Transmission Fundamentals

Handouts: Course information and tentative schedule.

- Signals
- Data rate and bandwidth
  - Nyquist sampling theorem
  - Shannon capacity theorem

1 Signals

Any electromagnetic signal (for that matter, essentially any function $f(t)$) can be expressed as a superposition of sinusoidal waves. A general sinusoidal signal is given by the function $s(t) = A\sin(2\pi ft + \phi)$, where A is the amplitude (usually in volts), f is the frequency, and ϕ is the phase. A more accurate way of stating this is using Fourier analysis and Fourier transforms, as follows.

$$\begin{aligned} s(t) &= \int_{-\infty}^{+\infty} S(f)e^{i2\pi ft}df \\ S(f) &= \int_{-\infty}^{+\infty} s(t)e^{-i2\pi ft}dt. \end{aligned}$$

The function $S(f)$ is referred to as the Fourier transform of $s(t)$, and captures how the energy of the signal $s(t)$ is distributed within the frequency spectrum.

Studying signals in the frequency domain greatly faciliates the study of how signals are transmitted. A channel takes as input an input signal $s(t)$ and outputs a signal $r(t)$. A channel can be typically described by its impulse response function $h(t)$ which is the channel output corresponding to an infinitesimally narrow pulse of unit area at time t. Given an input $s(t)$ to a channel with impulse response $h(t)$, the output $r(t)$ is given by

$$r(t) = \int_{-\infty}^{+\infty} s(\tau)h(t-\tau)d\tau.$$

Using the Fourier transform equations, we obtain:

$$\begin{aligned} r(t) &= \int_{-\infty}^{+\infty} \left(h(t-\tau)\int_{-\infty}^{+\infty} S(f)e^{i2\pi f\tau}df\right)d\tau \\ &= \int_{\infty}^{+\infty} \left(S(f)\int_{-\infty}^{+\infty} h(t-\tau)e^{i2\pi f\tau}d\tau\right)df \end{aligned}$$

$$
\begin{aligned}
&= \int_{\infty}^{+\infty} S(f)e^{i2\pi ft}\left(h(t-\tau)e^{-i2\pi f(t-\tau)}d\tau\right)df \\
&= \int_{\infty}^{+\infty} S(f)e^{i2\pi ft}\left(h(\tau)e^{-i2\pi f\tau)}d\tau\right)df \\
&= \int_{\infty}^{+\infty} S(f)H(f)e^{i2\pi ft}df.
\end{aligned}
$$

If $R(f)$ is the Fourier transform of $r(t)$, then we obtain the following simple input-output relation in the frequency domain

$$R(f) = S(f)H(f).$$

The absolute bandwidth of the signal $s(t)$ (resp., a channel with impulse response $h(t)$) is the range of frequencies f in which $S(f) \neq 0$ (resp., $H(f) \neq 0$). The effective bandwidth, or simply, the bandwidth of the signal is the range within which most of the energy is concentrated. The energy of a signal $s(t)$ (per unit resistance) is given by

$$\int_{-\infty}^{+\infty} s(t)^2 dt.$$

It can be shown that any periodic signal $s(t)$ with frequency f has non-zero frequency components only at frequencies that are multiples of f. That is, a periodic signal $s(t)$ with frequency f can be expressed as:

$$s(t) = \sum_{n=-\infty}^{\infty} S_n e^{2\pi inft}.$$

A periodic square wave with amplitude A and -A can be expressed as follows.

$$s(t) = A \times \frac{4}{\pi} \sum_{\text{odd } k} \frac{sin(2\pi kft)}{k}.$$

2 Data rate and bandwidth

Suppose we superpose sinusoids of frequencies f, $3f$, and $5f$, then we get a wave that resembles a square wave with 3 peaks per half cycle. The bandwidth of the signal is $5f - f = 4f$. In a time interval of $1/f$ seconds, two bits have been transmitted, achieving a bit rate of $2f$. If we double the frequency f, this doubles the bandwidth and hence also doubles the bit rate. If we reduce the bandwidth, say by using only sinusoids f and $3f$, then we can achieve the same bit rate, but the signal has a higher chance of being misinterpreted since it is "farther" from the desired square wave (and hence can cause more errors).

