2 An Information-Theoretic Approach to Analog-to-Digital Compression

Alon Kipnis, Yonina C. Eldar, and Andrea J. Goldsmith

Summary

Processing, storing, and communicating information that originates as an analog phenomenon involve conversion of the information to bits. This conversion can be described by the combined effect of sampling and quantization, as illustrated in Fig. 2.1. The digital representation in this procedure is achieved by first sampling the analog signal so as to represent it by a set of discrete-time samples and then quantizing these samples to a finite number of bits. Traditionally, these two operations are considered separately. The sampler is designed to minimize information loss due to sampling on the basis of prior assumptions about the continuous-time input [1]. The quantizer is designed to represent the samples as accurately as possible, subject to the constraint on the number of bits that can be used in the representation [2]. The goal of this chapter is to revisit this paradigm by considering the joint effect of these two operations and to illuminate the dependence between them.

2.1 Introduction

Consider the minimal sampling rate that arises in classical sampling theory due to Whittaker, Kotelnikov, Shannon, and Landau [1, 3, 4]. These works establish the Nyquist rate or the spectral occupancy of the signal as the critical sampling rate above which the signal can be perfectly reconstructed from its samples. This statement, however, focuses only on the condition for perfectly reconstructing a bandlimited signal from its infinite-precision samples; it does not incorporate the quantization precision of the samples and does not apply to signals that are not bandlimited. In fact, as follows from lossy source coding theory, it is impossible to obtain an exact representation of any continuous-amplitude sequence of samples by a digital sequence of numbers due to finite quantization precision, and therefore any digital representation of an analog signal is prone to error. That is, no continuous amplitude signal can be reconstructed from its quantized samples with zero distortion regardless of the sampling rate, even when the signal is bandlimited. This limitation raises the following question. In converting a signal to bits via sampling and quantization at a given bit precision, can the signal be

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

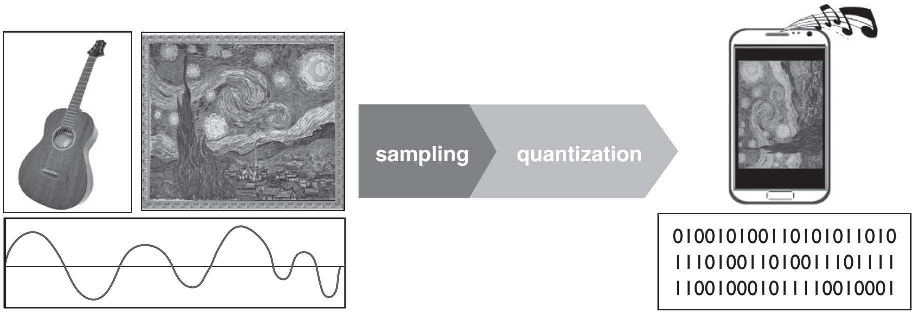

Figure 2.1 Analog-to-digital conversion is achieved by combining sampling and quantization.

reconstructed from these samples with minimal distortion based on sub-Nyquist sampling? One of the goals of this chapter is to discuss this question by extending classical sampling theory to account for quantization and for non-bandlimited inputs. Namely, for an arbitrary stochastic input and given a total bitrate budget, we consider the lowest sampling rate required to sample the signal such that reconstruction of the signal from a bit-constrained representation of its samples results in minimal distortion. As we shall see, without assuming any particular structure for the input analog signal, this sampling rate is often below the signal's Nyquist rate.

The minimal distortion achievable in recovering a signal from its representation by a finite number of bits per unit time depends on the particular way the signal is quantized or, more generally, encoded, into a sequence of bits. Since we are interested in the fundamental distortion limit in recovering an analog signal from its digital representation, we consider all possible encoding and reconstruction (decoding) techniques. As an example, in Fig. 2.1 the smartphone display may be viewed as a reconstruction of the real-world painting *The Starry Night* from its digital representation. No matter how fine the smartphone screen, this recovery is not perfect since the digital representation of the analog image is not accurate. That is, loss of information occurs during the transformation from analog to digital. Our goal is to analyze this loss as a function of hardware limitations on the sampling mechanism and the number of bits used in the encoding. It is convenient to normalize this number of bits by the signal's free dimensions, that is, the dimensions along which new information is generated. For example, the free dimensions of a visual signal are usually the horizontal and vertical axes of the frame, and the free dimension of an audio wave is time. For simplicity, we consider analog signals with a single free dimension, and we denote this dimension as *time*. Therefore, our restriction on the digital representation is given in terms of its *bitrate* – the number of bits per unit time.

For an arbitrary continuous-time random signal with a known distribution, the fundamental distortion limit due to the encoding of the signal using a limited bitrate is given by Shannon's distortion-rate function (DRF) [5–7]. This function provides the optimal trade-off between the bitrate of the signal's digital representation and the distortion in recovering the original signal from this representation. Shannon's DRF is described only in terms of the distortion criterion, the probability distribution of the continuous-time signal, and the maximal bitrate allowed in the digital representation. Consequently, the

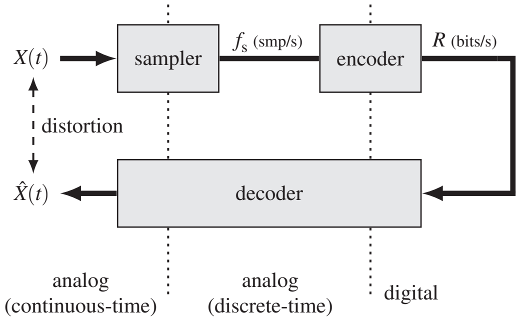

Figure 2.2 Analog-to-digital compression (ADX) and reconstruction setting. Our goal is to derive the minimal distortion between the signal and its reconstruction from any encoding at bitrate R of the samples of the signal taken at sampling rate f_s.

optimal encoding scheme that attains Shannon's DRF is a general mapping from the space of continuous-time signals to bits that does not consider any practical constraints in implementing such a mapping. In practice, the minimal distortion in recovering analog signals from their mapping to bits considers the digital encoding of the signal *samples*, with a constraint on both the *sampling rate* and the *bitrate* of the system [8–10]. Here the sampling rate f_s is defined as the number of samples per unit time of the continuous-time source signal; the bitrate R is the number of bits per unit time used in the representation of these samples. The resulting system describing our problem is illustrated in Fig. 2.2, and is denoted as the *analog-to-digital compression* (ADX) setting.

The digital representation in this setting is obtained by transforming a continuous-time continuous-amplitude random source signal $X(\cdot)$ through the concatenated operation of a *sampler* and an *encoder*, resulting in a bit sequence. The *decoder* estimates the original analog signal from this bit sequence. The *distortion* is defined to be the mean-squared error (MSE) between the input signal $X(\cdot)$ and its reconstruction $\widehat{X}(\cdot)$. Since we are interested in the fundamental distortion limit subject to a sampling constraint, we allow optimization over the encoder and the decoder as the time interval over which $X(\cdot)$ is sampled goes to infinity. When $X(\cdot)$ is bandlimited and the sampling rate f_s exceeds its Nyquist rate f_{Nyq}, the encoder can recover the signal using standard interpolation and use the optimal source code at bitrate R to attain distortion equal to Shannon's DRF of the signal [11]. Therefore, for bandlimited signals, a non-trivial interplay between the sampling rate and the bitrate arises only when f_s is below a signal's Nyquist rate. In addition to the optimal encoder and decoder, we also explore the optimal sampling mechanism, but limit ourselves to the class of linear and continuous deterministic samplers. Namely, each sample is defined by a bounded linear functional over a class of signals. Finally, in order to account for system imperfections or those due to external interferences, we assume that the signal $X(\cdot)$ is corrupted by additive noise prior to sampling. The noise-free version is obtained from our results by setting the intensity of this noise to zero.

The minimal distortion in the ADX system of Fig. 2.2 is bounded from below by two extreme cases of the sampling rate and the bitrate, as illustrated in Fig. 2.3: (1) when the bitrate R is unlimited, the minimal ADX distortion reduces to the minimal

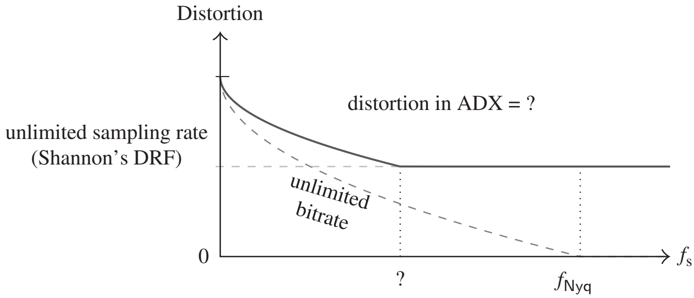

Figure 2.3 The minimal sampling rate for attaining the minimal distortion achievable under a bitrate-limited representation is usually below the Nyquist rate f_{Nyq}. In this figure, the noise is assumed to be zero.

MSE (MMSE) in interpolating a signal from its noisy samples at rate f_s [12, 13]. (2) When the sampling rate f_s is unlimited or above the Nyquist rate of the signal and when the noise is zero, the ADX distortion reduces to Shannon's DRF of the signal. Indeed, in this case, the optimal encoder can recover the original continuous-time source without distortion and then encode this recovery in an optimal manner according to the optimal lossy compression scheme attaining Shannon's DRF. When f_s is unlimited or above the Nyquist rate and the noise is not zero, the minimal distortion is the *indirect* (or *remote*) DRF of the signal given its noise-corrupted version, see Section 3.5 of [7] and [14]. Our goal is therefore to characterize the MSE due to the joint effect of a finite bitrate constraint and sampling at a sub-Nyquist sampling rate. In particular, we are interested in the minimal sampling rate for which Shannon's DRF, or the indirect DRF, describing the minimal distortion subject to a bitrate constraint, is attained. As illustrated in Fig. 2.3, and as will be explained in more detail below, this sampling rate is usually below the Nyquist rate of $X(\cdot)$, or, more generally, the spectral occupancy of $X(\cdot)$ when non-uniform or generalized sampling techniques are allowed. We denote this minimal sampling rate as the *critical sampling rate* subject to a bitrate constraint, since it describes the minimal sampling rate required to attain the optimal performance in systems operating under quantization or bitrate restrictions. The critical sampling rate extends the minimal-distortion sampling rate considered by Shannon, Nyquist, and Landau. It is only as the bitrate goes to infinity that sampling at the Nyquist rate is necessary to attain minimal (namely zero) distortion.

In order to gain intuition as to why the minimal distortion under a bitrate constraint may be attained by sampling below the Nyquist rate, we first consider in Section 2.2 a simpler version of the ADX setup involving the lossy compression of linear projections of signals represented as finite-dimensional random real vectors. Next, in Section 2.3 we formalize the combined sampling and source coding problem arising from Fig. 2.2 and provide basic properties of the minimal distortion in this setting. In Section 2.4, we fully characterize the minimal distortion in ADX as a function of the bitrate and sampling rate and derive the critical sampling rate that leads to optimal performance. We conclude this chapter in Section 2.5, where we consider uniform samplers, in particular single-branch and more general multi-branch uniform samplers, and show that these samplers attain the fundamental distortion limit.

