Chapter 4

Fourier Analysis and Sampling

Fourier analysis refers to a collection of tools that can be applied to express a function in terms of complex sinusoids, called basis elements, of different frequencies. The result of the decomposition is the amplitude and the phase to be imparted to each basis element in the reconstruction. This decomposition is termed the **frequency domain** representation of the original signal.

Fourier analysis is extremely useful in engineering, with a myriad of applications. Part of its appeal lies in the fact that basis elements are characteristic functions of linear time-invariant systems. This property, which may seem nebulous at this point, is instrumental in solving many challenging problems, and makes Fourier analysis a powerful methodology for the design of communication systems. We assume that the reader is familiar with basic Fourier analysis, and only review details that are pertinent to our treatment of communication systems. This is not intended to be a comprehensive treatment of the subject.

4.1 Fourier Series

Fourier series can be employed to express, as weighted sums of sinusoidal components, either periodic functions or functions that are time-limited. Suppose that the signal $s(t)$ is zero for all $|t| \geq \frac{T}{2}$, is integrable and satisfies

$$\int_{\mathbb{R}} |s(t)|^2 dt = \int_{-\frac{T}{2}}^{\frac{T}{2}} |s(t)|^2 dt < \infty.$$

Then, $s(t)$ possesses a **Fourier series** representation, which is defined by

$$s(t) = \begin{cases} \sum_{k=-\infty}^{\infty} \hat{s}_k e^{2\pi i \frac{k}{T} t}, & \text{if } |t| \leq \frac{T}{2} \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

where the Fourier series coefficients $\{\hat{s}_k : k \in \mathbb{Z}\}$ are given by

$$\hat{s}_k = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t) e^{-2\pi i \frac{k}{T} t} dt.$$

We can use the standard rectangular function $\mathrm{rect}(\cdot)$, defined by

$$\mathrm{rect}(t) = \begin{cases} 1, & \text{if } |t| < 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{4.2}$$

to simplify (4.1), and rewrite the Fourier representation of $s(t)$ as

$$s(t) = \sum_{k=-\infty}^{\infty} \hat{s}_k e^{2\pi i \frac{k}{T} t} \mathrm{rect}\left(\frac{t}{T}\right). \tag{4.3}$$

If $s(t)$ is periodic with $s(t + T) = s(t)$, instead of being zero for $|t| > \frac{T}{2}$, then the same result holds without the rectangular window function.

From a vector space perspective, (4.3) asserts that $s(t)$ can be expressed as a linear combination of basis elements $\{\theta_k(t) : k \in \mathbb{Z}\}$, where

$$\theta_k(t) = e^{2\pi i \frac{k}{T} t} \mathrm{rect}\left(\frac{t}{T}\right).$$

Furthermore, note that the collection of functions $\{\theta_k(t) : k \in \mathbb{Z}\}$ forms an orthogonal set under the standard inner product; that is,

$$\langle \theta_k(t), \theta_n(t) \rangle = \int_{-\infty}^{\infty} \theta_k(t) \theta_n^*(t) dt = \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{2\pi i \frac{k}{T} t} e^{-2\pi i \frac{n}{T} t} dt$$

$$= \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{2\pi i \frac{(k-n)}{T} t} dt = 0$$

for all $k \neq n$. An interesting and important aspect of Fourier series is that time-limited functions can be characterized using a discrete set of coefficients. This fact provides insight into the sampling theorem, which we will review shortly.

4.2 Fourier Transforms

The **Fourier transform** applies to functions that are not necessarily time-limited. A signal $x(t)$ is **square integrable** (or an **energy-type signal**) if

$$\|x(t)\|^2 \triangleq \int_{\mathbb{R}} |x(t)|^2 dt < \infty. \tag{4.4}$$

Then, we can express $x(t)$ using its frequency domain representation. The Fourier transform of $x(t)$, which we denote by $\hat{x}(f)$ or $\mathcal{F}[x(t)]$, is defined by

$$\hat{x}(f) = \mathcal{F}[x(t)] \triangleq \int_{\mathbb{R}} x(t)e^{-2\pi i f t} dt. \tag{4.5}$$

Using the inverse Fourier transform, the original function can also be expressed in terms of its decomposition with

$$x(t) = \mathcal{F}^{-1}[x(t)] \triangleq \int_{\mathbb{R}} \hat{x}(f)e^{2\pi i f t} df. \tag{4.6}$$

It is interesting to point out the duality between the Fourier transform and its inverse, $\mathcal{F}[\hat{x}(t)] = x(-f)$. This relation is rooted in the striking similarity between (4.5) and (4.6).

Definition 4.2.1. *The **sinc function** is defined by*

$$\operatorname{sinc}(t) \triangleq \frac{\sin(\pi t)}{\pi t}.$$

Example 4.2.2 (Rectangular Pulse). *The rectangular pulse* $\operatorname{rect}(\cdot)$, *defined in (4.2), can be used to constrain various signals in time or frequency. For $\alpha > 0$, one has $\|\operatorname{rect}(\alpha t)\|^2 = 1/\alpha < \infty$, which guarantees that Fourier analysis can be applied to this function. The Fourier transform of* $\operatorname{rect}(\alpha t)$ *can be computed as follows,*

$$\mathcal{F}[\operatorname{rect}(\alpha t)] = \int_{\mathbb{R}} \operatorname{rect}(\alpha t)e^{-2\pi i f t} dt = \int_{-\frac{1}{2\alpha}}^{\frac{1}{2\alpha}} e^{-2\pi i f t} dt$$

$$= \frac{1}{\pi f}\left(\frac{e^{\pi i f/\alpha} - e^{-\pi i f/\alpha}}{2i}\right) = \frac{\sin(\pi f/\alpha)}{\pi f}$$

$$= \frac{1}{\alpha}\operatorname{sinc}\left(\frac{f}{\alpha}\right).$$

Thus, the Fourier transform of $\operatorname{rect}(t)$ *is the aforementioned* $\operatorname{sinc}(f)$ *function, which plays a central role in the sampling and reconstruction of information signals.*

