)}{k + 1}.
\]

Definition 2.6  The entropy power of a random vector \( x \in \mathbb{R}^m \) is defined as

\[
N(x) = \frac{1}{2\pi e} 2^{\frac{1}{2} h(x)}.
\]

Entropy rate power generalizes entropy power to stochastic processes.

Definition 2.7  The entropy rate power of a stochastic process \( \{x_k\}, x_k \in \mathbb{R}^m \) is defined as

\[
N_\infty (x) \triangleq \frac{1}{2\pi e} 2^{\frac{1}{2} h_\infty (x)}.
\] (2.1)

The following lemma lists the key properties of entropy and mutual information that are relevant to our subsequent development [34].

Lemma 2.1  Let \( x \in \mathbb{R}^{m_1}, y \in \mathbb{R}^{m_2}, z \in \mathbb{R}^{m_3} \) be random vectors, and \( \{x_k\}, x_k \in \mathbb{R}^m \) be a vector process. Suppose that \( f (\cdot) \) is a measurable function defined on an appropriate space. Then, the following relations hold:

- \( I(x; y) = I(y; x) = h(x) - h(x|y) = h(y) - h(y|x) \);
- \( h(x, y) = h(x) + h(y|x) \);
- \( I(x; y) = h(x) + h(y) - h(x, y) \);
- \( h(x|y) \leq h(x|f(y)) \), and \( I(x; y) \geq I(x; f(y)) \), where the equalities hold if \( f(\cdot) \) is injective;
- \( I(x; y|z) = h(x|z) - h(x|y, z) \);
- \( h(x|y) = h(x + f(y)|y) \), \( h(x|y) = h(x|y, f(y)) \), and \( I(x; y|z) = I(x; y + f(z)|z) \);
- \( I(x; y, z) = I(x; y) + I(x; z|y) \);
- \( I(x; y) \leq I(x; y, z) \), and \( h(x|y, z) \geq h(x|y, z) \), where the equalities hold if and only if \( x \) and \( z \) are independent given \( y \);
- \( h(x_0, \ldots, k) = \sum_{i=0}^{k} h(x_i|x_0, \ldots, i-1) \);
- \( I(x_0, \ldots, k; y) = \sum_{i=0}^{k} I(x_i; y|x_0, \ldots, i-1) \);
- \( h(x_0, \ldots, k) \leq \sum_{i=0}^{k} h(x_i) \), where the equality holds if and only if \( x_0, \ldots, x_k \) are mutually independent;
- If \( c \in \mathbb{R}^{m_1} \) is a constant vector, then \( h(x + c) = h(x) \);
- If \( A \in \mathbb{R}^{m \times m_1} \) is a constant matrix and \( \det A \neq 0 \), then \( h(Ax) = h(x) + \log |\det A| \).

The next lemma [34, 115] relates the entropy of a random vector to its covariance matrix.
Lemma 2.2 Suppose that $x \in \mathbb{R}^m$ is a random vector with covariance matrix $\Sigma_x$. Then,

$$h(x) \leq \log \sqrt{(2\pi e)^m \det \Sigma_x},$$

where the equality holds if and only if $x$ is Gaussian.

The next lemma introduces the entropy power inequality [34], which will play an important role in the sequel.

Lemma 2.3 For any independent random vectors $x, y \in \mathbb{R}^m$,

$$2^\frac{1}{2} h(x+y) \geq 2^\frac{1}{2} h(x) + 2^\frac{1}{2} h(y),$$

where the equality holds if and only if $x$ and $y$ are Gaussian with proportional covariance matrices, i.e.,

$$\Sigma_y = \alpha \Sigma_x,$$

for some $\alpha > 0$.

The following lemma is adapted from [171].

Lemma 2.4 Consider the random vectors $x, y \in \mathbb{R}^m$. Suppose that $x$ is Gaussian, and that $x$ and $y$ are independent. Then, for any Gaussian random vector $y_G \in \mathbb{R}^m$ with the same covariance matrix as $y$,

$$I(x; x + y) \geq I(x; x + y_G).$$

The following lemma will also be useful.

Lemma 2.5 Consider the random vectors $x_1, y_1 \in \mathbb{R}^{m_1}$, $x_2, y_2 \in \mathbb{R}^{m_2}$. Suppose that $[x_1^T, x_2^T]^T$ and $[y_1^T, y_2^T]^T$ are independent, and that $y_1$ and $y_2$ are independent. Then,

$$I(x_1 + y_1; x_2 + y_2) \leq I(x_1; x_2),$$

or equivalently,

$$h(x_1 + y_1, x_2 + y_2) - h(x_1, x_2) \geq h(x_1 + y_1) + h(x_2 + y_2) - h(x_1) - h(x_2),$$

where the equalities hold if $x_1$ and $x_2$ are independent.

Proof Since $y_1$ and $x_2 + y_2$ are independent given $x_1$, we have

$$I(x_1 + y_1; x_2 + y_2) \leq I(x_1 + y_1, y_1; x_2 + y_2) = I(x_1, y_1; x_2 + y_2)$$

$$= I(x_1; x_2 + y_2) + I(y_1; x_2 + y_2 | x_1) = I(x_1; x_2 + y_2).$$
In the inequality above, the equality holds if \( x_1 \) and \( x_2 \) are independent. In addition, since \( x_1 \) and \( y_2 \) are independent given \( x_2 \), we have

\[
I(x_1; x_2 + y_2) \leq I(x_1; x_2 + y_2, y_2) = I(x_1; x_2, y_2) = I(x_1; x_2).
\]

The equality in the inequality above holds if \( x_1 \) and \( x_2 \) are independent. Hence,

\[
I(x_1 + y_1; x_2 + y_2) \leq I(x_1; x_2),
\]

and

\[
h(x_1 + y_1, x_2 + y_2) - h(x_1 + y_1) - h(x_2 + y_2) \geq h(x_1, x_2) - h(x_1) - h(x_2),
\]

where the equalities hold if \( x_1 \) and \( x_2 \) are independent. □

One interpretation of this result is that “adding totally independent information will not increase mutual information.”