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

2.2 Lossy Compression of Finite-Dimensional Signals

Let X^n be an n-dimensional Gaussian random vector with covariance matrix Σ_{X^n}, and let Y^m be a projected version of X^n defined by

$$Y^m = \mathbf{H}X^n, \tag{2.1}$$

where $\mathbf{H} \in \mathbb{R}^{m\times n}$ is a deterministic matrix and $m < n$. This projection of X^n into a lower-dimensional space is the counterpart of sampling the continuous-time analog signal $X(\cdot)$ in the ADX setting. We consider the normalized MMSE estimate of X^n from a representation of Y^m using a limited number of bits.

Without constraining the number of bits, the distortion in this estimation is given by

$$\mathsf{mmse}(X^n|Y^m) \triangleq \frac{1}{n}\operatorname{tr}(\Sigma_{X^n} - \Sigma_{X^n|Y^m}), \tag{2.2}$$

where $\Sigma_{X^n|Y^m}$ is the conditional covariance matrix. However, when Y^m is to be encoded using a code of no more than nR bits, the minimal distortion cannot be smaller than the indirect DRF of X^n given Y^m, denoted by $D_{X^n|Y^m}(R)$. This function is given by the following parametric expression [14]:

$$\begin{aligned} D(R_\theta) &= \operatorname{tr}(\Sigma_{X^n}) - \sum_{i=1}^{m}[\lambda_i(\Sigma_{X^n|Y^m}) - \theta]^+, \\ R_\theta &= \frac{1}{2}\sum_{i=1}^{m}\log^+[\lambda_i(\Sigma_{X^n|Y^m})/\theta], \end{aligned} \tag{2.3}$$

where $x^+ = \max\{x, 0\}$ and $\lambda_i(\Sigma_{X^n|Y^m})$ is the ith eigenvalue of $\Sigma_{X^n|Y^m}$.

It follows from (2.2) that X^n can be recovered from Y^m with zero MMSE if and only if

$$\lambda_i(\Sigma_{X^n}) = \lambda_i(\Sigma_{X^n|Y^m}), \tag{2.4}$$

for all $i = 1, \ldots, n$. When this condition is satisfied, (2.3) takes the form

$$\begin{aligned} D(R_\theta) &= \sum_{i=1}^{n}\min\{\lambda_i(\Sigma_{X^n}), \theta\}, \\ R_\theta &= \frac{1}{2}\sum_{i=1}^{n}\log^+[\lambda_i(\Sigma_{X^n})/\theta], \end{aligned} \tag{2.5}$$

which is Kolmogorov's reverse water-filling expression for the DRF of the vector Gaussian source X^n [15], i.e., the minimal distortion in encoding X^n using codes of rate R bits per source realization. The key insight is that the requirements for equality between (2.3) and (2.5) are not as strict as (2.4): all that is needed is equality among those eigenvalues that affect the value of (2.5). In particular, assume that for a point (R, D) on $D_{X^n}(R)$, only $\lambda_n(\Sigma_{X^n}), \ldots, \lambda_{n-m+1}(\Sigma_{X^n})$ are larger than θ, where the eigenvalues are organized in ascending order. Then we can choose the rows of $\mathbf{H}$ to be the m left eigenvectors corresponding to $\lambda_n(\Sigma_{X^n}), \ldots, \lambda_{n-m+1}(\Sigma_{X^n})$. With this choice of $\mathbf{H}$, the m largest eigenvalues of $\Sigma_{X^n|Y^m}$ are identical to the m largest eigenvalues of Σ_{X^n}, and (2.5) is equal to (2.3).

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

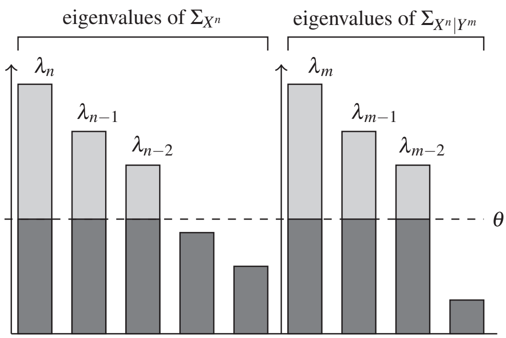

Figure 2.4 Optimal sampling occurs whenever $D_{X^n}(R) = D_{X^n|Y^m}(R)$. This condition is satisfied even for $m < n$, as long as there is equality among the eigenvalues of Σ_{X^n} and $\Sigma_{X^n|Y^m}$ which are larger than the water-level parameter θ.

Since the rank of the sampling matrix is now $m < n$, we effectively performed sampling below the "Nyquist rate" of X^n without degrading the performance dictated by its DRF. One way to understand this phenomenon is as an alignment between the range of the sampling matrix $\mathbf{H}$ and the subspace over which X^n is represented, according to Kolmogorov's expression (2.5). That is, when Kolmogorov's expression implies that not all degrees of freedom are utilized by the optimal distortion-rate code, sub-sampling does not incur further performance loss, provided that the sampling matrix is aligned with the optimal code. This situation is illustrated in Fig. 2.4. Sampling with an $\mathbf{H}$ that has fewer rows than the rank of Σ_{X^2} is the finite-dimensional analog of sub-Nyquist sampling in the infinite-dimensional setting of continuous-time signals.

In the rest of this chapter, we explore the counterpart of the phenomena described above in the richer setting of continuous-time stationary processes that may or may not be bandlimited, and whose samples may be corrupted by additive noise.

2.3 ADX for Continuous-Time Analog Signals

We now explore the fundamental ADX distortion in the setting of continuous-time stationary processes that may be corrupted by noise prior to sampling. We consider the system of Fig. 2.5 in which $X(\cdot) \triangleq \{X(t), t \in \mathbb{R}\}$ is a stationary process with power

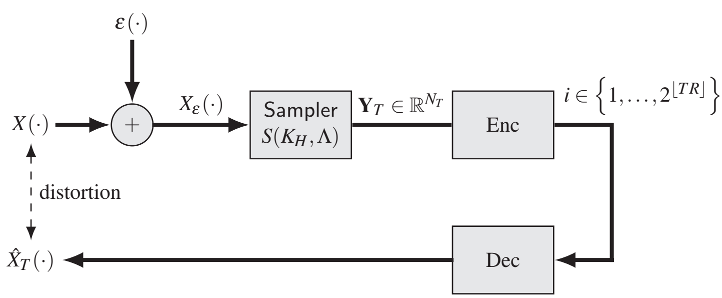

Figure 2.5 Combined sampling and source coding setting.

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

spectral density (PSD) $S_X(f)$. This PSD is assumed to be a real, symmetric, and absolute integrable function that satisfies

$$\mathbb{E}[X(t)X(s)] = \int_{-\infty}^{\infty} S_X(f) e^{2\pi j(t-s)f} df, \quad t, s \in \mathbb{R}. \tag{2.6}$$

The noise process $\varepsilon(\cdot)$ is another stationary process independent of $X(\cdot)$ with PSD $S_\varepsilon(f)$ of similar properties, so that the input to the sampler is the stationary process $X_\varepsilon(\cdot) \triangleq X(\cdot) + \varepsilon(\cdot)$ with PSD $S_{X_\varepsilon}(f) = S_X(f) + S_\varepsilon(f)$.

We note that, by construction, $X(\cdot)$ and $\varepsilon(\cdot)$ are regular processes in the sense that their spectral measure has an absolutely continuous density with respect to the Lebesgue measure. If in addition the support of $S_X(f)$, denoted by supp S_X, is contained[1] within a bounded interval, then we say that $X(\cdot)$ is bandlimited and denote by f_{Nyq} its Nyquist rate, defined as twice the maximal element in supp S_X. The *spectral occupancy* of $X(\cdot)$ is defined to be the Lebesgue measure of supp S_X.

Although this is not necessary for all parts of our discussion, we assume that the processes $X(\cdot)$ and $\varepsilon(\cdot)$ are Gaussian. This assumption leads to closed-form characterizations for many of the expressions we consider. In addition, it follows from [16, 17] that a lossy compression policy that is optimal under a Gaussian distribution can be used to encode non-Gaussian signals with matching second-order statistics, while attaining the same distortion as if the signals were Gaussian. Hence, the optimal sampler and encoding system we use to obtain the fundamental distortion limit for Gaussian signals attains the same distortion limit for non-Gaussian signals as long as the second-order statistics of the two signals are the same.

2.3.1 Bounded Linear Sampling

The sampler in Fig. 2.5 outputs a finite-dimensional vector of samples where, most generally, each sample is defined by a linear and bounded (hence, continuous) functional of the process $X_\varepsilon(\cdot)$. For this reason, we denote a sampler of this type as a *bounded linear sampler*. In order to consider this sampler in applications, it is most convenient to define it in terms of a bilinear kernel $K_H(t,s)$ on $\mathbb{R} \times \mathbb{R}$ and a discrete *sampling set* $\Lambda \subset \mathbb{R}$, as illustrated in Fig. 2.6. The kernel $K_H(t,s)$ defines a time-varying linear system on a suitable class of signals [18], and hence each element $t_n \in \Lambda$ defines a linear bounded functional $K(t_n, s)$ on this class by

$$Y_n \triangleq \int_{-\infty}^{\infty} X_\varepsilon(s) K_H(t_n, s) ds.$$

For a time horizon T, we denote by $\mathbf{Y}_T$ the finite-dimensional vector obtained by sampling at times $t_1, \ldots, t_n \in \Lambda_T$, where

$$\Lambda_T \triangleq \Lambda \cap [-T/2, T/2].$$

[1] Since the PSD is associated with an absolutely continuous spectral measure, sets defined in term of the PSD, e.g., supp S_X, are understood to be unique up to symmetric difference of Lebesgue measure zero.

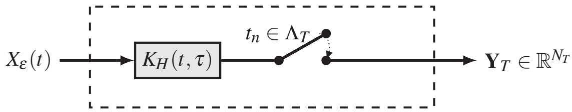

Figure 2.6 Bounded linear sampler ($N_T \triangleq |\Lambda_T|$).

We assume in addition that Λ is uniformly discrete in the sense that there exists $\varepsilon > 0$ such that $|t - s| > \varepsilon$ for every non-identical $t, s \in \Lambda$. The density of Λ_T is defined as the number of points in Λ_T divided by T and denoted here by $d(\Lambda_T)$. Whenever it exists, we define the limit

$$d(\Lambda) = \lim_{T\to\infty} d(\Lambda_T) = \lim_{T\to\infty} \frac{|\Lambda \cap [-T/2, T/2]|}{T}$$

as the *symmetric density* of Λ, or simply it's density.