The Fourier transform $\hat{x}(f)$ of a square-integrable signal $x(t)$ also allows one to pose the question: how much signal energy is contained in the spectral band between frequencies f_0 and f_1? The answer, for $f_0 \leq f_1$, is given by the integral

$$\int_{f_0}^{f_1} |\hat{x}(f)|^2 \, df.$$

This integral allows us to interpret the quantity $|\hat{x}(f)|^2$ as the **energy spectral density** of $x(t)$. For technical reasons, we actually define the energy spectral density later in more detail. In practice, the answer to the above question also depends on whether the signal is real or complex. For real signals, the integral is typically computed over the range $f_0 < |f| < f_1$.

When condition (4.4) is not satisfied, it may be hazardous to use Fourier analysis and frequency domain representations. Strictly speaking, the Fourier transform of a function may not exist if the function behaves wildly. Casually taking the Fourier transforms of arbitrary signals should be avoided. Having said that, there will be instances where we discuss the Fourier transforms of functions that do not fulfill (4.4). In such circumstances, the argument to the Fourier transform is carefully selected to remain meaningful from an engineering viewpoint; one such example appears below.

4.2.1   The Dirac Delta Function

The **Dirac delta function** $\delta(t)$ can be defined in a naive fashion with the operational rule

$$x(t) = \int_{\mathbb{R}} \delta(t - \tau)x(\tau)d\tau. \tag{4.7}$$

A more rigorous approach, based on generalized functions, is out of the scope of this class. So, these notes adopt a somewhat cavalier attitude towards the Fourier transform of $\delta(t)$ and rely on experience to avoid pitfalls. The benefit of this approach to Fourier analysis is that it rapidly leads to valuable engineering insight. On the downside, the reader is left with the burden of deciding whether a signal has a proper spectral representation, or if the definition of the Fourier transform is being applied loosely.

Starting with signal $x(t)$, we can write

$$x(t) = \int_{\mathbb{R}} \hat{x}(f) e^{2\pi i f t} df = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} x(\tau) e^{-2\pi i f \tau} d\tau \right] e^{2\pi i f t} df$$
$$= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} e^{2\pi i f(t-\tau)} df \right] x(\tau) d\tau, \tag{4.8}$$

where the second equality follows from (4.5) and the third equality is obtained by changing the order of integration. Since (4.8) holds for any time t, it follows from (4.7) that

$$\delta(t) = \int_{\mathbb{R}} e^{2\pi i f t} df$$

is one representation of $\delta(t)$ and hence the (cavalier) Fourier transform of the δ-function is $\mathcal{F}[\delta(t)] = 1$.

4.2.2   Periodic Signals

We can develop (cavalier) Fourier transform representations for periodic signals as well, thereby providing a unified treatment of periodic and aperiodic functions. Indeed, we can construct the Fourier transform of a periodic signal directly from its Fourier series representation. Let $x(t)$ be a signal with Fourier transform $\hat{x}(f) = \delta(f - f_0)$. To recover the signal $x(t)$, we can apply the inverse Fourier transform

$$x(t) = \mathcal{F}^{-1}[\delta(f - f_0)] = \int_{\mathbb{R}} \delta(f - f_0) e^{2\pi i f t} df = e^{2\pi i f_0 t}.$$

More generally, if $\hat{x}(f)$ is a linear combination of impulses equally spaced in frequency

$$\hat{x}(f) = \sum_{k=-\infty}^{\infty} \hat{s}_k \delta(f - k f_0), \tag{4.9}$$

then its inverse Fourier transform becomes

$$x(t) = \sum_{k=-\infty}^{\infty} \hat{s}_k e^{2\pi i k f_0 t}. \tag{4.10}$$

Note that (4.10) corresponds to the Fourier series representation of a periodic signal. Thus, the Fourier transform of a periodic signal with Fourier series coefficients $\{\hat{s}_k : k \in \mathbb{Z}\}$ can be interpreted as a train of impulses in the frequency domain.

A signal that will be useful in our analysis of sampling is the impulse train (or Dirac comb)

$$x(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT).$$

This is a special case of a periodic function, with period T. We can therefore apply a methodology similar to the one derived above to compute its Fourier transform. The Fourier series coefficients for the impulse train are obtained as

$$\hat{s}_k = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-2\pi i \frac{k}{T} t} dt = \frac{1}{T}.$$

Using (4.9), we get

$$\hat{x}(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} \delta\left(f - \frac{k}{T}\right). \tag{4.11}$$

Surprisingly, an impulse train in the time domain can be regarded as an impulse train in the frequency domain. A second representation for $x(t)$ is given by (4.10),

$$x(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT) = \frac{1}{T} \sum_{k=-\infty}^{\infty} e^{2\pi i \frac{k}{T} t}. \tag{4.12}$$

Which representation to use depends on the problem at hand.

4.2.3   Spectral Density

The energy of a deterministic signal $x(t)$ is given by (4.4). If the energy of $x(t)$ is finite, i.e. $\|x(t)\|^2 < \infty$, then we define its **autocorrelation function** by

$$R_x(\tau) \triangleq \int_{\mathbb{R}} x(t) x^*(t - \tau) dt.$$

Using this notation, we see that the energy of $x(t)$ is also given by $R_x(0)$.