### 2.2 Spectral Analysis

The average power of a stochastic process is defined as follows.

**Definition 2.8** The average power of a stochastic process \( \{x_k\}, x_k \in \mathbb{R}^m \) is defined as

\[
\text{pow}(x) = \lim_{k \to \infty} \frac{E \left[ \sum_{i=0}^{k} \sum_{j=1}^{m} x_i^2 \right]}{k + 1}. \tag{2.4}
\]

Stationarity and asymptotic stationarity are two important concepts in the study of stochastic processes [115].

**Definition 2.9** A stochastic process \( \{x_k\}, x_k \in \mathbb{R}^m \) is said to be (wide-sense) stationary if \( R_x(i, k) = E \left[ (x_i - E[x]) (x_{i+k} - E[x])^T \right] \) depends only on \( k \), where \( E[x] = \lim_{i \to \infty} E[x_i] \). For a stationary process \( \{x_k\} \), we write \( R_x(i, k) = R_x(k) \).

**Definition 2.10** The power spectrum (or power spectral density) of a stationary stochastic process \( \{x_k\}, x_k \in \mathbb{R}^m \) is defined as

\[
\Phi_x(\omega) = \sum_{k=-\infty}^{\infty} R_x(k) e^{-j\omega k}.
\]
It is well-known that $\Phi_x (\omega)$ is positive semidefinite. In addition,

$$R_x (k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_x (\omega) e^{j\omega k} d\omega,$$

and

$$R_x (0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_x (\omega) d\omega.$$

**Definition 2.11** A stochastic process $\{x_k\}, x_k \in \mathbb{R}^m$ is asymptotically stationary if the limit $E [x] = \lim_{i \to \infty} E [x_i]$ exists, and if the following limit exists for any integer $k$:

$$R_x (k) = \lim_{i \to \infty} E \left[ (x_i - E [x]) (x_{i+k} - E [x])^T \right].$$

**Definition 2.12** The asymptotic power spectrum of an asymptotically stationary stochastic process $\{x_k\}, x_k \in \mathbb{R}^m$ is defined as

$$\Phi_x (\omega) = \sum_{k=-\infty}^{\infty} R_x (k) e^{-j\omega k}.$$

Similarly, $\Phi_x (\omega)$ is positive semidefinite. Moreover,

$$R_x (k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_x (\omega) e^{j\omega k} d\omega,$$

and

$$R_x (0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_x (\omega) d\omega.$$

Throughout this book, we deal with asymptotically stationary processes. Since stationarity can be viewed as a special case of asymptotic stationarity, all the results derived for asymptotically stationary processes are also valid for stationary processes.

The following result [115] relates the power spectra of the input and output processes of a stable LTI system.

**Lemma 2.6** Consider an $m$-input, $m$-output stable LTI system with transfer function $L (z)$. If the input $\{x_k\}, x_k \in \mathbb{R}^m$ is an asymptotically stationary process, then the output $\{y_k\}, y_k \in \mathbb{R}^m$ is also an asymptotically stationary process. In addition,

$$\det \Phi_y (\omega) = \left| \det L (e^{j\omega}) \right|^2 \det \Phi_x (\omega),$$
where
\[ L(e^{j\omega}) = L(z) \big|_{z=e^{j\omega}}, \]
and \( \Phi_x(\omega) \) and \( \Phi_y(\omega) \) are the asymptotic power spectra of \( \{x_k\} \) and \( \{y_k\} \), respectively.

The following lemma [73] relates the entropy rate of a vector stochastic process to its asymptotic power spectrum.

**Lemma 2.7** Suppose that \( \{x_k\}, x_k \in \mathbb{R}^m \) is asymptotically stationary with asymptotic power spectrum \( \Phi_x(\omega) \). Then,
\[ h_\infty(x) \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \sqrt{(2\pi e)^m \det \Phi_x(\omega)} \, d\omega, \]
where the equality holds if and only if \( \{x_k\} \) is Gaussian.

The next lemma [73, 183] characterizes how the entropy rates of the input and output processes of a stable LTI system are related.

**Lemma 2.8** Consider an \( m \)-input, \( m \)-output stable LTI system with transfer function \( L(z) \), input process \( \{x_k\}, x_k \in \mathbb{R}^m \), and output process \( \{y_k\}, y_k \in \mathbb{R}^m \). Then,
\[ h_\infty(y) = h_\infty(x) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\det L(e^{j\omega})| \, d\omega. \]

In the scalar case, we have [115]
\[ h_\infty(y) = h_\infty(x) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |L(e^{j\omega})| \, d\omega. \]

Note that herein, the equalities do not require the input and output processes to be Gaussian. One may refer to [73, 115, 180] for a detailed proof. Interestingly, as early as in [142, 143], an alternative proof was provided by Claude Shannon.

The following lemma will also be useful.

**Lemma 2.9** Suppose that \( x, y \in \mathbb{R}^m \) are random vectors with positive definite covariance matrices \( \Sigma_x, \Sigma_y \), respectively. In addition, suppose that \( x \) and \( y \) are independent, and that \( y(1), \ldots, y(l) \) are mutually independent. Then,
\[ \frac{\det (\Sigma_x + \Sigma_y)}{\det \Sigma_x} \geq \prod_{j=1}^{m} \left[ \frac{\sigma_x^2(j) + \sigma_y^2(j)}{\sigma_x^2(j)} \right], \quad (2.5) \]
where \( \sigma_x^2(j), \sigma_y^2(j) \) denote the variances of \( x(j), y(j) \), respectively, and the equality holds if \( x(1), \ldots, x(l) \) are mutually independent.
Proof Suppose that \( x, y \) are Gaussian. It then follows that

\[
h(x) = \log \sqrt{(2\pi e)^m \det \Sigma_x}.
\]