Linear Time-Invariant Uniform Sampling

An important special case of the bounded linear sampler is that of a linear time-invariant (LTI) uniform sampler [1], illustrated in Fig. 2.7. For this sampler, the sampling set is a uniform grid $\mathbb{Z}T_s = \{nT_s, n \in \mathbb{Z}\}$, where $T_s = f_s^{-1} > 0$. The kernel is of the form $K_H(t,s) = h(t-s)$ where $h(t)$ is the impulse response of an LTI system with frequency response $H(f)$. Therefore, the entries of $\mathbf{Y}_T$ corresponding to sampling at times $nT_s \in \Lambda$ are given by

$$Y_n \triangleq \int_{-\infty}^{\infty} h(nT_s - s) X_\varepsilon(s) ds.$$

It is easy to check that $d(T_s\mathbb{Z}) = f_s$ and hence, in this case, the density of the sampling set has the usual interpretation of sampling rate.

Multi-Branch Linear Time-Invariant Uniform Sampling

A generalization of the uniform LTI sampler incorporates several of these samplers in parallel, as illustrated in Fig. 2.8. Each of the L branches in Fig. 2.8 consists of a LTI system with frequency response $H_l(f)$ followed by a uniform pointwise sampler with sampling rate f_s/L, so that the overall sampling rate is f_s. The vector $\mathbf{Y}_T$ consists of the concatenation of the vectors $Y_{1,T}, \ldots, Y_{L,T}$ obtained from each of the sampling branches.

2.3.2 Encoding and Reconstruction

For a time horizon T, the encoder in Fig. 2.5 can be any function of the form

$$f : \mathbb{R}^{N_T} \to \left\{1, \ldots, 2^{\lfloor TR \rfloor}\right\}, \tag{2.7}$$

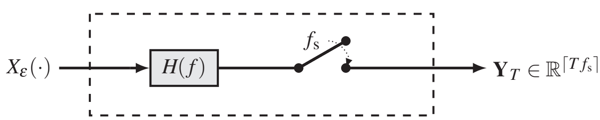

Figure 2.7 Uniform linear time-invariant sampler.

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

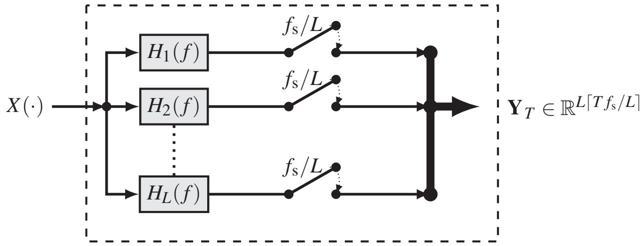

Figure 2.8 Multi-branch linear time-invariant uniform sampler.

where $N_T = \dim(\mathbf{Y}_T) = |\Lambda_T|$. That is, the encoder receives the vector of samples $\mathbf{Y}_T$ and outputs an index out of $2^{\lfloor TR \rfloor}$ possible indices. The decoder receives this index, and produces an estimate $\widehat{X}(\cdot)$ for the signal $X(\cdot)$ over the interval $[-T/2, T/2]$. Thus, it is a mapping

$$g : \left\{1, \ldots, 2^{\lfloor TR \rfloor}\right\} \to \mathbb{R}^{[-T/2, T/2]}. \tag{2.8}$$

The goal of the joint operation of the encoder and the decoder is to minimize the expected mean-squared error (MSE)

$$\frac{1}{T} \int_{-T/2}^{T/2} \mathbb{E}\left(X(t) - \widehat{X}(t)\right)^2 dt.$$

In practice, an encoder may output a finite number of samples that are then interpolated to the continuous-time estimate $\widehat{X}(\cdot)$. Since our goal is to understand the limits in converting signals to bits, this separation between decoding and interpolation, as well as the possible restrictions each of these steps encounters in practice, are not explored within the context of ADX.

Given a particular bounded linear sampler $S = (\Lambda, K_H)$ and a bitrate R, we are interested in characterizing the function

$$D_T(S, R) \triangleq \inf_{f,g} \frac{1}{T} \int_{-T/2}^{T/2} \mathbb{E}\left(X(t) - \widehat{X}(t)\right)^2 dt, \tag{2.9}$$

or its limit as $T \to \infty$, where the infimum is over all encoders and decoders of the form (2.7) and (2.8). The function $D_T(S, R)$ is defined only in terms of the sampler S and the bitrate R, and in this sense measures the minimal distortion that can be attained using the sampler S subject to a bitrate constraint R on the representation of the samples.

2.3.3 Optimal Distortion in ADX

From the definition of $D_T(S, R)$ and the ADX setting, it immediately follows that $D_T(S, R)$ is non-increasing in R. Indeed, any encoding into a set of $2^{\lfloor TR \rfloor}$ elements can be obtained as a special case of encoding to a set of $2^{\lfloor T(R+r) \rfloor}$ elements, for $r > 0$. In addition, by using the trivial encoder $g \equiv 0$ we see that $D_T(S, R)$ is bounded from above by the variance σ_X^2 of $X(\cdot)$, which is given by

$$\sigma_X^2 \triangleq \int_{-\infty}^{\infty} S_X(f) df.$$

In what follows, we explore additional important properties of $D_T(S, R)$.

Optimal Encoding

Denote by $\widetilde{X}_T(\cdot)$ the process that is obtained by estimating $X(\cdot)$ from the output of the sampler according to an MSE criterion. That is

$$\widetilde{X}_T(t) \triangleq \mathbb{E}[X(t)|\mathbf{Y}_T], \quad t \in \mathbb{R}. \tag{2.10}$$

From properties of the conditional expectation and MSE, under any encoder f we may write

$$\frac{1}{T}\int_{-T/2}^{T/2} \mathbb{E}\Big(X(t) - \widehat{X}(t)\Big)^2 dt = \mathsf{mmse}_T(S) + \mathsf{mmse}\Big(\widetilde{X}_T | f(\mathbf{Y}_T)\Big), \tag{2.11}$$

where

$$\mathsf{mmse}_T(S) \triangleq \frac{1}{T}\int_{-T/2}^{T/2} \mathbb{E}\Big(X(t) - \widetilde{X}_T(t)\Big)^2 dt \tag{2.12}$$

is the distortion due to sampling and

$$\mathsf{mmse}\Big(\widetilde{X}_T | f(\mathbf{Y}_T)\Big) \triangleq \frac{1}{T}\int_{-T/2}^{T/2} \mathbb{E}\Big(\widetilde{X}_T(t) - g(f(\mathbf{Y}_T))\Big)^2 dt$$

is the distortion associated with the lossy compression procedure, and depends on the sampler only through $\widetilde{X}_T(\cdot)$.

The decomposition (2.11) already provides important clues on an optimal encoder and decoder pair that attains $D_T(S,R)$. Specifically, it follows from (2.11) that there is no loss in performance if the encoder tries to describe the process $\widetilde{X}_T(\cdot)$ subject to the bitrate constraint, rather than the process $X(\cdot)$. Consequently, the optimal decoder outputs the conditional expectation of $\widetilde{X}_T(\cdot)$ given $f(\mathbf{Y}_T)$. The decomposition (2.11) was first used in [14] to derive the indirect DRF of a pair of stationary Gaussian processes, and later in [19] to derive indirect DRF expressions in other settings. An extension of the principle presented in this decomposition to arbitrary distortion measures is discussed in [20].

The decomposition (2.11) also sheds light on the behavior of the optimal distortion $D_T(S,R)$ under the two extreme cases of unlimited bitrate and unrestricted sampling rate, each of which is illustrated in Fig. 2.3. We discuss these two cases next.

Unlimited Bitrate

If we remove the bitrate constraint in the ADX setting (formally, letting $R \to \infty$), loss of information is only due to noise and sampling. In this case, the second term in the RHS of (2.11) disappears, and the distortion in ADX is given by $\mathsf{mmse}_T(S)$. Namely, we have

$$\lim_{R\to\infty} D_T(S,R) = \mathsf{mmse}_T(S).$$

The unlimited bitrate setting reduces the ADX problem to a classical problem in sampling theory: the MSE under a given sampling system. Of particular interest is the case of optimal sampling, i.e., when this MSE vanishes as $T \to \infty$. For example, by considering the noiseless case and assuming that $K_H(t,s) = \delta(t-s)$ is the identity operator, the sampler is defined solely in terms of Λ. The condition on $\mathsf{mmse}_T(S)$ to converge

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

to zero is related to the conditions for stable sampling in Paley–Wiener spaces studied by Landau and Beurling [21, 22]. In order to see this connection more precisely, note that (2.6) defines an isomorphism between the Hilbert spaces of finite-variance random variables measurable with respect to the sigma algebra generated by $X(\cdot)$ and the Hilbert space generated by the inverse Fourier transform of $\left\{e^{2\pi jtf}\sqrt{S_X(f)},\, t \in \mathbb{R}\right\}$ [23]. Specifically, this isomorphism is obtained by extending the map

$$X(t) \longleftrightarrow \mathcal{F}^{-1}\left\{e^{2\pi jtf}\sqrt{S_X(f)}\right\}(s)$$

to the two aforementioned spaces. It follows that sampling and reconstructing $X(\cdot)$ with vanishing MSE is equivalent to the same operation in the Paley–Wiener space of analytic functions whose Fourier transform vanishes outside supp S_X. In particular, the condition $\mathsf{mmse}_T(S) \overset{T\to\infty}{\longrightarrow} 0$ holds whenever Λ is a set of stable sampling in this Paley–Wiener space, i.e., there exists a universal constant $A > 0$ such that the L_2 norm of each function in this space is bounded by A times the energy of the samples of this function. Landau [21] showed that a necessary condition for this property is that the number of points in Λ that fall within the interval $[-T/2, T/2]$ is at least the spectral occupancy of $X(\cdot)$ times T, minus a constant that is logarithmic in T. For this reason, this spectral occupancy is termed the *Landau rate* of $X(\cdot)$, and we denote it here by f_{Lnd}. In the special case where supp S_X is an interval (symmetric around the origin since $X(\cdot)$ is real), the Landau and Nyquist rates coincide.