In general, the question relating bit rate and channel bandwidth is not about what bit rate can a given channel bandwidth support since over any finite length of time, there is a non-zero probability that every bit transmitted is in error. Consequently, any discussion of capacity of a channel has to be done together with the notion of an error rate. Shannon framed the capacity question as: given a particular channel bandwidth, what is the maximum rate at which information can be sent, for a given error rate? Shannon's channel coding theorem gives a surprising answer to this question, which implies that the maximum information rate achievable is independent of the error rate.

Theorem 1 (Nyquist Sampling Theorem) *[Sta05, Section 6A] Any signal $s(t)$ that contains no frequency component more than W Hz is completely determined by its values at times $j/2W$, where j is an integer in the range $[-\infty, +\infty]$.*

Proof: Consider a periodic pulse function $p(t)$ that consists of a unit pulse (Dirac delta function) every $T = 1/2W$ seconds. Since any periodic function can be expressed as a sum of sinusoidals with frequencies that are multiples of the function, we have

$$p(t) = \sum_{n=-\infty}^{\infty} P_n e^{i(2\pi n/T)t}$$

for some appropriate values of P_n. Let the input signal be $s(t)$ with Fourier transform $S(f)$. The sampled signal $x(t)$ can be written as

$$x(t) = s(t)p(t) = \sum_{n=-\infty}^{\infty} P_n s(t) e^{i(2\pi n/T)t}.$$

Therefore, the Fourier transform of the sampled signal is

$$\begin{aligned} X(f) &= \int_{-\infty}^{+\infty} \sum_{n=-\infty}^{\infty} P_n x(t) e^{i(2\pi n/T)t} e^{-i2\pi f t} dt \\ &= \frac{1}{T} \int_{-\infty}^{+\infty} \sum_{n=-\infty}^{\infty} P_n x(t) e^{-i2\pi(f-n/T)} dt. \end{aligned}$$

We also have

$$S(f - n/T) = \int_{-\infty}^{+\infty} x(t) e^{-i2\pi(f-n/T)} dt.$$

Therefore, we obtain

$$X(f) = \sum_{-\infty}^{+\infty} P_n S(f - n/T)$$

We know that $S(f) = 0$ for $|f| \geq W$. Since $1/T = 2W$, we obtain that for $-W \leq f \leq W$, $X(f) = P_0 S(f)$. Thus the frequency spectrum of s can be entirely determined by the frequency spectrum of the sampled signal x. ∎

There are several different proofs of the Nyquist sampling theorem. One equality that directly captures the equivalence of the band-limited signal s and the sample x is the following.

$$s(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin(\pi(2Wt - n))}{\pi(2Wt - n)} \tag{1}$$

Theorem 2 (Shannon Capacity Theorem) *[Sha49] Let P be the average transmitted power, and suppose the noise of white thermal noise of power N in the frequency band W. Then, there exists an encoding scheme that allows to transmit bits at a rate*

$$C = W \log_2 \frac{P + N}{N}$$

with as small an error probability as desired. It is not possible by any encoding method to send at a highed rate and have an arbitrarily low frequency of errors.

Proof Sketch: We consider a signal $s(t)$ over an interval of period T. Using Nyquist sampling, we specify the signal using $2TW$ samples x_1 through x_{2TW}. Using Equation 1 and the energy formula, one can show that the energy E (into unit resistance) of the signal $s(t)$ equals

$$\int_{-\infty}^{\infty} s(t)^2 dt = \frac{1}{2W} \sum_{n=1}^{2TW} x_n^2.$$

So if d is the distance of the point $(x_1, x_2, \ldots, x_n)$ from the origin, then we have

$$d^2 = 2WE = 2WPT,$$

where P is the average signal power during the period of length T. Thus any signal over time period T with average power at most P can be represented by a point in a ball of radius $\sqrt{2WPT}$ in $2WT$-dimensional space.