Furthermore, since \( x \) and \( y \) are independent, we have

\[
h(x + y) = \log \sqrt{(2\pi e)^m \det \Sigma_{x+y}} = \log \sqrt{(2\pi e)^m \det (\Sigma_x + \Sigma_y)}.
\]

On the other hand, using Lemma 2.5, we have

\[
h(x + y) - h(x) = h(x(1) + y(1), x(2) + y(2), \ldots, x(m) + y(m)) - h(x(1), x(2), \ldots, x(m)) \\
\geq h(x(1) + y(1)) + h(x(2) + y(2), \ldots, x(m) + y(m)) \\
- h(x(1)) - h(x(2), \ldots, x(m)) \\
\geq \ldots \\
\geq h(x(1) + y(1)) + \cdots + h(x(m) + y(m)) - h(x(1)) - \cdots - h(x(m)).
\]

Herein, if \( x(1), \ldots, x(m) \) are mutually independent, the lower bound becomes an equality as

\[
h(x + y) - h(x) = h(x(1) + y(1)) + \cdots + h(x(m) + y(m)) - h(x(1)) - \cdots - h(x(m)).
\]

As such,

\[
\log \sqrt{\frac{\det (\Sigma_x + \Sigma_y)}{\det (\Sigma_x)}} = \log \sqrt{(2\pi e)^m \det (\Sigma_x + \Sigma_y)} - \log \sqrt{(2\pi e)^m \det (\Sigma_x)} \\
= h(x + y) - h(x) \\
\geq h(x(1) + y(1)) + \cdots + h(x(m) + y(m)) - h(x(1)) - \cdots - h(x(m)) \\
= \log 2\pi e \sigma_{x(1)+y(1)}^2 + \cdots + \log 2\pi e \sigma_{x(m)+y(m)}^2 \\
- \log 2\pi e \sigma_{x(1)}^2 - \cdots - \log 2\pi e \sigma_{x(m)}^2 \\
= \log \prod_{j=1}^m \left[ \frac{\sigma_{x(j)}^2 + \sigma_{y(j)}^2}{\sigma_{x(j)}^2} \right].
\]
If $x, y$ are not Gaussian, then let $x_G$ and $y_G$ be Gaussian vectors with the same covariance matrices as $x$ and $y$, respectively. Similarly, we can prove that

$$\frac{\det (\Sigma + \Sigma_y)}{\det \Sigma_x} = \frac{\det (\Sigma_{x_G} + \Sigma_{y_G})}{\det \Sigma_{x_G}} \geq \prod_{j=1}^{m} \left[ \frac{\sigma_{x,G}(j)^2 + \sigma_{y,G}(j)^2}{\sigma_{x,G}(j)^2} \right] = \prod_{j=1}^{m} \left[ \frac{\sigma_{x}(j)^2 + \sigma_{y}(j)^2}{\sigma_{x}(j)^2} \right].$$

\[\square\]

### 2.3 Channel Capacity

Channel capacity [34, 142, 143, 159] is a well-known concept in information theory and communication engineering. Defined as the supremum of the mutual information (rate) between the channel input and output, it provides a tight upper bound on the rate at which information can be reliably transmitted over communication channels.

**Definition 2.13** Consider a general causal noisy channel, as depicted in Fig. 2.1, with input process $\{v_k\}, v_k \in \mathbb{R}^m$ and output process $\{u_k\}, u_k \in \mathbb{R}^m$. The channel capacity, measured in bits, is defined as

$$C = \sup_{p_x} I_\infty (v; u) = \sup_{p_x} \limsup_{k \to \infty} I (v_0, \ldots, v_k; u_0, \ldots, u_k),$$

where the supremum is taken over all possible densities $p_x$ of the input process allowed for the channel.

The next definition introduces feedback capacity, i.e., channel capacity with feedback.

**Fig. 2.1** A noisy channel with input $v$, noise $n$, and output $u$
Definition 2.14 Consider a general causal noisy channel with feedback as depicted in Fig. 2.2. The feedback capacity of the channel is defined as

\[
C_f = \sup_{p_f(v)} I_{\infty}(v; u) = \sup_{p_f(v)} \lim_{k \to \infty} \frac{I(v_{0,...,k}; u_{0,...,k})}{k + 1},
\]

where the supremum is taken over all possible densities \(p_f(v)\) of the input process allowed for the channel with feedback.

The difference between feedback capacity and capacity without feedback lies in that feedback capacity allows the current input of the channel to depend on the past values of the output. In other words, in the case of capacity without feedback, for \(i = 1, \ldots, k\), \(v_i\) is independent of \(u_{0,...,i-1}\). On the other hand, in the case of feedback capacity, for \(i = 1, \ldots, k\), \(v_i\) may be dependent on \(u_{0,...,i-1}\).

It is clear that feedback may increase channel capacity.

Lemma 2.10 Feedback capacity is always larger than or equal to channel capacity, i.e.,

\[
C_f \geq C.
\]

In what follows, we introduce three canonical classes of channels and characterize more explicitly their channel capacity for both SISO and MIMO cases.

We say that a channel is additive if

\[
u_k = v_k + n_k.
\]

Furthermore, if \(\{n_k\}\) is a white Gaussian process independent of \(\{v_k\}\), then we say that the channel is an additive white Gaussian noise (AWGN) channel.

Let the input of the channel be constrained in power as

\[
\lim_{k \to \infty} \frac{E\left[\sum_{i=0}^{k} \sum_{j=1}^{m} v_i^2(j)\right]}{k + 1} \leq P.
\] (2.6)
In the scalar case,
\[
\lim_{k \to \infty} \frac{E \left[ \sum_{i=0}^{k} v_i^2 \right]}{k + 1} \leq P. \tag{2.7}
\]

**Lemma 2.11** The channel capacity of a scalar AWGN channel with noise variance \( N \) and power constraint (2.7) is given by
\[
C = \frac{1}{2} \log \left( 1 + \frac{P}{N} \right).
\]

**Lemma 2.12** The feedback capacity of a scalar AWGN channel with noise variance \( N \) and power constraint (2.7) is given by
\[
C_f = C = \frac{1}{2} \log \left( 1 + \frac{P}{N} \right).
\]

The lemmas show that feedback does not increase capacity of scalar AWGN channels. Further results on feedback capacity can be found in [33, 34, 79] and the references therein. It can be seen that feedback increases capacity for other classes of channels.