Optimal Sampling

The other special case of the ADX setting is obtained when there is no loss of information due to sampling. For example, this is the case when $\mathsf{mmse}_T(S)$ goes to zero under the conditions mentioned above of zero noise, identity kernel, and sampling density exceeding the spectral occupancy. More generally, this situation occurs whenever $\widetilde{X}_T(\cdot)$ converges (in expected norm) to the MMSE estimator of $X(\cdot)$ from $X_\varepsilon(\cdot)$. This MMSE estimator is a stationary process obtained by non-causal Wiener filtering, and its PSD is

$$S_{X|X_\varepsilon}(f) \triangleq \frac{S_X^2(f)}{S_X(f) + S_\varepsilon(f)}, \tag{2.13}$$

where in (2.13) and in similar expressions henceforth we interpret the expression to be zero whenever both the numerator and denominator are zero. The resulting MMSE is given by

$$\mathsf{mmse}(X|X_\varepsilon) \triangleq \int_{-\infty}^{\infty} [S_X(f) - S_{X|X_\varepsilon}(f)]df. \tag{2.14}$$

Since our setting does not limit the encoder from computing $\widetilde{X}_T(\cdot)$, the ADX problem reduces in this case to the indirect source coding problem of recovering $X(\cdot)$ from a bitrate R representation of its corrupted version $X_\varepsilon(\cdot)$. This problem was considered and

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

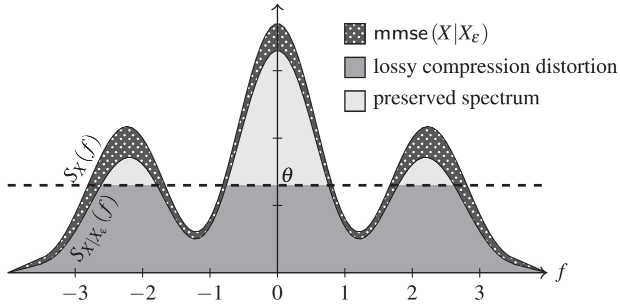

Figure 2.9 Water-filling interpretation of (2.15). The distortion is the sum of $\mathsf{mmse}(X|X_\varepsilon)$ and the lossy compression distortion.

solved by Dobrushin and Tsybakov in [14], where the following expression was given for the optimal trade-off between bitrate and distortion:

$$D_{X|X_\varepsilon}(R_\theta) \triangleq \mathsf{mmse}(X|X_\varepsilon) + \int_{-\infty}^{\infty} \min\{S_{X|X_\varepsilon}(f), \theta\} df, \tag{2.15a}$$

$$R_\theta = \frac{1}{2}\int_{-\infty}^{\infty} \log^+[S_{X|X_\varepsilon}(f)/\theta] df. \tag{2.15b}$$

A graphical water-filling interpretation of (2.15) is given in Fig. 2.9. When the noise $\varepsilon(\cdot)$ is zero, $S_{X|X_\varepsilon}(f) = S_X(f)$, and hence (2.15) reduces to

$$D_X(R_\theta) \triangleq \int_{-\infty}^{\infty} \min\{S_X(f), \theta\} df, \tag{2.16a}$$

$$R_\theta = \frac{1}{2}\int_{-\infty}^{\infty} \log^+[S_X(f)/\theta] df, \tag{2.16b}$$

which is Pinsker's expression [15] for the DRF of the process $X(\cdot)$, denoted here by $D_X(R)$. Note that (2.16) is the continuous-time counterpart of (2.3).

From the discussion above, we conclude that

$$D_T(S,R) \geq D_{X|X_\varepsilon}(R) \geq \max\{D_X(R), \mathsf{mmse}(X|X_\varepsilon)\}. \tag{2.17}$$

Furthermore, when the estimator $\mathbb{E}[X(t)|X_\varepsilon]$ can be obtained from $\mathbf{Y}_T$ as $T \to \infty$, we have that $D_T(S,R) \xrightarrow{T\to\infty} D_{X|X_\varepsilon}(R)$. In this situation, we say that the conditions for optimal sampling are met, since the only distortion is due to the noise and the bitrate constraint.

The two lower bounds in Fig. 2.3 describe the behavior of $D_T(S,R)$ in the two special cases of unrestricted bitrate and optimal sampling. Our goal in the next section is to characterize the intermediate case of non-optimal sampling and a finite bitrate constraint.

2.4 The Fundamental Distortion Limit

Given a particular bounded linear sampler $S = (\Lambda, K_H)$ and a bitrate R, we defined the function $D_T(S,R)$ as the minimal distortion that can be attained in the combined sampling and lossy compression setup of Fig. 2.5. Our goal in this section is to derive and analyze a function $D^\star(f_s, R)$ that bounds from below $D_T(S,R)$ for any such bounded

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

linear sampler with symmetric density of Λ not exceeding f_s. The achievability of this lower bound is addressed in the next section.

2.4.1 Definition of $D^\star(f_s,R)$

In order to define $D^\star(f_s,R)$, we let $F^\star(f_s)\subset\mathbb{R}$ be any set that maximizes

$$\int_F S_{X|X_\varepsilon}(f)df = \int_F \frac{S_X^2(f)}{S_X(f)+S_\varepsilon(f)}\,df \tag{2.18}$$

over all Lebesgue measurable sets F whose Lebesgue measure does not exceed f_s. In other words, $F^\star(f_s)$ consists of the f_s spectral bands with the highest energy in the spectrum of the process $\{\mathbb{E}[X(t)|X_\varepsilon(\cdot)],\ t\in\mathbb{R}\}$. Define

$$D^\star(f_s,R_\theta) = \mathsf{mmse}^\star(f_s) + \int_{F^\star(f_s)} \min\{S_{X|X_\varepsilon}(f),\theta\}df, \tag{2.19a}$$

$$R_\theta = \frac{1}{2}\int_{F^\star(f_s)} \log^+[S_{X|X_\varepsilon}(f)/\theta]df, \tag{2.19b}$$

where

$$\mathsf{mmse}^\star(f_s) \triangleq \sigma_X^2 - \int_{F^\star(f_s)} S_{X|X_\varepsilon}(f)df = \int_{-\infty}^{\infty}\left[S_X(f) - S_{X|X_\varepsilon}(f)\mathbf{1}_{F^\star(f_s)}\right]df.$$

Graphical interpretations of $D^\star(f_s,R)$ and $\mathsf{mmse}^\star(f_s)$ are provided in Fig. 2.10.

The importance of the function $D^\star(f_s,R)$ can be deduced from the following two theorems:

THEOREM 2.1 (converse) *Let $X(\cdot)$ be a Gaussian stationary process corrupted by a Gaussian stationary noise $\varepsilon(\cdot)$, and sampled using a bounded linear sampler $S = (K_H,\Lambda)$.*

(i) *Assume that for any $T>0$, $d(\Lambda_T)\le f_s$. Then, for any bitrate R,*

$$D_T(S,R) \ge D^\star(f_s,R).$$

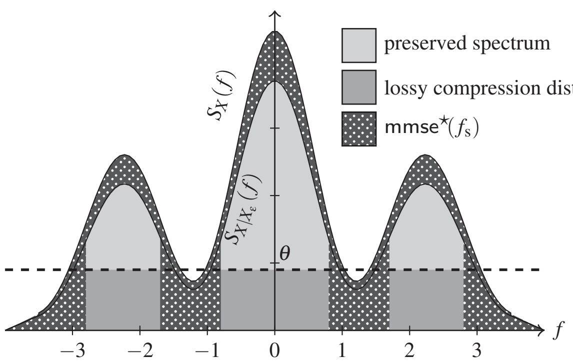

Figure 2.10 Water-filling interpretation of $D^\star(f_s,R_\theta)$: the fundamental distortion limit under any bounded linear sampling. This distortion is the sum of the fundamental estimation error $\mathsf{mmse}^\star(f_s)$ and the lossy compression distortion.

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

(ii) *Assume that the symmetric density of Λ exists and satisfies $d(\Lambda) \leq f_s$. Then, for any bitrate R,*

$$\liminf_{T\to\infty} D_T(S,R) \geq D^\star(f_s,R).$$

In addition to the negative statement of Theorem 2.1, we show in the next section the following positive coding result.

THEOREM 2.2 (achievability) *Let $X(\cdot)$ be a Gaussian stationary process corrupted by a Gaussian stationary noise $\varepsilon(\cdot)$. Then, for any f_s and $\varepsilon > 0$, there exists a bounded linear sampler S with a sampling set of symmetric density not exceeding f_s such that, for any R, the distortion in ADX attained by sampling $X_\varepsilon(\cdot)$ using S over a large enough time interval T, and encoding these samples using $\lfloor TR \rfloor$ bits, does not exceed $D^\star(f_s,R)+\varepsilon$.*

A full proof of Theorem 2.1 can be found in [24]. Intuition for Theorem 2.1 may be obtained by representing $X(\cdot)$ according to its Karhunen–Loève (KL) expansion over $[-T/2,T/2]$, and then using a sampling matrix that keeps only $N_T \triangleq \lfloor T f_s \rfloor$ of these coefficients. The function $D^\star(f_s,R)$ arises as the limiting expression in the noisy version of (2.5), when the sampling matrix is tuned to keep those KL coefficients corresponding to the N_T largest eigenvalues in the expansion.

In Section 2.5, we provide a constructive proof of Theorem 2.2 that also shows that $D^\star(f_s,R)$ is attained using a multi-branch LTI uniform sampler with an appropriate choice of pre-sampling filters. The rest of the current section is devoted to studying properties of the minimal ADX distortion $D^\star(f_s,R)$.

2.4.2 Properties of $D^\star(f_s,R)$

In view of Theorems 2.1 and 2.2, the function $D^\star(f_s,R)$ trivially satisfies the properties mentioned in Section 2.3.3 for the optimal distortion in ADX. It is instructive to observe how these properties can be deduced directly from the definition of $D^\star(f_s,R)$ in (2.19).

Unlimited Bitrate

As $R \to \infty$, the parameter θ goes to zero and (2.19a) reduces to $\mathsf{mmse}^\star(f_s)$. This function describes the MMSE that can be attained by any bounded linear sampler with symmetric density at most f_s. In particular, in the non-noisy case, $\mathsf{mmse}^\star(f_s) = 0$ if and only if f_s exceeds the Landau rate of $X(\cdot)$. Therefore, in view of the explanation in Section 2.3.3 and under unlimited bitrate, zero noise, and the identity pre-sampling operation, Theorem 2.1 agrees with the necessary condition derived by Landau for stable sampling in the Paley–Wiener space [21].

Optimal Sampling

The other extreme in the expression for $D^\star(f_s,R)$ is when f_s is large enough that it does not impose any constraint on sampling. In this case, we expect the ADX distortion to coincide with the function $D_{X|X_\epsilon}(R)$ of (2.15), since the latter is the minimal distortion only due to noise and lossy compression at bitrate R. From the definition of $F^\star(f_s)$, we

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

observe that $F^\star(f_s) = \operatorname{supp} S_X$ (almost everywhere) whenever f_s is equal to or greater than the Landau rate of $X(\cdot)$. By examining (2.19), we see that this equality implies that

$$D^\star(f_s, R) = D_{X|X_\varepsilon}(R). \tag{2.20}$$

In other words, the condition $f_s \geq f_{\mathsf{Lnd}}$ means that there is no loss due to sampling in the ADX system. This property of the minimal distortion is not surprising. It merely expresses the fact anticipated in Section 2.3.3 that, when (2.10) vanishes as T goes to infinity, the estimator $\mathbb{E}[X(t)|X_\varepsilon]$ is obtained from the samples in this limit and thus the only loss of information after sampling is due to the noise.

In the next section we will see that, under some conditions, the equality (2.20) is extended to sampling rates smaller than the Landau rate of the signal.