Definition 4.2.3. *The **energy spectral density** of an energy-type signal $x(t)$, denoted by $\mathcal{G}_x(f)$, is defined to be the Fourier transform of its autocorrelation function,*

$$\mathcal{G}_x(f) = \mathcal{F}[R_x(\tau)] = |\hat{x}(f)|^2.$$

Intuitively, the energy spectral density captures the frequency content of a signal and helps identify how its energy is distributed across frequencies.

A signal $x(t)$ is a **power-type signal** if the limit

$$P_x = \lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt$$

exists and $0 < P_x < \infty$. The autocorrelation function of a power-type signal is defined accordingly as

$$R_x(\tau) \triangleq \lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)x^*(t - \tau)dt.$$

We note that the Fourier transform of a power-type signal $x(t)$ need not exist because it may fail to satisfy condition (4.4). Nevertheless, if truncated versions of $x(t)$, defined by

$$x_T(t) \triangleq x(t)\text{rect}\left(\frac{t}{T}\right),$$

are energy-type signals, then we can define the Fourier transforms

$$\hat{x}_T(f) \triangleq \mathcal{F}\left[x(t)\text{rect}\left(\frac{t}{T}\right)\right].$$

From this, the **power spectral density**, which represents the power (per unit of bandwidth) present at each frequency of the signal, can be defined by

$$\mathcal{S}_x(f) \triangleq \lim_{T \to \infty} \frac{1}{T}|\hat{x}_T(f)|^2.$$

Intuitively, the power spectral density captures the frequency content of a signal and helps identify how its power is distributed across frequencies.

Notice how the truncated signal is used to overcome the difficulty of dealing with infinite-energy signals. This is a common and valuable trick.

Definition 4.2.4. *The power spectral density of a power-type signal $x(t)$ is also given by the Fourier transform of its autocorrelation function,*

$$\mathcal{S}_x(f) = \mathcal{F}[R_x(\tau)].$$

These two definitions of power spectral density are equivalent under mild conditions on $x(t)$.

The **spectral bandwidth** of a signal $x(t)$ is the smallest value of W such that its spectral density is zero for all $|f| > W$. An energy-type signal $x(t)$

is **bandwidth-limited** to W if it can be obtained as the inverse Fourier transform of a function $\hat{x}(f)$, where $\hat{x}(f)$ is identically zero for all $|f| > W$. Likewise, a power-type signal is bandwidth-limited to W if its power-spectral density $S_x(f)$ is identically zero for all $|f| > W$.

4.2.4 Linear Time-Invariant Filters

The importance of the Fourier transform comes, partly, from its ability to capture the effects of linear time-invariant filters on deterministic signals. Suppose that the input to a linear time-invariant filter is $x(t)$, then its output is given by

$$y(t) = x(t) * h(t),$$

where $h(t)$ is the impulse response of the linear filter and $*$ denotes the convolution operator. If we use $\hat{h}(f)$ to represent the Fourier transform of impulse response $h(t)$, then the output signal in the frequency domain becomes

$$\hat{y}(f) = \hat{x}(f)\hat{h}(f).$$

That is, convolution in the time domain becomes multiplication in the frequency domain, a much simpler operation. The output signal can then be recovered by taking the inverse Fourier transform of $\hat{y}(f)$,

$$y(t) = \mathcal{F}^{-1}[\hat{y}(f)] = \mathcal{F}^{-1}[\hat{x}(f)\hat{h}(f)].$$

This also implies that the spectral density, $\mathcal{G}_y(f)$, of $y(t)$ satisfies

$$\mathcal{G}_y(f) = \mathcal{G}_x(f)\mathcal{G}_h(f).$$

4.3 Sampling Deterministic Signals

A great deal of inuition about sampling can be gained using the Fourier transform representation for periodic signals developed in Section 4.2.2. Let $x_s(t)$ denote the result of sampling $x(t)$ by impulses at times $\{nT : n \in \mathbb{Z}\}$,

$$x_s(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) = \sum_{n=-\infty}^{\infty} x(nT)\delta(t - nT).$$

Looking at the sampled signal in the frequency domain, we get

$$\hat{x}_{\mathrm{s}}(f) = \hat{x}(f) * \mathcal{F}\left[\sum_{n=-\infty}^{\infty} \delta(t - nT)\right] = \hat{x}(f) * \frac{1}{T}\sum_{n=-\infty}^{\infty}\delta\left(f - \frac{n}{T}\right)$$
$$= \frac{1}{T}\sum_{n=-\infty}^{\infty}\hat{x}(f) * \delta\left(f - \frac{n}{T}\right) = \frac{1}{T}\sum_{n=-\infty}^{\infty}\int_{\mathbb{R}}\hat{x}(\xi) * \delta\left(f - \xi - \frac{n}{T}\right)d\xi$$
$$= \frac{1}{T}\sum_{n=-\infty}^{\infty}\hat{x}\left(f - \frac{n}{T}\right),$$

where we have used (4.11) to express the Fourier transform of an impulse train. When the sampling rate is fast enough, the translated copies of $\hat{x}(f)$ contained in the transform $\hat{x}_{\mathrm{s}}(f)$ do not overlap, and the original signal can be recovered using an ideal lowpass filter. However, when the sampling period T is too small, the various copies of $\hat{x}(f)$ overlap and the content of the original is partially destroyed. This is know as **aliasing**.
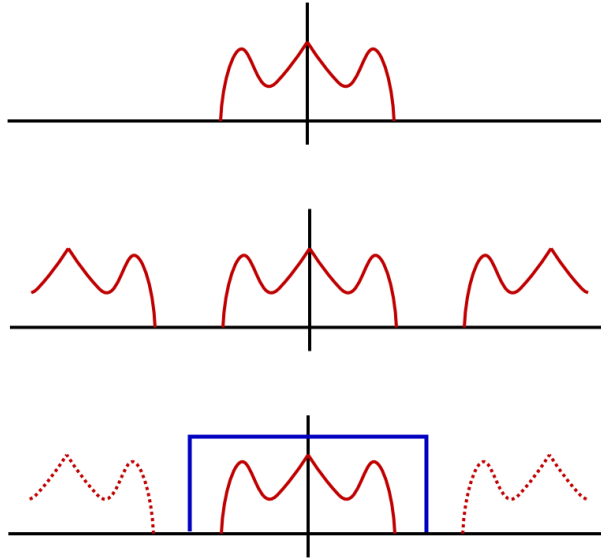