We continue with the second class of channels. For an additive channel
\[
u_k = v_k + n_k,
\]
if \( \{n_k\} \) is a stationary colored Gaussian process with a non-constant power spectrum independent of \( \{v_k\} \), then we say that the channel is an additive colored Gaussian noise (ACGN) channel.

**Lemma 2.13** Consider a scalar ACGN channel. Let the noise \( \{n_k\} \) be a stationary and additive colored Gaussian noise with power spectrum \( N(\omega) \). The capacity of the channel with power constraint (2.7) is given by
\[
C = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \left[ 1 + \frac{P(\omega)}{N(\omega)} \right] d\omega,
\]
where the power to be used at frequency \( \omega \) is
\[
P(\omega) = \max \{0, \zeta - N(\omega)\},
\]
and \( \zeta \) is such a normalizing scalar that
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) d\omega = P.
\]
More generally, for an MIMO AWGN channel, we have \( u_k = v_k + n_k \), with \( u_k, v_k, n_k \in \mathbb{R}^m \), where \( \{n_k\} \) is a white Gaussian process, and \( \{n_k\} \) and \( \{v_k\} \) are independent.

**Lemma 2.14** Consider a set of \( m \) parallel AWGN channels with noise covariance matrix \( \Sigma_n = U_n \Lambda_n U_n^T \), where \( \Lambda_n = \text{diag} (\Lambda_1, \ldots, \Lambda_m) \) and \( U_n \) is an orthogonal matrix. Assume that the power constraint inequality given in (2.6) holds. Then, the channel capacity is given by

\[
C = \sum_{j=1}^{m} \frac{1}{2} \log \left( 1 + \frac{P_j}{N_j} \right),
\]

where the power \( P_j \) to be used in the channel \( j \) is

\[
P_j = \max \left\{ 0, \zeta - N_j \right\},
\]

with \( \zeta \) satisfying

\[
\sum_{j=1}^{m} P_j = P.
\]

This result represents the famed “water-filling” power allocation policy [34]. Loosely speaking, this optimal policy mandates that more power is to be delivered to less noisy channels.

MIMO ACGN channels constitute a direct generalization of MIMO AWGN channels. In this channel model, the noise \( \{n_k\} \) is assumed to be a stationary and additive colored Gaussian process with a non-constant power spectrum, and \( \{n_k\} \) and \( \{v_k\} \) are still assumed to be independent.

**Lemma 2.15** Let the noise \( \{n_k\}, n_k \in \mathbb{R}^m \) be a stationary and additive Gaussian colored noise with power spectrum \( \Phi_n (\omega) = U_n (\omega) \Lambda_n (\omega) U_n^T (\omega) \), where \( \Lambda_n (\omega) = \text{diag} (N_1 (\omega), \ldots, N_m (\omega)) \) and \( U_n (\omega) \) is an orthogonal matrix. Then, the channel capacity with power constraint (2.6) is given by

\[
C = \sum_{j=1}^{m} \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \left[ 1 + \frac{P_j (\omega)}{N_j (\omega)} \right] d\omega,
\]

where the power to be used in channel \( j \) at frequency \( \omega \) is given by

\[
P_j (\omega) = \max \left\{ 0, \zeta - N_j (\omega) \right\},
\]
and $\zeta$ satisfies

$$\sum_{j=1}^{m} \frac{1}{2\pi} \int_{-\pi}^{\pi} P_j(\omega) \, d\omega = P.$$  

The third class of channels is the fading channels [159]. In the scalar case, the fading channel can be described by the general channel input-output relation

$$u_k = h_k v_k + n_k,$$

where $\{v_k\}, v_k \in \mathbb{R}$ is the channel input, $\{h_k\}, h_k \in \mathbb{R}$ is the fading process, and $\{n_k\}, n_k \in \mathbb{R}$ is an additive white Gaussian process. Moreover, $\{n_k\}, \{v_k\},$ and $\{h_k\}$ are mutually independent. A fading channel is said to be fast fading if the fading time of $\{h_k\}$ is relatively small compared to transmission duration, i.e., if the codeword length spans many coherence periods. A fading channel is said to be with transmitter/receiver side information if the transmitter/receiver know the values of the realization of $\{h_k\}$ for each $k$.

**Lemma 2.16** Consider the fast fading channel with power constraint (2.7) and noise variance $N$. If the channel is only with receiver side information, then its channel capacity is given by

$$C = E_k \left[ \frac{1}{2} \log \left( 1 + \frac{|h_k|^2 P}{N} \right) \right].$$

If the fast fading channel is with both receiver side and transmitter side information, then its channel capacity is given by

$$C = E_k \left[ \frac{1}{2} \log \left( 1 + \frac{|h_k|^2 P_k}{N} \right) \right],$$

where

$$P_k = \max \left\{ 0, \zeta - \frac{N}{|h_k|^2} \right\},$$

and $\zeta$ is a normalizing scalar satisfying

$$E_k [P_k] = P.$$  

We give below two examples of fading channels.
Example 2.1  Let the channel gain $h_k$ have a Rayleigh distribution with probability density function

$$f(x) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, & \text{if } x \geq 0, \\ 0, & \text{otherwise}. \end{cases}$$