2.4.3 Optimal Sampling Subject to a Bitrate Constraint

We now explore the minimal sampling rate f_s required in order to attain equality in (2.20), that is, the minimal sampling rate at which the minimal distortion in ADX equals the indirect DRF of $X(\cdot)$ given $X_\varepsilon(\cdot)$, describing the minimal distortion subject only to a bitrate R constraint and additive noise. Intuition for this sampling rate is obtained by exploring the behavior of $D^\star(f_s, R)$ as a function of f_s for a specific PSD and a fixed bitrate R. For simplicity, we explore this behavior under the assumption of zero noise ($\varepsilon \equiv 0$) and signal $X(\cdot)$ with a unimodal PSD as in Fig. 2.11. Note that in this case we

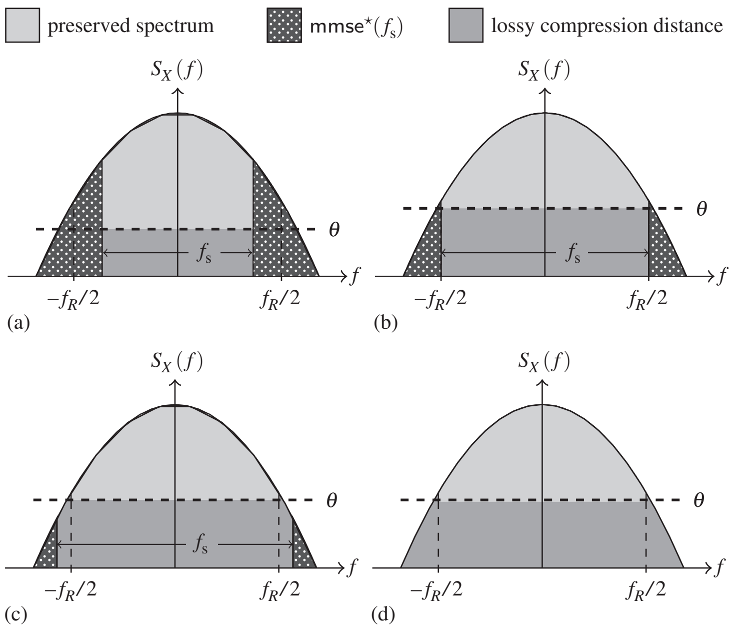

Figure 2.11 Water-filling interpretation for the function $D^\star(f_s, R)$ under zero noise, a fixed bitrate R, and three sampling rates: (a) $f_s < f_R$, (b) $f_s = f_R$, and (c) $f_s > f_R$. (d) corresponds to the DRF of $X(\cdot)$ at bitrate R. This DRF is attained whenever $f_s \geq f_R$, where f_R is smaller than the Nyquist rate.

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

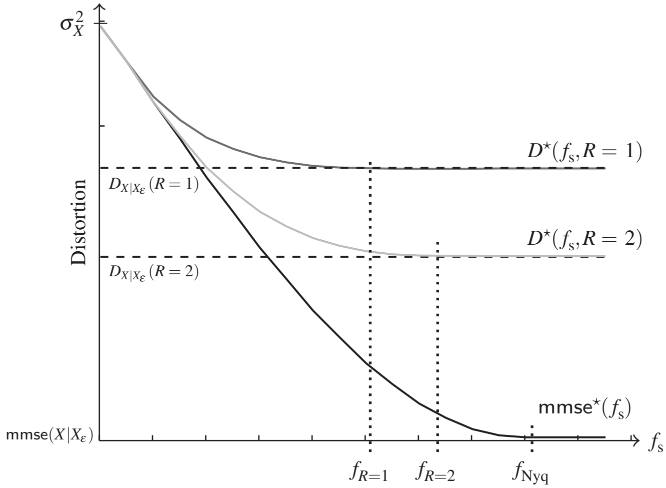

Figure 2.12 The function $D^\star(f_s, R)$ for the PSD of Fig. 2.11 and two values of the bitrate R. Also shown is the DRF of $X(\cdot)$ at these values that is attained at the sub-Nyquist sampling rates marked by f_R.

have $S_{X|X_\varepsilon}(f) = S_X(f)$ since the noise is zero, $f_{\mathsf{Lnd}} = f_{\mathsf{Nyq}}$ since $S_X(f)$ has a connected support, and $F^\star(f_s)$ is the interval of length f_s centered around the origin since $S_X(f)$ is unimodal. In all cases in Fig. 2.11 the bitrate R is fixed and corresponds to the preserved part of the spectrum through (2.19b). The distortion $D^\star(f_s, R)$ changes with f_s, and is given by the sum of two terms in (2.19a): $\mathsf{mmse}^\star(f_s)$ and the lossy compression distortion. For example, the increment in f_s from (a) to (b) reduces $\mathsf{mmse}^\star(f_s)$ and increases the lossy compression distortion, although the overall distortion decreases due to this increment. However, the increase in f_s leading from (b) to (c) is different: while (c) shows an additional reduction in $\mathsf{mmse}^\star(f_s)$ compared with (b), the sum of the two distortion terms is identical in both cases and, as illustrated in (d), equals the DRF of $X(\cdot)$ from (2.16). It follows that, in the case of Fig. 2.11, the optimal ADX performance is attained at some sampling rate f_R that is smaller than the Nyquist rate, and depends on the bitrate R through expression (2.16). The full behavior of $D^\star(f_s, R)$ as a function of f_s is illustrated in Fig. 2.12 for two values of R.

The phenomenon described above and in Figs. 2.11 and 2.12 can be generalized to any Gaussian stationary process with arbitrary PSD and noise in the ADX setting, according to the following theorem.

THEOREM 2.3 (optimal sampling rate [24]) *Let $X(\cdot)$ be a Gaussian stationary process with PSD $S_X(f)$ corrupted by a Gaussian noise $\varepsilon(\cdot)$. For each point (R, D) on the graph of $D_{X|X_\varepsilon}(R)$ associated with a water-level θ via* (2.15)*, let f_R be the Lebesgue measure of the set*

$$F_\theta \triangleq \{f : S_{X|X_\varepsilon}(f) \geq \theta\}.$$

Then, for all $f_s \geq f_R$,

$$D^\star(f_s, R) = D_{X|X_\varepsilon}(R).$$

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

The proof of Theorem 2.3 is relatively straightforward and follows from the definition of F_θ and $D^\star(f_s, R)$.

We emphasize that the critical frequency f_R depends only on the PSDs $S_X(f)$ and $S_\varepsilon(f)$, and on the operating point on the graph of $D^\star(f_s, R)$. This point may be parametrized by D, R, or the water-level θ using (2.15). Furthermore, we can consider a version of Theorem 2.3 in which the bitrate is a function of the distortion and the sampling rate, by inverting $D^\star(f_s, R)$ with respect to R. This inverse function, $R^\star(f_s, D)$, is the minimal number of bits per unit time one must provide on the samples of $X_\varepsilon(\cdot)$, obtained by any bounded linear sampler with sampling density not exceeding f_s, in order to attain distortion not exceeding D. The following representation of $R^\star(f_s, D)$ in terms of f_R is equivalent to Theorem 2.3.

THEOREM 2.4 (rate-distortion lower bound) *Consider the samples of a Gaussian stationary process $X(\cdot)$ corrupted by a Gaussian noise $\varepsilon(\cdot)$ obtained by a bounded linear sampler of maximal sampling density f_s. The bitrate required to recover $X(\cdot)$ with MSE at most $D > \mathsf{mmse}^\star(f_s)$ is at least*

$$R^\star(f_s, D) = \begin{cases} \frac{1}{2}\int_{F^\star(f_s)} \log^+\left(\frac{f_s S_{X|X_\epsilon}(f)}{D - \mathsf{mmse}^\star(f_s)}\right) df, & f_s < f_R, \\ R_{X|X_\epsilon}(D), & f_s \geq f_R, \end{cases} \tag{2.21}$$

where

$$R_{X|X_\varepsilon}(D_\theta) = \frac{1}{2}\int_{-\infty}^{\infty} \log^+[S_{X|X_\varepsilon}(f)/\theta] df$$

is the indirect rate-distortion function of $X(\cdot)$ given $X_\varepsilon(\cdot)$, and θ is determined by

$$D_\theta = \mathsf{mmse}^\star(f_s) + \int_{-\infty}^{\infty} \min\{S_{X|X_\varepsilon}(f), \theta\} df.$$

Theorems 2.3 and 2.4 imply that the equality in (2.20), which was previously shown to hold for $f_s \geq f_{\mathsf{Lnd}}$, is extended to all sampling rates above $f_R \leq f_{\mathsf{Lnd}}$. As R goes to infinity, $D^\star(f_s, R)$ converges to $\mathsf{mmse}^\star(f_s)$, the water-level θ goes to zero, the set F_θ coincides with the support of $S_X(f)$, and f_R converges to f_{Lnd}. Theorem 2.3 then implies that $\mathsf{mmse}^\star(f_s) = 0$ for all $f_s \geq f_{\mathsf{Lnd}}$, a fact that agrees with Landau's characterization of sets of sampling for perfect recovery of signals in the Paley–Wiener space, as explained in Section 2.3.3.

An intriguing way to explain the critical sampling rate subject to a bitrate constraint arising from Theorem 2.3 follows by considering the degrees of freedom in the representation of the analog signal pre- and post-sampling and with lossy compression of the samples. For stationary Gaussian signals with zero sampling noise, the degrees of freedom in the signal representation are those spectral bands in which the PSD is non-zero. When the signal energy is not uniformly distributed over these bands, the optimal lossy compression scheme described by (2.16) calls for discarding those bands with the lowest energy, i.e., the parts of the signal with the lowest uncertainty.

The degree to which the new critical rate f_R is smaller than the Nyquist rate depends on the energy distribution of $X(\cdot)$ across its spectral occupancy. The more uniform this

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

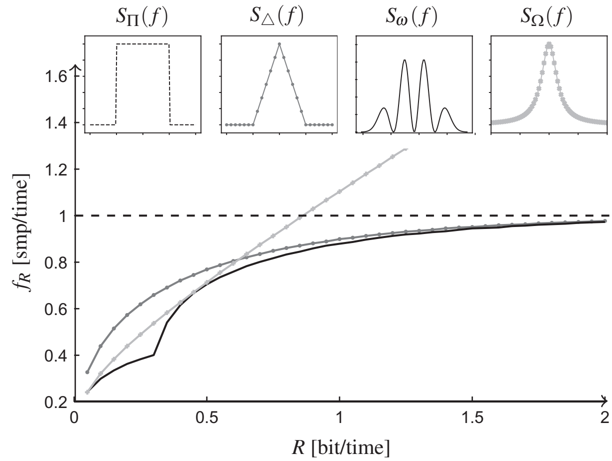

Figure 2.13 The critical sampling rate f_R as a function of the bitrate R for the PSDs given in the small frames at the top of the figure. For the bandlimited PSDs $S_\Pi(f)$, $S_\triangle(f)$, and $S_\omega(f)$, the critical sampling rate is always at or below the Nyquist rate. The critical sampling rate is finite for any R, even for the non-bandlimited PSD $S_\Omega(f)$.

distribution, the more degrees of freedom are required to represent the lossy compressed signal and therefore the closer f_R is to the Nyquist rate. In the examples below we derive the precise relation between f_R and R for various PSDs. These relations are illustrated in Fig. 2.13 for the signals $S_\Pi(f)$, $S_\triangle(f)$, $S_\omega(f)$, and $S_\Omega(f)$ defined below.