Figure 4.1: The sampling and reconstruction of a bandwidth-limited signal. When the sampling rate exceeds twice the bandwidth of the original signal, this signal can be reconstituted from its sampled values.

A succession of power spectral densities can be found in Figure 4.1. The top component shows the power spectral density of the original signal. The

density of the sampled signal appears below. Finally, the reconstruction operation where a lowpass filter is employed to recovered the original function is illustrated at the bottom of the figure. In contrast, Figure 4.2 exhibits a case where the sampling frequency is too low. Aliasing in the frequency domain
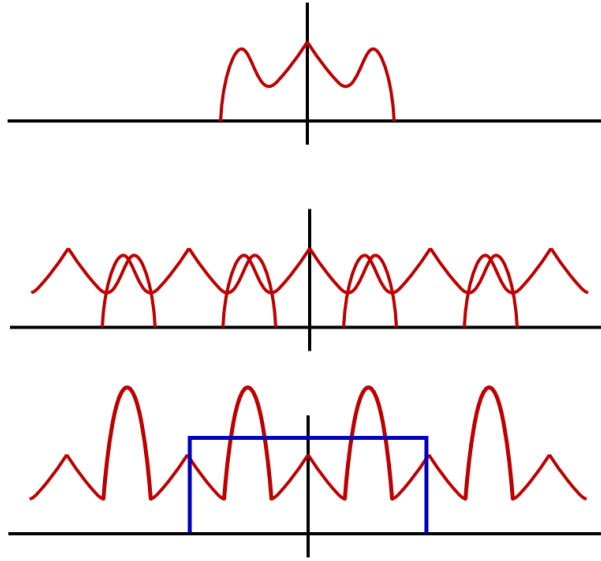

Figure 4.2: A low sampling frequency leads to aliasing, thereby preventing reconstruction of the original signal.

prevents the original signal from being retrieved.

Sampling and aliasing are also important in film and video. The illusion of a moving image in video is achieved by displaying a rapid succession of still pictures over time. Films are typically shot at a rate of twenty-four frames per second, whereas the minimum frame rate required to create the appearance of a moving image is about fifteen frames per second. The human eye acts as a lowpass filter and transforms the succession of images into a live video. High-speed cameras are used to record slow-motion playback movies. As a consequence, they must run at much higher frame-rates than normal cameras.

4.3.1   The Sampling Theorem

The sampling theorem is one of the most significant results in digital communication and signal processing. Many digital communication systems rely on

the validity of this theorem and on the design insights it provides for proper operation.

The basic idea behind the sampling theorem can be summarized in a few words. If a signal $x(t)$ is bandwidth-limited to W, then this signal can be reconstructed from a collection of samples so long as the samples are taken at periodic intervals of $T \leq \frac{1}{2W}$. A formal version of the sampling theorem appears below.

Theorem 4.3.1 (Sampling Theorem). *Let signal $x(t)$ be a bandwidth-limited function with bandwidth W. If $x(t)$ is sampled at times $\{nT : n \in \mathbb{Z}\}$ where $T \leq \frac{1}{2W}$, then it is possible to reconstruct the original signal $x(t)$ from its sampled points $\{x(nT) : n \in \mathbb{Z}\}$. Specifically, if $T \leq \frac{1}{2W}$ then*

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT)\mathrm{sinc}\left(\frac{t}{T} - n\right). \tag{4.13}$$

Proof. The signal $x(t)$ is bandwidth-limited with bandwidth W. It follows that $x(t)$ is the inverse Fourier transform of a function $\hat{x}(f)$, where $\hat{x}(f) = 0$ for all frequencies such that $|f| > W$. For convenience, we define $F = \frac{1}{T}$ and we stress that $W \leq \frac{1}{2T} = \frac{F}{2}$. Thus, $\hat{x}(f) = 0$ whenever $|f| > \frac{F}{2}$. We can apply the theory of Fourier series introduced in Section 4.1 to express $\hat{x}(f)$ as

$$\hat{x}(f) = \sum_{k=-\infty}^{\infty} s_k e^{2\pi i \frac{k}{F} f} \mathrm{rect}\left(\frac{f}{F}\right)$$

where the coefficients $\{s_k : k \in \mathbb{Z}\}$ are equal to

$$s_k = \frac{1}{F} \int_{-\frac{F}{2}}^{\frac{F}{2}} \hat{x}(f) e^{-2\pi i \frac{k}{F} f} df.$$

Special care should be taken when reading these equations because we are applying Fourier series analysis to a function in the frequency domain. This can get confusing.

