Data Conversion

Topics: A/D and D/A conversion, sampling, aliasing, reconstruction, filters

DSP systems usually have one or both types of data conversion: from analog to digital (A/D) and from digital to analog (D/A). A system with both is shown in Figure 1:

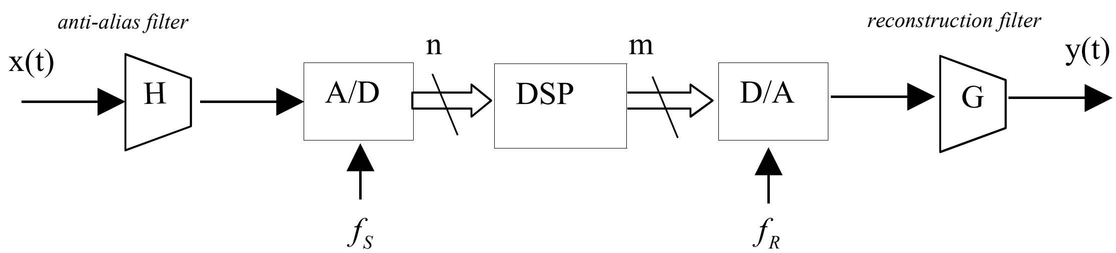

Figure 1: A DSP system having both analog input and analog output. Here, f_S is the rate at which samples are taken for A/D conversion, and f_R is the rate of D/A conversion.

Note that there are purely digital systems which do not use an analog input, and there are digital output DSP systems.

It is important to understand the data conversion process for DSP. In fact, it is sometimes said that DSP as a subject began when Harry Nyquist established the importance of the sample rate f_S in processing analog signals. Nyquist, in 1928, demonstrated a limitation inherent in sampling: no input frequency equal to or greater than $f_S/2$ can be completely reconstructed from its samples[1]. Therefore, whatever information the input signal carries at frequencies greater than or equal to $f_S/2$ cannot be recovered. Therefore, the sampling rate of a DSP system must exceed $2f_{MAX}$, where f_{MAX} is the maximum frequency in the input signal. Today, the value of $f_S/2$ is called the *Nyquist frequency* or simply *Nyquist*, and the value $2f_{MAX}$ is called the *Nyquist rate*. Only frequencies below Nyquist frequency are unambiguously preserved after sampling, and in order to prevent loss of information, the sampling rate must be greater than the Nyquist rate. We now examine the basics of sampling which explain Nyquist's results.

Sampling theory

Suppose that the input signal is a single sinusoid: $x(t) = A\cos(2\pi ft + \theta)$. If this signal is sampled at a rate of f_S, then the samples are spaced $t_S = 1/f_S$ seconds apart. The n-th sample is

[1] Nyquist's 1928 paper is called "Certain topics in telegraph transmission theory". Note that the telegraph, which uses only short and long pulse or "dots and dashes", is a digital transmission system.

$$x[n] = x(nt_S) = A\cos(2\pi f n t_S + \theta) = A\cos\left(2\pi \frac{f}{f_S} n + \theta\right)$$

There are three simple but important observations we can make about this expression. First, the original frequency f is replaced by the ratio of frequencies f / f_S after sampling. Let

$$F = \frac{f}{f_S}$$

If f is in units of cycles/second (Hz), and f_S is the number of samples per second (Sa/s), then the units of the normalized frequency F are

$$\frac{\text{cycles}}{\text{second}} \times \left(\frac{\text{samples}}{\text{second}}\right)^{-1} = \frac{\text{cycles}}{\text{sample}}$$

Second, because a sinusoid is periodic, i.e., $\cos(\psi + 2\pi) = \cos(\psi)$ for any ψ, there is a potential for two different normalized frequencies to produce the same samples. Speci-fically, note that

$$x[n] = A\cos\left(2\pi Fn + \theta\right) = A\cos(2\pi Fn + \theta + 2\pi) = A\cos(2\pi(F+1)n + \theta)$$

Therefore, the two normalized frequencies F and $F+1$ cycles/sample are exactly the sample. For example, $F = 0$ cycles/sample is actually a constant signal. However, so is $F = 1$ cycles/sample (repeating every sample). This implies that the two different input frequencies $f = Ff_S$ and $f_1 = (F+1)f_S = f + f_S$ produce identical samples, and therefore cannot be distinguished from one another after sampling. For example, if $f_S = 1000$ Sa/s then the two frequencies 100 Hz and 1,100 Hz produce exactly the same samples. This is called *aliasing*: two different input signals appear the same if their frequencies are spaced apart by the sampling frequency. We can also say that 1,100 Hz is *aliased* to 100 Hz, or that 100 Hz is an *alias* of 1,000 Hz.