Then, this channel is called a Rayleigh fast fading channel [159]. Suppose that the channel input satisfies the constraint (2.7), and the noise has a variance $N$. If the channel is only with receiver side information, then its channel capacity can be found as

$$C = \frac{N}{2P\sigma^2} E_1 \left( \frac{N}{2P\sigma^2} \right),$$

where

$$E_1(x) = \int_1^\infty \frac{e^{-xt}}{t} dt.$$

Example 2.2  Let $h_k$ have a Weibull distribution [129] with probability density function

$$f(x) = \begin{cases} \frac{k}{\lambda} \left( \frac{x}{\lambda} \right)^{k-1} e^{-\frac{x}{\lambda}}, & \text{if } x \geq 0, \\ 0, & \text{otherwise}, \end{cases}$$

where $k > 0, \lambda > 0$. We call this channel a Weibull fast fading channel [129]. Suppose that the channel input satisfies the constraint (2.7), and the noise has a variance $N$. If the channel is only with receiver side information, then its channel capacity can be found as

$$C = \int_0^\infty \left[ \frac{1}{2} \log (1 + y) \right] \frac{k}{2\lambda \sqrt{P_N/N}} \left( \frac{1}{\lambda} \sqrt{\frac{Ny}{P}} \right)^{k-1} e^{-\frac{1}{2} \sqrt{\frac{Ny}{P}}} dy.$$  

An MIMO fading channel [159] can generally be described by the channel input-output relation

$$u_k = H_k v_k + n_k,$$

where $\{v_k\}, v_k \in \mathbb{R}^m$ is the channel input, $\{H_k\}, H_k \in \mathbb{R}^{m \times m}$ is the fading process, and $\{n_k\}, n_k \in \mathbb{R}^m$ is an additive white Gaussian process. Moreover, $\{n_k\}, \{v_k\}$, and $\{H_k\}$ are mutually independent.
Lemma 2.17 Consider the MIMO fast fading channel with power constraint (2.6) and covariance matrix $\Sigma_n$. Let $H_k^{-1} \Sigma_n H_k^{-T} = U_n \Lambda_n U_n^T$, for $k \in \mathbb{N}$, where $\Lambda_n = \text{diag} (N_k(1), \ldots, N_k(m))$, and $U_n$ is an orthogonal matrix. If the channel is with both transmitter and receiver side information, i.e., both the transmitter and the receiver know the channel gain $\{H_k\}$, then the channel capacity is given by

$$C = \mathbb{E}_k \left[ \sum_{j=1}^{m} \frac{1}{2} \log \left[ 1 + \frac{P_k(j)}{N_k(j)} \right] \right],$$

where the power to be used in channel $j$ at time $k$ is

$$P_k(j) = \max \{0, \zeta - N_k(j)\},$$

and $\zeta$ satisfies

$$\mathbb{E}_k \left[ \sum_{j=1}^{m} P_k(j) \right] = P.$$

2.4 Negentropy and Negentropy Rate

We now generalize the notion of negentropy of random variables to stochastic processes. As a measure of non-Gaussianity for random variables, negentropy has been used extensively in independent component analysis [68].

Definition 2.15 The negentropy (or negative entropy) of a random vector $x \in \mathbb{R}^m$ is defined as

$$J(x) = h(x_G) - h(x),$$

where $x_G$ is a Gaussian vector with the same covariance matrix as $x$.

It is known that $J(x) \geq 0$, and that $J(x) = 0$ if and only if $x$ is Gaussian. As such, negentropy is a measure of non-Gaussianity for random vectors. Let the covariance matrix of $x$ be given by $\Sigma_x$. It follows that

$$h(x_G) = \log \sqrt{(2\pi e)^m} \det \Sigma_{x_G} = \log \sqrt{(2\pi e)^m} \det \Sigma_x,$$

and

$$J(x) = h(x_G) - h(x) = \log \sqrt{(2\pi e)^m} \det \Sigma_x - h(x).$$
Consider the linear transformation \( y = Ax \), where \( x, y \in \mathbb{R}^m \), \( A \in \mathbb{R}^{m \times m} \), and \( \det A \neq 0 \). Then,

\[
J (y) = \log \sqrt{(2\pi e)^m \det \Sigma_y} - h(y) = \log \sqrt{(2\pi e)^m \det A \Sigma_x A^T} - h(x) - \log |\det A|
\]

\[
= \log \sqrt{(2\pi e)^m \det \Sigma_x} - h(x) = J (x) .
\]

(2.8)

In other words, negentropy is invariant under linear transformations, which in general is not true for differential entropy.

Consider next a random variable \( x \in \mathbb{R} \) distributed uniformly from \(-a\) to \(a\) with \(a > 0\). Let its density be

\[
p_x (x) = \begin{cases} 
\frac{1}{2a}, & \text{if } -a \leq x \leq a, \\
0, & \text{otherwise}.
\end{cases}
\]

As a result, \( \sigma_x^2 = a^2 / 3 \), and

\[
h (x) = - \int_{-a}^{a} \frac{1}{2a} \log \left( \frac{1}{2a} \right) \, dx = \log (2a) .
\]

In this case,

\[
J (x) = \log \sqrt{2\pi e \sigma_x^2} - h (x) = \log \sqrt{\frac{\pi e}{6}}.
\]

(2.9)

In other words, \( J (x) \) is independent of \( a \). To put it another way, all uniform distributions have the same degree of non-Gaussianity (Gaussianity), which is approximately 0.25461. Note that the logarithm is with base 2 herein.

Table 2.1 gives the negentropy for a wide class of random variables. The distributions include the triangular distribution, the exponential distribution, the double exponential distribution, the Rayleigh distribution, the Erlang distribution, and the Gamma distribution.

We then define negentropy rate for asymptotically stationary processes.

**Definition 2.16** The negentropy rate (or negative entropy rate) of an asymptotically stationary process \( \{ x_k \} \), \( x_k \in \mathbb{R}^m \) is defined as

\[
J_\infty (x) \triangleq h_\infty (x_G) - h_\infty (x) ,
\]

(2.10)

where \( \{ x_G (k) \} \), \( x_G (k) \in \mathbb{R}^m \) is a Gaussian process with the same asymptotic power spectrum as \( \{ x_k \} \).