We can then write $x(t)$ in terms of basis elements,

$$x(t) = \mathcal{F}^{-1}\left[\hat{x}(f)\right] = \mathcal{F}^{-1}\left[\sum_{k=-\infty}^{\infty} s_k e^{2\pi i \frac{k}{F} f} \text{rect}\left(\frac{f}{F}\right)\right]$$

$$= \sum_{k=-\infty}^{\infty} s_k \mathcal{F}^{-1}\left[e^{2\pi i \frac{k}{F} f} \text{rect}\left(\frac{f}{F}\right)\right]$$

$$= \sum_{k=-\infty}^{\infty} \frac{s_k}{T} \text{sinc}\left(\frac{t}{T} + k\right).$$

Above, we have successively used the scaling and time-shift properties of the Fourier transform. We can obtain the values of $\{s_k : k \in \mathbb{Z}\}$ explicitly by exploiting the characteristics of the $\text{sinc}(\cdot)$ function,

$$x(nT) = \sum_{k=-\infty}^{\infty} \frac{s_k}{T} \text{sinc}\left(\frac{nT}{T} + k\right) = \sum_{k=-\infty}^{\infty} \frac{s_k}{T} \text{sinc}(n + k) = \frac{s_{-n}}{T}.$$

Thus, we have $s_n = Tx(-nT)$ and formula (4.13) follows. The sampling rate $F = 2W$ associated with the sampling period $T = \frac{1}{2W}$ is the minimum rate at which perfect reconstruction is possible. It is called the **Nyquist rate** in honor of Swedish-American engineer Harry Nyquist. $\qquad\square$

4.3.2   Imperfect Sampling and Reconstruction

In practice, it is impossible to measure (i.e., sample) a signal $x(t)$ instantaneously. A more realistic model is to assume that the sample value is obtained through the integral

$$u(t) = \int_{\mathbb{R}} x(\tau) p(\tau - t) d\tau,$$

for some sampling waveform $p(t)$. In this case, the resulting samples are identical to the perfect sampling of the filtered waveform $u(t) = x(t) * p(-t)$. That is, the sample values are given by

$$u(nT) = \int_{\mathbb{R}} x(\tau) p(\tau - nT) d\tau.$$

It also follows that, if no aliasing occurs, this degradation can be eliminated completely using a discrete-time filter to equalize the resulting samples.

A similar imperfection occurs during reconstruction. In practice, it is not possible to weight each sample by the exact sinc interpolation waveform. Instead, the samples $x(nT)$ are weighted by a pulse shape $q(t)$. In this case, the reconstruction output is given by

$$y(t) = \sum_{n=-\infty}^{\infty} x(nT)q(t-nT).$$

Since $q(t-nT) = \delta(t-nT) * q(t)$, we can see this instead as perfect reconstruction followed by filtering and write

$$y(t) = \left(\sum_{n=-\infty}^{\infty} x(nT)\delta(t-nT) \right) * q(t) = x(t) * q(t).$$

Again, it follows that this imperfection can be eliminated completely by post-filtering. In practice, the main advantage of this observation is that one can jointly optimize $p(t)$, $q(t)$ and the post-filter to provide good performance while using inexpensive components.

4.4 Sampling Bandlimited Processes*

We know from Theorem 4.3.1 that a bandwidth-limited signal can be perfectly reconstructed from its samples provided that the sampling rate exceeds twice the bandwidth of the original signal. At this point, one may wonder whether it is possible to extend the sampling theorem to bandwidth-limited stochastic processes. This question is answered in the affirmative below.

Theorem 4.4.1. *Suppose that $X(t)$ is a wide-sense stationary bandwidth-limited process with bandwidth W and power spectral density $\mathcal{S}_X(f)$. Let $\tilde{X}(t)$ be an approximation for $X(t)$ built from the sampled values $\{X(nT) : n \in \mathbb{Z}\}$,*

$$\tilde{X}(t) = \sum_{n=-\infty}^{\infty} X(nT)\mathrm{sinc}(2W(t-nT)),$$

where $T = \frac{1}{2W}$ denotes the sampling interval. Then the mean-squared error between the original random process and the reconstructed version vanishes,

$$\left\| X(t) - \tilde{X}(t) \right\|^2 = \mathrm{E}\left[\left| X(t) - \sum_{n=-\infty}^{\infty} X(nT)\mathrm{sinc}(2W(t-nT)) \right|^2 \right] = 0.$$

$$(4.14)$$

The expectation in (4.14) *is over all possible realizations of* $X(t)$.

Proof. To establish this result, we expand the mean-squared error of (4.14),

$$\left\| X(t) - \tilde{X}(t) \right\|^2 = \mathrm{E}\left[\left| X(t) - \sum_{n=-\infty}^{\infty} X(nT)\mathrm{sinc}(2W(t - nT)) \right|^2 \right]$$

$$= R_X(0) - \sum_{n=-\infty}^{\infty} [R_X(t - nT) + R_X(t - nT)]\mathrm{sinc}(2W(t - nT))$$

$$+ \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} R_X((m - n)T)\mathrm{sinc}(2W(t - mT))\mathrm{sinc}(2W(t - nT)).$$

The double summation above can be rewritten as

$$\sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} R_X(kT)\mathrm{sinc}(2W(t - kT - nT))\mathrm{sinc}(2W(t - nT))$$

$$\sum_{n=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} R_X(kT)\mathrm{sinc}(2W(t - kT - nT)) \right) \mathrm{sinc}(2W(t - nT))$$

$$= \sum_{n=-\infty}^{\infty} R_X(t - nT)\mathrm{sinc}(2W(t - nT)),$$

where the last equality follows from the sampling theorem for deterministic signals (Theorem 4.3.1). Putting these results together, we get

$$\left\| X(t) - \tilde{X}(t) \right\|^2 = R_X(0) - \sum_{n=-\infty}^{\infty} R_X^*(t - nT)\mathrm{sinc}(2W(t - nT)).$$

Applying Theorem 4.3.1 one more time and noticing that $R_X(0) = R_X^*(0)$, we obtain $\|X(t) - \tilde{X}(t)\|^2 = 0$, as desired.   □

Theorem 4.4.1 is important because it confirms that the design insights gained from analyzing deterministic signals hold for random signals as well.