Third, because the cosine function is symmetric, i.e., $\cos(\psi) = \cos(2\pi - \psi)$, there is another ambiguity possible in sampling. Specifically,

$$x[n] = A\cos\left(2\pi Fn + \theta\right) = A\cos(2\pi - 2\pi Fn - \theta) = A\cos(2\pi(1-F)n - \theta)$$

This relationship implies that a sinuosoid with frequency F and phase θ is indistin-guishable from a sinusoid with frequency 1-F and phase $-\theta$. For example, if $f_S = 1000$ Sa/s then frequencies 100 Hz and 900 Hz are indistinguishable if the latter has the negative of phase of the former. This is called *folding*: two different input signals appear to be same if their frequencies add up to the sampling frequency *and* their phases are negatives of each other. We can say that 900 Hz *folds* to 100 Hz with negative phase.

With either sampling or aliasing, it is important to realize that we can never tell whether they occurred: if $f_S = 1000$ Sa/s and we observe a signal with $F = 0.1$ and phase $\theta = \pi/2$, then we *cannot* say whether this is truly an input of $f = 100$ Hz, or $f = 1100$ Hz with the same phase (aliasing), or $f = 900$ Hz with phase $\theta = -\pi/2$ (folding). The situation is summarized in Figure 2.

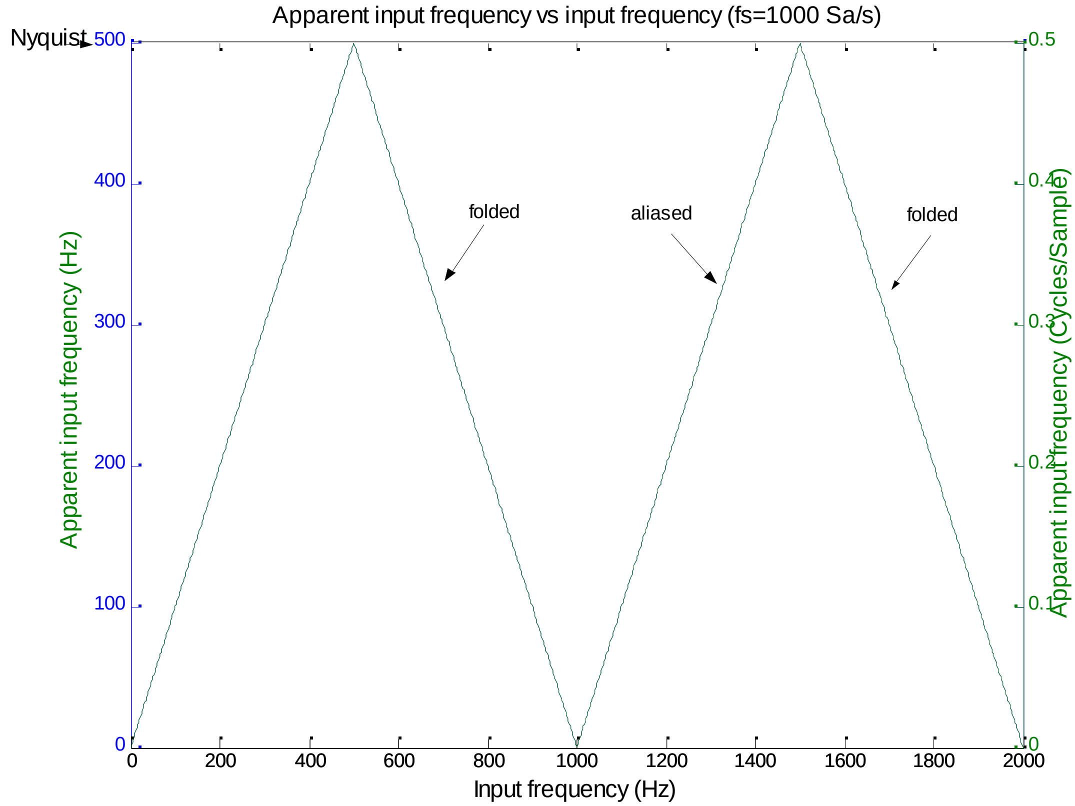

Figure 2: A plot of apparent input frequency for various input frequencies, showing aliasing and folding (negative phase).

The discussion so far leads to this conclusion: only those input frequencies that are ≥ 0 (DC) and are $< f_S/2$ (Nyquist frequency) are unambiguously determined after sampling. All other input frequencies suffer folding or aliasing. Therefore, if a signal has maximum frequency f_{MAX}, the sampling rate must be $f_S > 2 f_{MAX}$ (Nyquist rate).

Suppose that a signal is sampled at less than the Nyquist rate. What happens to the signal's spectral energy $\geq f_S/2$? This energy does not disappear. Rather, it is folded or aliased and then adds with the energy that exists DC and between $f_S/2$. This can be seen with the aid of a simple example: suppose that the input signal is

$$x(t) = 3\cos(2\pi \cdot 10 \cdot t) + \cos(2\pi \cdot 90 \cdot t + \pi/2) + 0.25\cos(2\pi \cdot 110 \cdot t)$$

After sampling at a rate of $f_S = 100$ Sa/s the result is

$$x[n] = 3\cos(2\pi \cdot 0.1 \cdot n) + \cos(2\pi \cdot 0.1 \cdot n - \pi/2) + 0.25\cos(2\pi \cdot 0.1 \cdot n)$$
$$= \sqrt{3.25^2 + 1.0^2}\cos(2\pi \cdot 0.1 \cdot n - 0.2985)$$

Therefore, the signal component at frequency $F = 0.1$ cycles/sample contains both folded and aliased energies.