2.4.4 Examples

Example 2.1 Consider the Gaussian stationary process $X_\triangle(\cdot)$ with PSD

$$S_\triangle(f) \triangleq \sigma^2 \frac{[1-|f/W|]^+}{W},$$

for some $W > 0$. Assuming that the noise is zero,

$$F_\theta = [W(W\theta - 1), W(1 - W\theta)]$$

and thus $f_R = 2W(1 - W\theta)$. The exact relation between f_R and R is obtained from (2.19b) and found to be

$$R = \frac{1}{2}\int_{-f_R/2}^{f_R/2} \log\left(\frac{1-|f/W|}{1-f_R/2W}\right) df = W \log\left(\frac{2W}{2W-f_R}\right) - \frac{f_R}{2\ln 2}.$$

In particular, note that $R \to \infty$ leads to $f_R \to f_{\mathsf{Nyq}} = 2W$, as anticipated.

Example 2.2 Let $X_\Pi(\cdot)$ be the process with PSD

$$S_\Pi(f) = \frac{\mathbf{1}_{|f|<W}(f)}{2W}. \tag{2.22}$$

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

Assume that $\varepsilon(\cdot)$ is noise with a flat spectrum within the band $(-W,W)$ such that $\gamma \triangleq S_\Pi(f)/S_\varepsilon(f)$ is the SNR at the spectral component f. Under these conditions, the water-level θ in (2.15) satisfies

$$\theta = \sigma^2 \frac{\gamma}{1+\gamma} 2^{-R/W},$$

and hence

$$\frac{D_{X|X_\varepsilon}R)}{\sigma^2} = \frac{1}{1+\gamma} + \frac{\gamma}{1+\gamma} 2^{-R/W}. \tag{2.23}$$

In particular, $F_\theta = [-W,W]$, so that $f_R = 2W = f_{\mathsf{Nyq}}$ for any bitrate R and $D^\star(f_s,R) = D_{X|X_\varepsilon}(R)$ only for $f_s \geq f_{\mathsf{Nyq}}$. That is, for the process $X_\Pi(\cdot)$, optimal sampling under a bitrate constraint occurs only at or above its Nyquist rate.

Example 2.3 Consider the PSD

$$S_\Omega(f) \triangleq \frac{\sigma^2/f_0}{(\pi f/f_0)^2+1},$$

for some $f_0 > 0$. The Gaussian stationary process $X_\Omega(\cdot)$ corresponding to this PSD is a Markov process, and it is in fact the unique Gaussian stationary process that is also Markovian (a.k.a. the Ornstein–Uhlenbeck process). Since the spectral occupancy of $X_\Omega(\cdot)$ is the entire real line, its Nyquist and Landau rates are infinite and it is impossible to recover it with zero MSE using any bounded linear sampler. Assuming $\varepsilon(\cdot) \equiv 0$ and noting that $S_\Omega(f)$ is unimodal, $F^\star(f_s) = (-f_2/2, f_s/2)$, and thus

$$\mathsf{mmse}^\star(f_s) = 2\int_{f_s/2}^{\infty} S_\Omega(f)df = \sigma^2\left[1 - \frac{2}{\pi}\arctan\left(\frac{\pi f_s}{2f_0}\right)\right].$$

Consider now a point (R,D) on the DRF of $X_\Omega(\cdot)$ and its corresponding water-level θ determined from (2.16). It follows that $f_R = (2f_s/\pi)\sqrt{1/\theta f_s - 1}$, so that Theorem 2.3 implies that the distortion cannot be reduced below D by sampling above this rate. The exact relation between R and f_R is found to be

$$R = \frac{1}{\ln 2}\left(f_R - \frac{2f_0}{\pi}\arctan\left(\frac{\pi f_R}{2f_0}\right)\right). \tag{2.24}$$

Note that, although the Nyquist rate of $X_\Omega(\cdot)$ is infinite, for any finite R there exists a critical sampling frequency f_R satisfying (2.24) such that $D_{X_\Omega}(R)$ is attained by sampling at or above f_R.

The asymptotic behavior of (2.24) as R goes to infinity is given by $R \sim f_R/\ln 2$. Thus, for R sufficiently large, the optimal sampling rate is linearly proportional to R. The ratio R/f_s is the average number of bits per sample used in the resulting digital representation. It follows from (2.24) that, asymptotically, the "right" number of bits per sample converges to $1/\ln 2 \approx 1.45$. If the number of bits per sample is below this value, then

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

the distortion in ADX is dominated by the DRF $D_{X_\Omega}(\cdot)$, as there are not enough bits to represent the information acquired by the sampler. If the number of bits per sample is greater than this value, then the distortion in ADX is dominated by the sampling distortion, as there are not enough samples for describing the signal up to a distortion equal to its DRF.

As a numerical example, assume that we encode $X_\Omega(t)$ using two bits per sample, i.e., $f_s = 2R$. As $R \to \infty$, the ratio between the minimal distortion $D^\star(f_s, R)$ and $D_{X_\Omega}(R)$ converges to approximately 1.08, whereas the ratio between $D^\star(f_s, R)$ and $\mathsf{mmse}^\star(f_s)$ converges to approximately 1.48. In other words, it is possible to attain the optimal encoding performance within an approximately 8% gap by providing one sample per each two bits per unit time used in this encoding. On the other hand, it is possible to attain the optimal sampling performance within an approximately 48% gap by providing two bits per each sample taken.

2.5 ADX under Uniform Sampling

We now analyze the distortion in the ADX setting of Fig. 2.5 under the important class of single- and multi-branch LTI uniform samplers. Our goal in this section is to show that for any source and noise PSDs, $S_X(f)$ and $S_\varepsilon(f)$, respectively, the function $D^\star(f_s, R)$ describing the fundamental distortion limit in ADX is attainable using a multi-branch LTI uniform sampler. By doing so, we also provide a proof of Theorem 2.2.

We begin by analyzing the ADX system of Fig. 2.5 under an LTI uniform sampler. As we show, the asymptotic distortion in this case can be obtained in a closed form that depends only on the signal and noise PSDs, the sampling rate, the bitrate, and the pre-sampling filter $H(f)$. We then show that, by taking $H(f)$ to be a low-pass filter with cutoff frequency $f_s/2$, we can attain the fundamental distortion limit $D^\star(f_s, R)$ whenever the function $S_{X|X_\varepsilon}(f)$ of (2.13) attains its maximum at the origin. In the more general case of an arbitrarily shaped $S_{X|X_\varepsilon}(f)$, we use multi-branch sampling in order to achieve $D^\star(f_s, R)$.

2.5.1 Single-Branch LTI Uniform Sampling

Assume that the sampler S is the LTI uniform sampler defined in Section 2.3.1 and illustrated in Fig. 2.7. This sampler is characterized by its sampling rate f_s and the frequency response $H(f)$ of the pre-sampling filter.

In Section 2.3.3, we saw that, for any bounded linear sampler, optimal encoding in ADX is obtained by first forming the estimator $\widetilde{X}_T(\cdot)$ from $\mathbf{Y}_T$, and then encoding $\widetilde{X}_T(\cdot)$ in an optimal manner subject to the bitrate constraint. That is, the encoder performs estimation under an MSE criterion followed by optimal source coding for this estimate. Under the LTI uniform sampler, the process $\widetilde{X}_T(\cdot)$ has an asymptotic distribution described by the conditional expectation of $X(\cdot)$ given the sigma

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

algebra generated by $\{X_\varepsilon(n/f_s), n \text{ in } \mathbb{Z}\}$. Using standard linear estimation techniques, this conditional expectation has a representation similar to that of a Wiener filter given by [12]:

$$\widetilde{X}(t) \triangleq \mathbb{E}[X(t)|\{X(n/f_s), n \in \mathbb{Z}\}] = \sum_{n\in\mathbb{Z}} X_\varepsilon(n/f_s)w(t-n/f_s), \quad t \in \mathbb{R}, \tag{2.25}$$

where the Fourier transform of $w(t)$ is

$$W(f) = \frac{S_X(f)|H(f)|^2}{\sum_{k\in\mathbb{Z}} S_{X_\varepsilon}(f-kf_s)|H(f-kf_s)|^2}.$$

Moreover, the resulting MMSE, which is the asymptotic value of $\mathsf{mmse}_T(S)$, is given by

$$\mathsf{mmse}_H(f_s) \triangleq \sum_{n\in\mathbb{Z}} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \left[S_X(f-nf_s) - \widetilde{S}_X(f)\right] df, \tag{2.26}$$

where

$$\widetilde{S}_X(f) \triangleq \frac{S_X^2(f-f_s n)|H(f-f_s n)|^2}{\sum_{n\in\mathbb{Z}} S_{X_\varepsilon}(f-f_s n)|H(f-f_s n)|^2}. \tag{2.27}$$

From the decomposition (2.11), it follows that, when S is an LTI uniform sampler, the distortion can be expressed as

$$D_H(f_s,R) \triangleq \liminf_{T\to\infty} D_T(S,R) = \mathsf{mmse}_H(f_s) + D_{\widetilde{X}}(R),$$

where $D_{\widetilde{X}}(R)$ is the DRF of the Gaussian process $\widetilde{X}(\cdot)$ defined by (2.25), satisfying the law of the process $\widetilde{X}_T(\cdot)$ in the limit as T goes to infinity.

Note that, whenever $f_s \geq f_{\mathsf{Nyq}}$ and supp S_X is included within the passband of $H(f)$, we have that $\widetilde{S}_X(f) = S_{X|X_\varepsilon}(f)$ and thus $\mathsf{mmse}_H(f_s) = \mathsf{mmse}(X|X_\varepsilon)$, i.e., no distortion due to sampling. Moreover, in this situation, $\widetilde{X}(t) = \mathbb{E}[X(t)|X_\varepsilon(\cdot)]$ and

$$D_H(f_s,R) = D_{X|X_\varepsilon}(R). \tag{2.28}$$

The equality (2.28) is a special case of (2.20) for LTI uniform sampling, and says that there is no loss due to sampling in ADX whenever the sampling rate exceeds the Nyquist rate of $X(\cdot)$.

When the sampling rate is below f_{Nyq}, (2.25) implies that the estimator $\widetilde{X}(\cdot)$ has the form of a stationary process modulated by a deterministic pulse, and is therefore a *block-stationary* process, also called a *cyclostationary* process [25]. The DRF for this class of processes can be described by a generalization of the orthogonal transformation and rate allocation that leads to the water-filling expression (2.16) [26]. Evaluating the resulting expression for the DRF of the cyclostationary process $\widetilde{X}(\cdot)$ leads to a closed-form expression for $D_H(f_s,R)$, which was initially derived in [27].