The following proposition provides a key link between entropy domain analysis and frequency domain analysis of asymptotically stationary processes, by relating the negentropy rate to the asymptotic power spectrum.
Negentropy and Negentropy Rate

Table 2.1 Negentropy of typical distributions

<table>
<thead>
<tr>
<th>Distribution of $x$</th>
<th>Density function $p_x(x)$</th>
<th>Negentropy $J(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triangular</td>
<td>$\begin{cases} a + ax^2, &amp; \text{if } -\frac{1}{a} \leq x \leq 0 \ a - ax^2, &amp; \text{if } 0 \leq x \leq \frac{1}{a} \ 0, &amp; \text{otherwise} \end{cases}$</td>
<td>$\log \sqrt{\frac{\pi}{3}}$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$ae^{-ax}$, if $x \geq 0$</td>
<td>$\log \sqrt{\frac{2\pi}{e}}$</td>
</tr>
<tr>
<td>Double exponential</td>
<td>$\frac{1}{2}ae^{-a</td>
<td>x</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>$\begin{cases} \frac{xe^{-x^2/2a^2}}{a^2}, &amp; \text{if } x \geq 0 \ 0, &amp; \text{otherwise} \end{cases}$</td>
<td>$\log \sqrt{\frac{2\pi(4-\pi)}{e^{\mu+1}}}$</td>
</tr>
<tr>
<td>Erlang</td>
<td>$\begin{cases} \frac{d^{n+1}x^n e^{-ax}}{n!}, &amp; \text{if } x \geq 0 \ 0, &amp; \text{otherwise} \end{cases}$</td>
<td>$\log \sqrt{\frac{2\pi(n+1)}{(n!)^2}e^{-2\mu(\theta)\psi(\theta)-(1-\theta)\psi(\theta)-\log[\Gamma(\theta)]}}$</td>
</tr>
<tr>
<td>Gamma</td>
<td>$\begin{cases} \frac{a^\theta x^{\theta-1}e^{-ax}}{\Gamma(\theta)}, &amp; \text{if } x \geq 0 \ 0, &amp; \text{otherwise} \end{cases}$</td>
<td>$\log \sqrt{2\pi e\theta - \theta - \log[\Gamma(\theta)] - (1-\theta)\psi(\theta)}$</td>
</tr>
</tbody>
</table>

$a > 0$, $n \in \mathbb{N}$ and $n > 1$, $\theta > 0$ and $\Gamma(\theta)$ is the Gamma function, $\mu$ is Euler’s constant

**Proposition 2.1** Suppose that the asymptotic power spectrum of $\{x_k\}, x_k \in \mathbb{R}^m$ is given by $\Phi_x(\omega)$. Then,

$$J_\infty (x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \sqrt{(2\pi e)^m \det \Phi_x(\omega)} d\omega - h_\infty (x). \tag{2.11}$$

In addition, $J_\infty (x) \geq 0$, and $J_\infty (x) = 0$ if and only if $\{x_k\}$ is Gaussian.

**Proof** Since

$$h_\infty (x) \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \sqrt{(2\pi e)^m \det \Phi_x(\omega)} d\omega,$$

where the equality holds if and only if $\{x_k\}$ is Gaussian, we have

$$h_\infty (x_G) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \sqrt{(2\pi e)^m \det \Phi_{x_G}(\omega)} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \sqrt{(2\pi e)^m \det \Phi_x(\omega)} d\omega.$$

As a result,

$$J_\infty (x) = h_\infty (x_G) - h_\infty (x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \sqrt{(2\pi e)^m \det \Phi_x(\omega)} d\omega - h_\infty (x).$$
Clearly, $J_\infty (x) \geq 0$. Furthermore, $J_\infty (x) = 0$ if and only if $\{x_k\}$ is Gaussian.

Hence, negentropy rate provides a measure of non-Gaussianity for asymptotically stationary processes. We now show that the negentropy rate of a white process is equal to the negentropy of any of its sample, independent of time.

**Proposition 2.2** If $\{x_k\}$ is white with covariance matrix $\Sigma_x$, then

$$J_\infty (x) = \log \sqrt{2\pi e^m \det \Sigma_x} - h(x_k) = J(x_k), \forall k \in \mathbb{N}. \quad (2.12)$$

**Proof** Since $\{x_k\}$ is white, we have

$$h_\infty (x) = \lim_{k \to \infty} \sup h(x_0) + \cdots + h(x_k) = h(x_k),$$

and

$$R_x (l) = \lim_{i \to \infty} E[x_i x_i^T] = R_x (l) \delta (l),$$

where $\delta (\cdot)$ is the Dirac delta function. Hence,

$$J_\infty (x) = \log \sqrt{2\pi e^m \det R_x (0)} - h(x_k) = \log \sqrt{2\pi e^m \det \Sigma_x} - h(x_k) = J(x_k).$$

In what follows, we show that stable LTI systems do not change the non-Gaussianity of asymptotically stationary processes.

**Proposition 2.3** Consider an $m$-input, $m$-output stable LTI system $L(z)$. If the input process $\{x_k\}$ is asymptotically stationary, then the output process $\{y_k\}$ is also asymptotically stationary. Furthermore,

$$J_\infty (y) = J_\infty (x). \quad (2.13)$$

**Proof** It is known from Lemma 2.8 that

$$h_\infty (y) = h_\infty (x) + \frac{1}{2\pi} \int_{-\pi}^\pi \log |\det L(e^{i\omega})| d\omega.$$ 

Hence, since $\det \Phi_y (\omega) = |\det L(e^{i\omega})|^2 \det \Phi_x (\omega)$, we have

$$J_\infty (y) = \frac{1}{2\pi} \int_{-\pi}^\pi \log \sqrt{(2\pi e)^m |\det L(e^{i\omega})|^2 \det \Phi_x (\omega)} d\omega - h_\infty (y)$$

$$= \frac{1}{2\pi} \int_{-\pi}^\pi \log \sqrt{(2\pi e)^m \det \Phi_x (\omega)} d\omega - h_\infty (x) = J_\infty (x).$$
2.5 Gaussianity-Whiteness

Spectral flatness [36] is an important tool useful for describing the shape of power spectral density of an asymptotically stationary process by a single value, which also provides a measure of how white such a process may be.