4.5   Bandpass Signals and Processes*

One possible application of sampling is to take a continuous-time signal and to transform it into a discrete-time signal. For instance, this operation gives the

information coming out of a source a format more suitable for digital communications. This prime application of sampling served as the original motivation for our study of the subject. A second possible application of sampling is the processing of received waveforms at the output of communication channels. In digital communications, the data often assumes the form of an analog carrier signal modulated by a digital bit stream. Mathematically, this situation is captured by the equation

$$y(t) = x(t) \cos(2\pi f_c t).$$

The signal $y(t)$ is a special form of a **bandpass signal**. Its Fourier transform $\hat{y}(f)$ is non-zero only for frequencies contained in a small neighborhood of carrier frequency f_c. That is, $\hat{y}(f) = 0$ for all frequencies such that $|f - f_c| \geq W$. To apply the sampling tools derived above to the information bearing signal $x(t)$, we need to shift the corresponding spectrum to the origin.

The Fourier transform of $y(t)$ is given by

$$\hat{y}(f) = \frac{1}{2}\hat{x}(f + f_c) + \frac{1}{2}\hat{x}(f - f_c).$$

Our strategy is to first eliminate $\frac{1}{2}\hat{x}(f + f_c)$ from $\hat{y}(f)$, and then to scale and shift $\frac{1}{2}\hat{x}(f + f_c)$ back to the origin. Define the **step function** by

$$\text{step}(t) = \frac{1}{2} + \frac{1}{2}\text{sign}(t).$$

Taking the (cavalier) Fourier transform of $\text{step}(t)$, we get

$$\mathcal{F}[\text{step}(t)] = \mathcal{F}\left[\frac{1}{2} + \frac{1}{2}\text{sign}(t)\right]$$
$$= \frac{1}{2}\delta(f) - \frac{1}{2}\int_{-\infty}^{0} e^{-2\pi i f t} dt + \frac{1}{2}\int_{0}^{\infty} e^{-2\pi i f t} dt$$
$$= \frac{1}{2}\delta(f) + \frac{1}{2\pi i f}.$$

Using the duality property of the Fourier transform, we get

$$\mathcal{F}^{-1}[\text{step}(f)] = \frac{1}{2}\delta(t) + \frac{i}{2\pi t}.$$

And, by construction, we obtain $\hat{x}(f - f_c) = 2\text{step}(f)\hat{y}(f)$. We can therefore recover the original lowpass signal $x(t)$ using the frequency-shift property of

the Fourier transform,

$$x(t) = \left[y(t) * \left(\delta(t) + \frac{i}{\pi t} \right) \right] e^{-2\pi i f_c t} = \left[y(t) + i \left(y(t) * \frac{1}{\pi t} \right) \right] e^{-2\pi i f_c t}.$$

The second component of this signal,

$$y(t) * \frac{1}{\pi t},$$

is called the **Hilbert transform** of $y(t)$. Once $x(t)$ is brought back to baseband, the standard sampling theorem applies and a discrete-time version of the signal can be produced.

4.6   Stochastic Signals

A **random process** (or **stochastic process**) is an extension of the concept of random variable to the situation where the values of a signal are not known beforehand. Mathematically, a stochastic process can be viewed in two different ways. First, the process can be thought of as an instantiation of a random experiment where the outcome is selected from a collection of time functions. Alternatively, a stochastic process can be viewed as a collection of random variables indexed by time. If the index set corresponds to the real numbers, then the process is a **continuous-time random process**. Whereas if the index set is discrete, then it is a **discrete-time random process**. The viewpoint where a stochastic process is regarded as a collection of random variables tends to prevail in the study of digital communications.

Random processes are frequently employed in the design of communication systems. For example, they can be used to model the data originating from a source, channel variations, noise and interference. Their importance will become evident as we progress through these notes. In general, it is difficult to provide a complete mathematical description for a random process. For now, we restrict our attention to **stationary** and **ergodic** random processes.

Definition 4.6.1 (Stationarity)**.** *A random process $X(t)$ is wide-sense stationary (WSS) if its **mean***
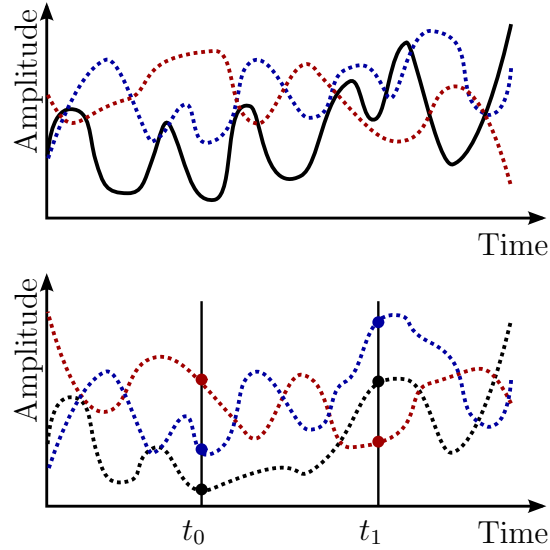
$$m_X(t) \triangleq \mathrm{E}[X(t)]$$

Figure 4.3: Two distinct abstractions of a random process. It can be viewed as the output of an experiment where a function is selected at random. Alternatively, a random process may be taken as a set of random variables indexed by time.