THEOREM 2.5 (achievability for LTI uniform sampling) *Let $X(\cdot)$ be a Gaussian stationary process corrupted by a Gaussian stationary noise $\varepsilon(\cdot)$. The minimal distortion in*

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

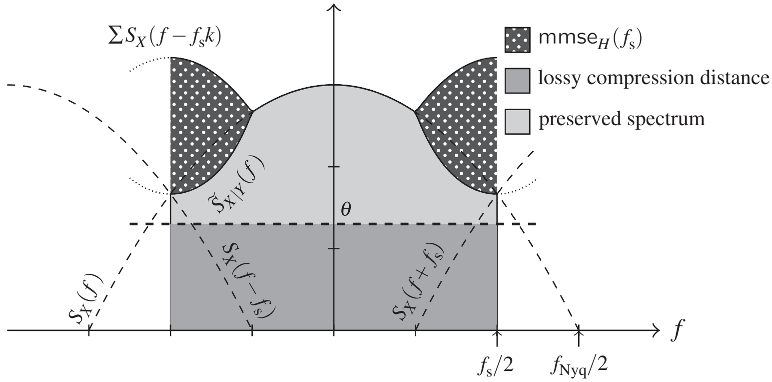

Figure 2.14 Water-filling interpretation of (2.29) with an all-pass filter $H(f)$. The function $\sum_{k\in\mathbb{Z}} S_X(f-f_s k)$ is the aliased PSD that represents the full energy of the original signal within the discrete-time spectrum interval $(-f_s/2, f_s/2)$. The part of the energy recovered by the $\widetilde{X}(\cdot)$ is $\widetilde{S}_X(f)$. The distortion due to lossy compression is obtained by water-filling over the recovered energy according to (2.29a). The overall distortion $D_H(f_s, R)$ is the sum of the sampling distortion and the distortion due to lossy compression.

ADX at bitrate R with an LTI uniform sampler with sampling rate f_s and pre-sampling filter $H(f)$ is given by

$$D_H(f_s, R_\theta) = \text{mmse}_H(f_s) + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min\left\{\widetilde{S}_X(f), \theta\right\} df, \tag{2.29a}$$

$$R_\theta = \frac{1}{2}\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log_2^+\left[\widetilde{S}_X(f)/\theta\right] df, \tag{2.29b}$$

where $\text{mmse}_H(f_s)$ *and* $\widetilde{S}_X(f)$ *are given by* (2.26) *and* (2.27)*, respectively.*

A graphical water-filling interpretation of (2.29) is provided in Fig. 2.14. This expression combines the MMSE (2.26), which depends only on f_s and $H(f)$, with the expression for the indirect DRF of (2.15), which also depends on the bitrate R. The function $\widetilde{S}_X(f)$ arises in the MMSE estimation of $X(\cdot)$ from its samples and can be interpreted as an average over the PSD of polyphase components of the cyclostationary process $\widetilde{X}(\cdot)$ [26].

As is implicit in the analysis above, the coding scheme that attains $D_H(f_s, R)$ is described by the decomposition of the non-causal MSE estimate of $X(\cdot)$ from its samples $\mathbf{Y}_T$. This estimate is encoded using a codebook with $2^{\lfloor TR \rfloor}$ elements that attains the DRF of the Gaussian process $\widetilde{X}(\cdot)$ at bitrate R, and the decoded codewords (which is a waveform over $[-T/2, T/2]$) are used as the reconstruction of $X(\cdot)$. For any $\rho > 0$, the MSE resulting from this process can be made smaller than $D_H(f_s, R-\rho)$ by taking T to be sufficiently large.

Example 2.4 (continuation of Example 2.2) As a simple example for using formula (2.29), consider the process $X_\Pi(\cdot)$ of Example 2.2. Assuming that the noise $\varepsilon(\cdot) \equiv 0$ (equivalently, $\gamma \to \infty$) and that $H(f)$ passes all frequencies $f \in [-W, W]$, the relation between the distortion in (2.29a) and the bitrate in (2.29b) is given by

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

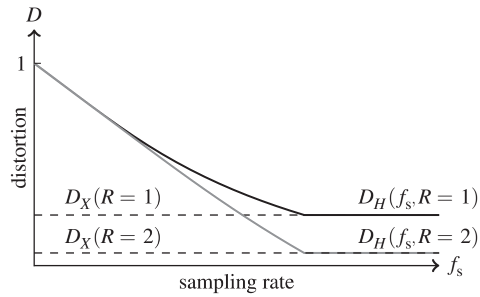

Figure 2.15 Distortion as a function of sampling rate for the source with PSD $S_\Pi(f)$ of (2.22), zero noise, and source coding rates $R=1$ and $R=2$ bits per time unit.

$$D_H(f_s,R) = \begin{cases} \mathsf{mmse}_H(f_s) + \sigma^2 \frac{f_s}{2W} 2^{-\frac{2R}{f_s}}, & f_s < 2W, \\ \sigma^2 2^{-\frac{R}{W}}, & f_s \geq 2W, \end{cases} \tag{2.30}$$

where $\mathsf{mmse}_H(f_s) = \sigma^2[1-f_s/2W]^+$. Expression (2.30) is shown in Fig. 2.15 for two fixed values of the bitrate R. It has a very intuitive structure: for frequencies below $f_{\mathsf{Nyq}} = 2W$, the distortion as a function of the bitrate increases by a constant factor due to the error as a result of non-optimal sampling. This factor completely vanishes once the sampling rate exceeds the Nyquist frequency, in which case $D_H(f_s,R)$ coincides with the DRF of $X(\cdot)$.

In the noisy case when $\gamma = S_\Pi(f)/S_\varepsilon(f)$, we have $\mathsf{mmse}_H(f_s) = \sigma^2(1 - f_s/(2W(1+\gamma)))$ and the distortion takes the form

$$D^\star(f_s,R) = \sigma^2 \begin{cases} \mathsf{mmse}_H(f_s) + (f_s/2W)(\gamma/(1+\gamma))2^{-2R/f_s}, & f_s < 2W, \\ \mathsf{mmse}(X|X_\varepsilon) + (\gamma/(1+\gamma))2^{-R/W}, & f_s \geq 2W, \end{cases} \tag{2.31}$$

where $\mathsf{mmse}(X|X_\varepsilon) = 1/(1+\gamma)$.

Next, we show that, when $S_{X|X_\varepsilon}(f)$ is unimodal, an LTI uniform sampler can be used to attain $D^\star(f_s,R)$.

2.5.2 Unimodal PSD and Low-Pass Filtering

Under the assumption that $H(f)$ is an ideal low-pass filter with cutoff frequency $f_s/2$, (2.29) becomes

$$D_{\mathrm{LPF}}(f_s,R_\theta) = \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} [S_X(f) - S_{X|X_\varepsilon}(f)]df + \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min\{S_{X|X_\varepsilon}(f),\theta\}df, \tag{2.32a}$$

$$R_\theta = \frac{1}{2}\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log_2^+[S_{X|X_\varepsilon}(f)/\theta]df. \tag{2.32b}$$

Comparing (2.32) with (2.19), we see that the two expressions coincide whenever the interval $[-f_s/2, f_s/2]$ minimizes (2.18). Therefore, we conclude that, when the function $S_{X|X_\varepsilon}(f)$ is unimodal in the sense that it attains its maximal value at the origin, the

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

fundamental distortion in ADX is attained using an LTI uniform sampler with a low-pass filter of cutoff frequency $f_s/2$ as its pre-sampling operation.

An example for a PSD for which (2.32) describes its fundamental distortion limit is the one in Fig. 2.11. Note the LPF with cutoff frequency $f_s/2$ in cases (a)–(c) there. Another example of this scenario for a unimodal PSD is given in Example 2.4 above.

Example 2.5 (continuation of Examples 2.2 and 2.4) In the case of the process $X_\Pi(\cdot)$ with a flat spectrum noise as in Examples 2.2 and 2.4, (2.32) leads to (2.31). It follows that the fundamental distortion limit in ADX with respect to $X_\Pi(\cdot)$ and a flat spectrum noise is given by (2.31), which was obtained from (2.29). Namely, the fundamental distortion limit in this case is obtained using any pre-sampling filter whose passband contains $[-W, W]$, and using an LPF is unnecessary.

In particular, the distortion in (2.31) corresponding to $f_s \geq f_{\mathsf{Nyq}}$ equals the indirect DRF of $X(\cdot)$ given $X_\varepsilon(\cdot)$, which can be found directly from (2.15). Therefore, (2.31) implies that optimal sampling for $X_\Pi(\cdot)$ under LTI uniform sampling occurs only at or above its Nyquist rate. This conclusion is not surprising since, according to Example 2.2, super-Nyquist sampling of $X_\Pi(\cdot)$ is necessary for (2.20) to hold under any bounded linear sampler.

The analysis above implies in particular that a distortion of $D^\star(f_s, R)$ is achievable using LTI uniform sampling for any signal $X(\cdot)$ with a unimodal PSD in the noiseless case or unimodal $S_{X|X_\varepsilon}(f)$ in the noisy case. Therefore, Theorem 2.5 implies Theorem 2.2 in these special cases. In what follows, we use multi-branch sampling in order to show that Theorem 2.2 holds for an arbitrary $S_{X|X_\varepsilon}(f)$.

2.5.3 Multi-Branch LTI Uniform Sampling

We now consider the ADX system of Fig. 2.5 where the sampler is the multi-branch sampler defined in Section 2.3.1 and illustrated in Fig. 2.8. This sampler is characterized by L filters $H_1(f), \ldots, H_L(f)$ and a sampling rate f_s.

The generalization of Theorem 2.5 under this sampler is as follows [27].

THEOREM 2.6 *Let $X(\cdot)$ be a Gaussian stationary process corrupted by a Gaussian stationary noise $\varepsilon(\cdot)$. The minimal distortion in ADX at bitrate R with a multi-branch LTI uniform sampler is given by*

$$D_{H_1,\ldots,H_L}(f_s, R) = \mathsf{mmse}_{H_1,\ldots,H_L}(f_s) + \sum_{l=1}^{L} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \min\{\lambda_l(f)\} df, \tag{2.33a}$$

$$R_\theta = \frac{1}{2} \sum_{l=1}^{L} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log_2^+[\lambda_l(f)/\theta] df, \tag{2.33b}$$

https://doi.org/10.1017/9781108616799.003 Published online by Cambridge University Press

where $\lambda_1, \dots, \lambda_L$ are the eigenvalues of the matrix

$$\widetilde{\mathbf{S}}_X(f) = \left(\mathbf{S}_Y^{-\frac{1}{2}}(f)\right)^H \mathbf{K}(f)\mathbf{S}_Y^{-\frac{1}{2}}(f),$$

with

$$(\mathbf{S}_Y(f))_{i,j} = \sum_{n\in\mathbb{Z}} S_{X_\varepsilon}(f - f_s n)H_i(f - f_s n)H_j^*(f - f_s n), \quad i, j = 1, \dots, L,$$

$$(\mathbf{K}(f))_{i,j} = \sum_{n\in\mathbb{Z}} S_X^2(f - f_s n)H_i(f - f_s n)H_j^*(f - f_s n), \quad i, j = 1, \dots, L.$$

In addition,

$$\mathsf{mmse}_{H_1,\dots,H_L}(f_s) \triangleq \sigma_X^2 - \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} tr\left(\widetilde{\mathbf{S}}_X(f)\right) df$$

is the minimal MSE in estimating $X(\cdot)$ from the combined output of the L sampling branches as T approaches infinity.