**Definition 2.17** The spectral flatness (or Wiener entropy) of an asymptotically stationary process \( \{ x_k \} \), \( x_k \in \mathbb{R} \) is defined as

\[
\gamma^2_x = \frac{2}{\pi} \int_{-\pi}^{\pi} \log S_x(\omega) d\omega,
\]

where \( S_x(\omega) \) is the asymptotic power spectrum of \( \{ x_k \} \).

It is known that \( 0 \leq \gamma^2_x \leq 1 \), and that \( \gamma^2_x = 1 \) if and only if \( \{ x_k \} \) is white. Thus, \( \gamma^2_x \) may measure how much an asymptotically stationary process may deviate from a white process.

**Example 2.3** Consider an asymptotically stationary process for which

\[
S_x(\omega) = \begin{cases} a_1, & \text{if } 0 \leq \omega < \omega_0, \\ a_2, & \text{if } \omega_0 \leq \omega \leq \pi. \end{cases}
\]

Then,

\[
\gamma^2_x = \frac{2}{\pi} \int_{0}^{\omega_0} a_1 d\omega + \frac{1}{\pi} \int_{\omega_0}^{\pi} a_2 d\omega = \frac{a_1 \omega_0}{\pi} + \frac{a_2 (\pi - \omega_0)}{\pi} = \frac{a_1}{\pi} a_2 \frac{\omega_0}{a_1} + \frac{\pi - \omega_0}{\pi} a_2.
\]

(2.14)

It can be verified that \( \gamma^2_x \to 0 \) as \( a_2/a_1 \to \infty \) or \( a_1/a_2 \to \infty \). \( \square \)

**Example 2.4** Suppose that

\[
S_x(\omega) = |L(e^{i\omega})|^2,
\]

(2.15)

where \( L(z) \) is a stable transfer function with relative degree \( \nu \). Denote

\[
\rho = \lim_{z \to \infty} [z^{\nu} L(z)].
\]

(2.16)
Then, by invoking Jensen’s formula \([128]\), we obtain
\[
\gamma_x^2 = \frac{\frac{1}{2\pi} \int_{\pi}^{\pi} \log S_x(\omega) \, d\omega}{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) \, d\omega} = \frac{|\rho|^2 \prod_i \max \{1, |\phi_i|^2\}}{\|L(z)\|_2^2},
\]
(2.17)
where \(\phi_i\) denote the zeros of \(L(z)\).

Consider now a memoryless system \(y_k = l x_k, l \neq 0\), where if \(\{x_k\}, x_k \in \mathbb{R}\) is asymptotically stationary, then \(\{y_k\}, y_k \in \mathbb{R}\) is also asymptotically stationary. Moreover, we have \(S_y(\omega) = l^2 S_x(\omega)\), and  
\[
\gamma_y^2 = \gamma_x^2.
\]
(2.18)

In other words, spectral flatness is scale invariant.

The next proposition relates the spectral flatness of the input with that of the output of a stable LTI system.

**Proposition 2.4** Consider an SISO stable LTI system \(L(z)\). If the input \(\{x_k\}, x_k \in \mathbb{R}\) is an asymptotically stationary process, then the output \(\{y_k\}, y_k \in \mathbb{R}\) is also an asymptotically stationary process. Furthermore,
\[
\gamma_y^2 = \frac{\frac{1}{2\pi} \int_{\pi}^{\pi} \log |L(e^{j\omega})|^2 \, d\omega}{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) \, d\omega} \gamma_x^2 \geq \gamma_L^2 \gamma_x^2,
\]
(2.19)
where \(S_x(\omega)\) is the asymptotic power spectrum of \(\{x_k\}\), and \(\gamma_L^2\) is defined as
\[
\gamma_L^2 \triangleq \frac{\frac{1}{2\pi} \int_{\pi}^{\pi} \log |L(e^{j\omega})|^2 \, d\omega}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |L(e^{j\omega})|^2 \, d\omega}.
\]
(2.20)

**Proof** Since
\[
S_y(\omega) = |L(e^{j\omega})|^2 S_x(\omega),
\]
we have

\[
\gamma_y^2 = \frac{2}{\pi} \int_{-\pi}^{\pi} \log S_y(\omega) d\omega = \frac{2}{\pi} \int_{-\pi}^{\pi} \left| L(e^{j\omega}) \right|^2 S_y(\omega) d\omega
\]

\[
\geq \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| L(e^{j\omega}) \right|^2 S_y(\omega) d\omega \right] \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| L(e^{j\omega}) \right|^2 S_x(\omega) d\omega \right]^{1/2}
\]

\[
= \frac{2}{\pi} \int_{-\pi}^{\pi} \log \left| L(e^{j\omega}) \right|^2 d\omega
\]

\[
\geq \frac{1}{2\pi} \int_{-\pi}^{\pi} S_y(\omega) d\omega
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_y(\omega) d\omega
\]

\[
= \gamma_y^2.
\]

\[
\square
\]

Note that $\gamma_L^2$ can be viewed as a flatness measure of the system $L(z)$.

We now generalize the spectral flatness to multivariate processes.

**Definition 2.18** The spectral flatness of an asymptotically stationary multivariate process $\{x_k\}, x_k \in \mathbb{R}^m$ is defined as

\[
\gamma_x^2 \triangleq \frac{2}{\pi} \int_{-\pi}^{\pi} \log \det \Phi_x(\omega) d\omega
\]

\[
\det \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_x(\omega) d\omega \right]
\]

(2.21)

It is easy to show that $0 \leq \gamma_x^2 \leq 1$ using the property that $\log \det (\cdot)$ is concave for positive semidefinite matrices. In addition, $\gamma_x^2 = 1$ if and only if $\{x_k\}$ is white. As such, $\gamma_x^2$ is a measurement of whiteness for multivariate asymptotically stationary processes.