*is independent of time, and its **autocorrelation function**, defined by*

$$R_X(t_1, t_2) \triangleq \mathrm{E}[X(t_1)X^*(t_2)],$$

only depends on the difference between t_1 and t_2. With a slight abuse of notation, we can denote the mean and autocorrelation of a stationary process respectively by m_X and $R_X(\tau)$, where $\tau = t_1 - t_2$.

Definition 4.6.2 (Ergodic). *An **ergodic theorem** asserts that, under certain conditions, the time average of a random process*

$$\langle g(X(t)) \rangle \triangleq \lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} g(X(t))dt$$

*exists and is equal to the ensemble average $\mathrm{E}[g(X(t))]$ for almost all trajectories of the random process. When a stochastic process fulfills these conditions, it is called **ergodic**.*

One of the important characteristics of an ergodic process is that it suffices to look at one realization of the process to infer many of its statistical
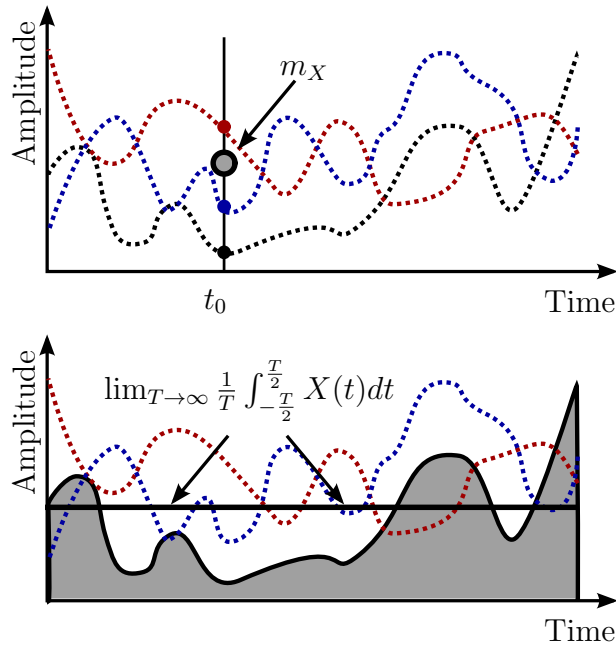
Figure 4.4: For ergodic processes, the time average of a function along a trajectory is equal to the ensemble average.

attributes. Ergodicity is a very strong property, and it is hard to test and validate. Rather, it is frequently taken as a premise in the design of communication systems. For instance, most information sources are assumed to be stationary and ergodic. Such a postulate appears reasonable, especially considering the many successful communication systems implemented based on this assumption.

4.6.1   Power Spectral Density

For a stochastic signal, the definition of the **power spectral density** is somewhat more intricate because it must account for uncertainty in the process. Let $X(t)$ be a wide-sense stationary random process with $R_X(0) < \infty$. Then, we can define

$$\hat{X}_T(f) \triangleq \mathcal{F}\left[X(t)\mathrm{rect}\left(\frac{t}{T}\right)\right]$$

and

$$\mathcal{S}_X(f) \triangleq \lim_{T \to \infty} \frac{1}{T}\mathrm{E}\left[|\hat{X}_T(f)|^2\right].$$

As we will soon see, the power spectral density plays an instrumental role in the sampling theorem for random signals. First, we provide a means to compute $\mathcal{S}_X(f)$ from its statistical attributes.

Theorem 4.6.3 (Wiener-Khinchin). *The power spectral density $\mathcal{S}_X(f)$ of a wide-sense stationary random process $X(t)$ is equal to the Fourier transform of its autocorrelation function, $\mathcal{S}_X(f) = \mathcal{F}[R_X(\tau)]$.*

Proof. For a wide-sense stationary process, we have

$$
\begin{aligned}
\mathcal{S}_X(f) &= \lim_{T\to\infty} \frac{1}{T} \mathrm{E}\left[|\hat{X}_T(f)|^2\right] = \lim_{T\to\infty} \frac{1}{T} \mathrm{E}\left[\hat{X}_T(f)\hat{X}_T^*(f)\right] \\
&= \lim_{T\to\infty} \frac{1}{T} \mathrm{E}\left[\int_{-\frac{T}{2}}^{\frac{T}{2}} X(t_1)e^{-2\pi i f t_1}dt_1 \int_{-\frac{T}{2}}^{\frac{T}{2}} X^*(t_2)e^{2\pi i f t_2}dt_2\right] \\
&= \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{-\frac{T}{2}}^{\frac{T}{2}} \mathrm{E}\left[X(t_1)X^*(t_2)\right] e^{-2\pi i f(t_1-t_2)}dt_1 dt_2 \\
&= \int_{\mathbb{R}} R_X(\tau)e^{-2\pi i f\tau}d\tau = \mathcal{F}[R_X(\tau)].
\end{aligned}
$$

The fourth equality is obtained by interchanging the expectation and the integrals, while the sixth equality follows from a change of variables and the fact that $X(t)$ is wide-sense stationary. To guarantee that the former operation is legitimate, $\tau R_X(\tau)$ must remain finite for all τ. $\square$

In some cases, it may be possible to estimate the power spectral density from a single realization of $X(t)$. For example, the autocorrelation of $X(t)$ is ergodic if it satisfies

$$
\langle X(t)X^*(t-\tau)\rangle \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X(t)X^*(t-\tau)dt = \mathrm{E}\left[X(t)X^*(t-\tau)\right].
$$

In this case, the expectation in the theorem above can be replaced by a time average and the result can be computed from a single realization.

4.6.2 Filtering Stochastic Processes

We discussed in Section 4.2.4 how the Fourier transform can simplify the analysis of the effects of linear time-invariant filters on deterministic signals. In this section, we consider the operation of such filters in the context of random processes.