The most interesting feature in the extension of (2.29) provided by (2.33) is the dependences between samples obtained over different branches, expressed in the definition of the matrix $\widetilde{\mathbf{S}}_X(f)$. In particular, if $f_s \geq f_{\mathsf{Nyq}}$, then we may choose the bandpasses of the L filters to be a set of L disjoint intervals, each of length at most f_s/L, such that the union of their supports contains the support of this choice, the matrix $\widetilde{\mathbf{S}}_X(f)$ is diagonal and its eigenvalues are

$$\lambda_l = \widetilde{S}_l(f) \triangleq \frac{\sum_{n\in\mathbb{Z}} S_X^2(f - f_s n)}{\sum_{n\in\mathbb{Z}} S_{X+\varepsilon}(f - f_s n)} \mathbf{1}_{\mathrm{supp} H_l}(f).$$

Since the union of the filters' support contains the support of $S_{X|X_\varepsilon}$, we have

$$D_{H_1,\dots,H_L}(f_s, R) = D_{X|X_\varepsilon}(R).$$

While it is not surprising that a multi-branch sampler attains the optimal sampling distortion when f_s is above the Nyquist rate, we note that at each branch the sampling rate can be as small as f_{Nyq}/L. This last remark suggests that a similar principle may be used under sub-Nyquist sampling to sample those particular parts of the spectrum of maximal energy whenever $S_{X|X_\varepsilon}(f)$ is not unimodal.

Our goal now is to prove Theorem 2.2 by showing that, for any PSDs $S_X(f)$ and $S_\varepsilon(f)$, the distortion in (2.33) can be made arbitrarily close to the fundamental distortion limit $D^\star(f_s, R)$ with an appropriate choice of the number of sampling branches and their filters. Using the intuition gained above, given a sampling rate f_s we cover the set of maximal energy $F^\star(f_s)$ of (2.18) using L disjoint intervals, such that the length of each interval does not exceed f_s/L. For any $\varepsilon > 0$, it can be shown that there exists L large enough such $\int_\Delta S_{X|X_\varepsilon}(f)df < \varepsilon$, where Δ is the part that is not covered by the L intervals [28].

From this explanation, we conclude that, for any PSD $S_{X|X_\varepsilon}(f)$, $f_s > 0$, and $\varepsilon > 0$, there exists an integer L and a set of L pre-sampling filters $H_1(f), \ldots, H_L(f)$ such that, for every bitrate R,

$$D_{H_1,\ldots,H_L}(f_s, R) \leq D^\star(f_s, R) + \varepsilon. \tag{2.34}$$

Since $D_{H_1,\ldots,H_L}(f_s, R)$ is obtained in the limit as T approaches infinity of the minimal distortion in ADX under the aforementioned multi-branch uniform sampler, the fundamental distortion limit in ADX is achieved up to an arbitrarily small constant.

The description starting from Theorem 2.6 and ending in (2.34) sketches the proof of the achievability side of the fundamental ADX distortion (Theorem 2.2). Below we summarize the main points in the procedure described in this section.

(i) Given a sampling rate f_s, use a multi-branch LTI uniform sampler with a sufficient number of sampling branches L that the effective passband of all branches is close enough to $F^\star$, which is a set of Lebesgue measure f_s that maximizes (2.18).
(ii) Estimate the signal $X(\cdot)$ under an MSE criterion, leading to $\widetilde{X}_T(\cdot)$ defined in (2.10). As $T \to \infty$ this process converges in $\mathbf{L}_2$ norm to $\widetilde{X}(\cdot)$ defined in (2.25).
(iii) Given a bitrate constraint R, encode a realization of $\widetilde{X}_T(\cdot)$ in an optimal manner subject to an MSE constraint as in standard source coding [7]. For example, for $\rho > 0$ arbitrarily small, we may use a codebook consisting of $2^{\lfloor T(R+\rho)\rfloor}$ waveforms of duration T generated by independent draws from the distribution defined by the preserved part of the spectrum in Fig. 2.10. We then use minimum distance encoding with respect to this codebook.

2.6 Conclusion

The processing, communication, and digital storage of an analog signal requires first representing it as a bit sequence. Hardware and modeling constraints in processing analog information imply that the digital representation is obtained by first sampling the analog waveform and then quantizing or encoding its samples. That is, the transformation from analog signals to bits involves the composition of sampling and quantization or, more generally, lossy compression operations.

In this chapter we explored the minimal sampling rate required to attain the fundamental distortion limit in reconstructing a signal from its quantized samples subject to a strict constraint on the bitrate of the system. We concluded that, when the energy of the signal is not uniformly distributed over its spectral occupancy, the optimal signal representation can be attained by sampling at some critical rate that is lower than the Nyquist rate or, more generally, the Landau rate, in bounded linear sampling. This critical sampling rate depends on the bitrate constraint, and converges to the Nyquist or Landau rates in the limit of infinite bitrate. This reduction in the optimal sampling rate under finite bit precision is made possible by designing the sampling mechanism to sample only those parts of the signals that are not discarded due to optimal lossy compression.

The information-theoretic approach to analog-to-digital compression explored in this chapter can be extended in various directions. First, while we considered the minimal sampling rate and resulting distortion under an ideal encoding of the samples, such an encoding is rarely possible in practice. Indeed, in most cases, the encoding of the samples is subject to additional constraints in addition to the bit resolution, such as complexity, time delay, or limited information on the distribution of the signal and the noise. It is therefore important to characterize the optimal sampling rate and resulting distortion under these limitations. In addition, the reduction in the optimal sampling rate under the bitrate constraint from the Nyquist rate to f_R can be understood as the result of a reduction in degrees of freedom in the compressed signal representation compared with the original source. It is interesting to consider whether a similar principle holds for non-stationary [29] or non-Gaussian [30, 31] signal models (e.g., sparse signals).

References

[1] Y. C. Eldar, *Sampling theory: Beyond bandlimited systems*. Cambridge University Press, 2015.

[2] R. M. Gray and D. L. Neuhoff, “Quantization,” *IEEE Trans. Information Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.

[3] C. E. Shannon, “Communication in the presence of noise,” *IRE Trans. Information Theory*, vol. 37, pp. 10–21, 1949.

[4] H. Landau, “Sampling, data transmission, and the Nyquist rate,” *Proc. IEEE*, vol. 55, no. 10, pp. 1701–1706, 1967.

[5] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical J.*, vol. 27, pp. 379–423, 623–656, 1948.

[6] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” *IRE National Convention Record*, vol. 4, no. 1, pp. 142–163, 1959.

[7] T. Berger, *Rate-distortion theory: A mathematical basis for data compression*. Prentice-Hall, 1971.

[8] R. Walden, “Analog-to-digital converter survey and analysis,” *IEEE J. Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, 1999.

[9] J. Candy, “A use of limit cycle oscillations to obtain robust analog-to-digital converters,” *IEEE Trans. Communications*, vol. 22, no. 3, pp. 298–305, 1974.

[10] B. Oliver, J. Pierce, and C. Shannon, “The philosophy of PCM,” *IRE Trans. Information Theory*, vol. 36, no. 11, pp. 1324–1331, 1948.

[11] D. L. Neuhoff and S. S. Pradhan, “Information rates of densely sampled data: Distributed vector quantization and scalar quantization with transforms for Gaussian sources,” *IEEE Trans. Information Theory*, vol. 59, no. 9, pp. 5641–5664, 2013.

[12] M. Matthews, “On the linear minimum-mean-squared-error estimation of an undersampled wide-sense stationary random process,” *IEEE Trans. Signal Processing*, vol. 48, no. 1, pp. 272–275, 2000.

[13] D. Chan and R. Donaldson, “Optimum pre- and postfiltering of sampled signals with application to pulse modulation and data compression systems,” *IEEE Trans. Communication Technol.*, vol. 19, no. 2, pp. 141–157, 1971.

[14] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Information Theory*, vol. 8, no. 5, pp. 293–304, 1962.

[15] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Information Theory*, vol. 2, no. 4, pp. 102–108, 1956.

[16] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Information Theory*, vol. 43, no. 1, pp. 38–47, 1997.

[17] I. Kontoyiannis and R. Zamir, "Mismatched codebooks and the role of entropy coding in lossy data compression," *IEEE Trans. Information Theory*, vol. 52, no. 5, pp. 1922–1938, 2006.

[18] A. H. Zemanian, *Distribution theory and transform analysis: An introduction to generalized functions, with applications*. Courier Corporation, 1965.

[19] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Information Theory*, vol. 16, no. 4, pp. 406–411, 1970.

[20] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Information Theory*, vol. 26, no. 5, pp. 518–521, 1980.

[21] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, no. 1, pp. 37–52, 1967.

[22] A. Beurling and L. Carleson, *The collected works of Arne Beurling: Complex analysis*. Birkhäuser, 1989, vol. 1.

[23] F. J. Beutler, "Sampling theorems and bases in a Hilbert space," *Information and Control*, vol. 4, nos. 2–3, pp. 97–117, 1961.

[24] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Fundamental distortion limits of analog-to-digital compression," *IEEE Trans. Information Theory*, vol. 64, no. 9, pp. 6013–6033, 2018.

[25] W. Bennett, "Statistics of regenerative digital transmission," *Bell Labs Technical J.*, vol. 37, no. 6, pp. 1501–1542, 1958.

[26] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "The distortion rate function of cyclostationary Gaussian processes," *IEEE Trans. Information Theory*, vol. 64, no. 5, pp. 3810–3824, 2018.

[27] A. Kipnis, A. J. Goldsmith, Y. C. Eldar, and T. Weissman, "Distortion rate function of sub-Nyquist sampled Gaussian sources," *IEEE Trans. Information Theory*, vol. 62, no. 1, pp. 401–429, 2016.

[28] A. Kipnis, "Fundamental distortion limits of analog-to-digital compression," Ph.D. dissertation, Stanford University, 2018.

[29] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "The distortion-rate function of sampled Wiener processes," in *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 482–499, Jan. 2019. doi: 10.1109/TIT.2018.2878446

[30] A. Kipnis, G. Reeves, and Y. C. Eldar, "Single letter formulas for quantized compressed sensing with Gaussian codebooks," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 71–75.

[31] A. Kipnis, G. Reeves, Y. C. Eldar, and A. J. Goldsmith, "Compressed sensing under optimal quantization," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2148–2152.