If $\{x_k\}, x_k \in \mathbb{R}^m$ is asymptotically stationary, then $\{y_k\}, y_k = Ax_k, A \in \mathbb{R}^{m \times m}, y_k \in \mathbb{R}^m$ is also asymptotically stationary, where $\det A \neq 0$. Moreover, we have $\Phi_y(\omega) = A \Phi_x(\omega) A^T$. It can then be shown that

\[
\gamma_y^2 = \gamma_x^2.
\]

(2.22)

In other words, the spectral flatness is invariant under linear transformation.

We now introduce a new measure termed Gaussianity-whiteness, which is defined by combining negentropy rate and spectral flatness in a non-trivial way.
**Definition 2.19** Consider an asymptotically stationary process \( \{x_k\} \), \( x_k \in \mathbb{R}^m \) with spectral flatness \( \gamma_x^2 \) and negentropy rate \( J_\infty(x) \). Its Gaussianity-whiteness is defined as

\[
GW_x \triangleq \left[ 2^{-2J_\infty(x)} \right] \gamma_x^2 .
\] (2.23)

Since \( J_\infty(x) \geq 0 \) and \( 0 \leq \gamma_x^2 \leq 1 \), we have \( 0 \leq GW_x \leq 1 \). It is easy to verify that \( GW_x = 1 \) if and only if \( \{x_k\} \) is white Gaussian, i.e., \( J_\infty(x) = 0 \) and \( \gamma_x^2 = 1 \) at the same time. Furthermore, since \( 2^{-2J_\infty(x)} \) is a measure of Gaussianity (\( J_\infty(x) \) is a measure of non-Gaussianity), and \( \gamma_x^2 \) is a measure of whiteness, \( GW_x \) can be well viewed as a measure of Gaussianity and whiteness. This definition of Gaussianity-whiteness shows explicitly how Gaussianity and whiteness are traded off, e.g., how the increase in Gaussianity can be compensated by the decrease in whiteness, and vice versa. Considering the fact that \( J_\infty(x) \) is based on Shannon entropy and \( \gamma_x^2 \) is Wiener entropy, the Gaussianity-whiteness \( GW_x = \left[ 2^{-2J_\infty(x)} \right] \gamma_x^2 \) can be interpreted as a joint Shannon-Wiener entropy. Note that a similar notion was proposed as a generalized spectral flatness measure for non-Gaussian processes in [36].

It is also worth mentioning that there exist more than one way to integrate the Gaussianity measure and the whiteness measure to define a Gaussianity-whiteness measure. The Gaussianity-whiteness \( GW_x \) proposed herein reveals the underlying connection between the asymptotic variance (signal power) and entropy rate power (cf. Definition 2.6) of an asymptotically stationary process. To be more specific, consider an asymptotically stationary process \( \{x_k\} \), \( x_k \in \mathbb{R} \) with asymptotic variance \( \sigma_x^2 \) and entropy rate power \( N_\infty(x) \). Then,

\[
\sigma_x^2 GW_x = N_\infty(x) .
\] (2.24)

In the multivariate case, consider an asymptotically stationary process \( \{x_k\} \), \( x_k \in \mathbb{R}^m \) with asymptotic covariance matrix \( \Sigma_x \) and entropy rate power \( N_\infty(x) \). Then,

\[
(det \Sigma_x) GW_x = N_\infty^m(x) .
\] (2.25)

Since both negentropy rate and spectral flatness are invariant under linear transformation, Gaussianity-whiteness is also invariant under linear transformation.

The next proposition relates the Gaussianity-whiteness of the input with that of the output of a stable LTI system.

**Proposition 2.5** Consider an SISO stable LTI system \( L(z) \). Let \( \{x_k\} \) be its input and \( \{y_k\} \) its output. Suppose that \( \{x_k\} \) is asymptotically stationary. Then,

\[
GW_y = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |L(e^{j\omega})|^2 d\omega GW_x \geq \gamma_L^2 GW_x .
\] (2.26)
In the scalar case, if \{x_k\} is white with variance \(\sigma^2_x\), then \(\gamma^2_x = 1\), and

\[ J_{\infty}(x) = J(x) = \log \sqrt{2\pi e \sigma^2_x} - h(x). \]  (2.27)

Therefore,

\[ GW_x = \left[ 2^{-2J_{\infty}(x)} \right] \gamma^2_x = 2^{-2J(x)} = \frac{2^{2h(x)}}{2\pi e \sigma^2_x} = \frac{N(x)}{\sigma^2_x}. \]  (2.28)

Similarly, the Gaussianity-whiteness of a multivariate white process \{x_k\}, \(x_k \in \mathbb{R}^m\) with covariance \(\Sigma_x\) is found to be

\[ GW_x = \frac{N^m(x)}{\det \Sigma_x}. \]  (2.29)

Notes and References

Information theory originates from Claude Shannon’s classical work [142, 143]. For the introduction to stochastic processes and information theory, one may refer to [34, 115]. A thorough treatment of the basic concepts in information theory, e.g., entropy, mutual information, and channel capacity, can be found in the book [34]. Regarding the channel capacity, we refer to [34] for AWGN and ACGN channels and to [159] for fading channels. More discussions on negentropy and spectral flatness can be found in [36, 68], respectively. Negentropy rate was first proposed in [42], and Gaussianity-whiteness was first introduced in [40].
Towards Integrating Control and Information Theories
From Information-Theoretic Measures to Control Performance Limitations
Fang, S.; Chen, J.; Ishii, H.
2017, XII, 190 p. 30 illus., Hardcover
ISBN: 978-3-319-49288-9