Theorem 4.6.4. *If a wide-sense stationary process $X(t)$ with mean m_X and autocorrelation function $R_X(\tau)$ is passed through a linear time-invariant filter with impulse response $h(t)$, then the output process $Y(t)$ has mean*

$$m_Y = m_X \int_{\mathbb{R}} h(t)dt$$

and its autocorrelation is equal to

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h^*(-\tau).$$

Proof. The output process at time t is given by $Y(t) = \int_{\mathbb{R}} X(t-\xi)h(\xi)d\xi$. We can therefore obtain the expectation of $Y(t)$ as follows,

$$m_Y(t) = \mathrm{E}\left[\int_{\mathbb{R}} X(t-\xi)h(\xi)d\xi\right] = \int_{\mathbb{R}} \mathrm{E}\left[X(t-\xi)\right]h(\xi)d\xi$$

$$= m_X \int_{\mathbb{R}} h(\xi)d\xi.$$

We emphasize that m_Y is independent of time.

To derive the autocorrelation function for $Y(t)$, we first compute the cross-correlation between $X(t)$ and $Y(t)$,

$$
\begin{aligned}
\mathrm{E}[X(t_1)Y^*(t_2)] &= \mathrm{E}\left[X(t_1)\int_{\mathbb{R}} X^*(\xi)h^*(t_2-\xi)d\xi\right] \\
&= \int_{\mathbb{R}} \mathrm{E}\left[X(t_1)X^*(\xi)\right]h^*(t_2-\xi)d\xi \\
&= \int_{\mathbb{R}} R_X(t_1-\xi)h^*(t_2-\xi)d\xi \\
&= R_X(\tau) * h^*(-\tau).
\end{aligned}
\tag{4.15}
$$

This shows that the cross-correlation between $X(t)$ and $Y(t)$ depends only on τ; we can therefore express it as $R_{XY}(\tau)$. We are now ready to compute the autocorrelation function for $Y(t)$.

$$
\begin{aligned}
\mathrm{E}[Y(t_1)Y^*(t_2)] &= \mathrm{E}\left[\int_{\mathbb{R}} X(\xi)h(t_1-\xi)d\xi Y^*(t_2)\right] \\
&= \int_{\mathbb{R}} \mathrm{E}\left[X(\xi)Y^*(t_2)\right]h(t_1-\xi)d\xi \\
&= \int_{\mathbb{R}} R_{XY}(\xi-t_2)h(t_1-\xi)d\xi \\
&= R_{XY}(\tau) * h(\tau).
\end{aligned}
\tag{4.16}
$$

Substituting $R_{XY}(\tau)$ by the equivalent expression $R_X(\tau)*h^*(-\tau)$ from (4.15), we get the desired result. We observe that the autocorrelation of the process $Y(t)$ only depends on the difference between t_1 and t_2, and hence $Y(t)$ is also wide-sense stationary. □

Obtaining an expression for the autocorrelation function corresponding to the output of a linear time-invariant filter allows us to characterize the power spectral density of the output process. In terms of the frequency representation, we get $m_Y = m_X \hat{h}(0)$ and

$$
\begin{aligned}
\mathcal{S}_Y(f) &= \mathcal{F}[R_Y(\tau)] \\
&= \mathcal{F}\left[R_X(\tau) * h(\tau) * h^*(-\tau)\right] \\
&= \mathcal{S}_X(f)|\hat{h}(f)|^2.
\end{aligned}
$$

A linear time-invariant filter can be employed to shape the spectrum of a stochastic process, and to constrain its bandwidth. This is an important result, as linear filters can be used to reduce the bandwidth of a random signal before sampling or to reconstruct a random signal from its samples.

4.6.3 Gaussian Processes

A stochastic process $X(t)$ is a **Gaussian process** if, for any finite set of sampling times $t_1, t_2, \ldots, t_n$, the resulting random variables $X(t_1), X(t_2), \ldots, X(t_n)$ are jointly Gaussian random variables. The distribution of jointly Gaussian random variables is completely determined by their mean vector and correlation matrix. Therefore, a Gaussian process is completely determined by its mean function $m_X(t)$ and its autocorrelation function $R_{XX}(t_1, t_2)$. Of course, if the process is also wide-sense stationary, then this reduces to its mean value m_X and its autocorrelation function $R_X(\tau)$.

More generally, one finds that, for any finite set of energy-type signals $s_1(t), s_2(t), \ldots, s_n(t)$, the random variables

$$
Z_i = \int_{-\infty}^{\infty} s(t)X(t)dt
$$

are jointly Gaussian. The intuition behind this is that each integral is defined as the limit of a sequence of Riemann sums. Since each Riemann sum

is simply the sum of Gaussian random variables, it is also Gaussian. This argument can be extended to show that the mean vector and correlation matrix of the vector $(Z_1, Z_2, \ldots, Z_n)$ is also jointly Gaussian. In fact, combining this with Theorem 4.6.4, implies that the output of linear time-invariant system, whose input is a wide-sense stationary Gaussian process, is also a wide-sense stationary Gaussian process.

The most common random process in signal processing and communications is the **Gaussian white-noise** process $N(t)$. This process is defined to be a zero-mean wide-sense stationary Gaussian process whose power spectral density $S_N(f)$ is constant (e.g., 1). Since this process has infinite power, it is not particularly well-defined mathematically. Still, we can use the cavalier Fourier transform $\mathcal{F}[1] = \delta(t)$ to argue that the autocorrelation function of $N(t)$ should be $R_N(\tau) = \delta(\tau)$. In practice, this is not a problem because $N(t)$ is only used as the input to a linear filter whose output process is guaranteed to have finite power.