We now consider the effect of additive white Gaussian noise (AWGN) on this signal. An AWGN signal has the property that each sample is perturbed independently of all others and the distribution of the perturbation is Gaussian with standard deviation $\sqrt{N}$ where N is the average noise power.

It can be shown that for any desired error probability ε, if T is sufficiently large, then the noise causes the perturbation to lie within a ball of radius $\sqrt{2WNT}$ from the signal point $(x_1, \ldots, x_{2WT})$ with probability $1 - \varepsilon$.

When transmitting the information, we need to make sure that two signals carrying different information should not be decoded as the same. So we need to select signal points such that the balls of radius $\sqrt{2WNT}$ around them do not intersect. How many such signal points can we select? The answer is essentially the ratio of the volume of the ball of radius $\sqrt{2WT(N+P)}$ to that of $\sqrt{2WNT}$. For $2WT$-dimensional space, this ratio equals

$$M = \left(\frac{\sqrt{2WT(N+P)}}{\sqrt{2WNT}} \right)^{2TW} = \left(\frac{P+N}{N} \right)^{TW}.$$

It is fairly easy to see, using a ball-packing argument, that the right-hand-side is an upper bound on M. To see the lower bound, Shannon gave a much more involved argument. It relies on random coding, in which we map each of M pieces of an information to a random point in the ball of radius $\sqrt{2WNT}$. Shannon then goes on to show that the probability that any two balls overlap can be made arbitrarily close to zero by setting T sufficiently high.

The term M is the number of different pieces of information we can send in T time units. So the information rate is given by

$$C = \frac{\log_2 M}{T} = W \log_2 \frac{P+N}{N}.$$

■

3 Analog and digital data, signals, and transmissions

Simply put, the term *analog* refers to "continuous" functions, over time, while *digital* refers to "discrete" functions over time. Examples of analog data include voice, video, and graphics, while examples of digital data include text and stored images. Analog signals are continuously varying signals; unguided wireless media transmit analog signals. A digital signal is a sequence of voltage pulses; it is generally cheaper and less susceptible to noise, though suffers more attenuation than analog signals.

We can have either analog or digital data transmitted using analog or digital signals. The process of converting data into signals that can be carried by a given transmission medium is called *modulation* and the reverse process is called *demodulation*. We briefly discuss the four possibilities here; these will be elaborated in more detail in the following weeks.

Analog-to-analog. A canonical example of transmitting analog data using analog signals is the telephone. In this case, the frequency spectrum of the underlying data (voice) matches that of the carrier (telephone wire) – this is referred to as baseband transmission. For several other applications (e.g., cellular telephony), one needs to modulate the analog data so that the frequency spectrum of the resulting signal matches that of the carrier.

Digital-to-analog. When digital data are sent using analog signals, the discrete data is encoded in terms of the carrier signals of the underlying transmission medium. Some transmission media such as wireless only admit analog signals, so transmitting digital content through them requires digital-to-analog conversion. Another example is digital data over telephone wires.

Analog-to-digital. When analog data are transmitted using digital signals, the data is essentially digitized to obtain a bit stream, which is transmitted using a digital signal. Digitization may be done even in cases where the final transmission medium admits analog signals; e.g., digitzation of voice signals prior to transmission over wireless media, to improve quality.

Digital-to-digital. Finally, when digital data is transmitted using digital signals, a simple encoding scheme can be used that essentially represents the two bit values using two different voltage levels. There are other encoding schemes that are used when the signal needs to satisfy some other properties (e.g., synchronization) which are not satisfied by this simple scheme.

References

[Sha49] C. Shannon. Communication in the presence of noise. *Proc. Inst. of Radio Engineers*, 37:10–21, 1949.

[Sta05] William Stallings. *Wireless Communications and Networks.* Prentice Hall Professional Technical Reference, 2005. 2nd edition.