This example shows how to account for the spectral amplitude that we see after sampling, and motivates the following principle:

*The spectral amplitude that is visible after sampling is the vector sum of the amplitude at the true corresponding frequency below Nyquist, as well as **all** folded components above Nyquist (with negative phase), and **all** aliased components above the sampling rate.*

OR

Spectral amplitude at frequency F = Vector sum of the following:

Spectral amplitude at frequency $f = Ff_S$
+ Spectral amplitudes at frequencies $f = nf_S - Ff_S$
with negative phase (folding), $n = 1, 2, 3, \ldots$
+ Spectral amplitudes at frequencies $f = nf_S + Ff_S$
aliasing $n = 1, 2, 3, \ldots$

Note that vector sums are being used, so the spectral amplitude at F after sampling can be *less* than the amplitude at the true corresponding frequency $f = Ff_S$. This is illustrated by the example shown in Figure 3.

In this example, a spectral amplitude distribution $A(f)$ and a phase distribution $\theta(f)$ is assumed. This means that the sinusoidal signal component at any frequency is

$$A(f)\cos\left(2\pi ft + \theta(f)\right)$$

For Figure 3, the distributions are

$$A(f) = \exp(-f^2/4) \quad \theta(f) = 2\pi f$$

Asssuming that the sampling rate is $f_S = 4$ Sa/s, then the vector sum of complex amplitudes for the original and folded components is

$$\left|A(f)e^{j\theta(f)} + A(4-f)e^{j\theta(4-f)}\right| = \left|e^{-f^2/4}e^{j2\pi f} + e^{-(4-f)^2/4}e^{j2\pi(4-f)}\right| = \left|e^{-f^2/4} + e^{-(4-f)^2/4}e^{-j4\pi f}\right|$$

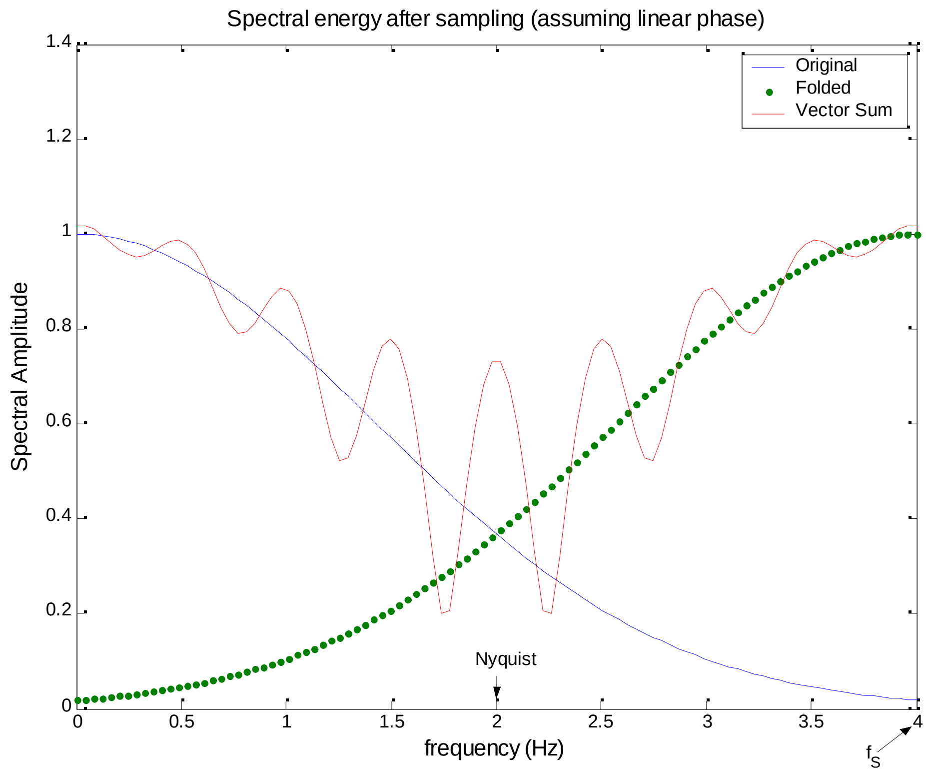

Figure 3: *Illustration of vector sum of original and folded spectral amplitudes.*

Oversampling method

There are advantages to oversampling, which is sampling at much higher than Nyquist rate. The oversampling ratio is defined as

$$OSR = \frac{f_S}{(2 * f_{max})}$$

When the OSR >> 1, the difference between samples is small. If large enough, e.g., OSR = 64, the difference $\Delta = new\ sample - old\ level$ can be recorded in a binary bit q as follows:

$$q = \begin{Bmatrix} 1, if\ \Delta \geq 0 \\ 0, else \end{Bmatrix}$$

One simple way to encode samples is to count the number of "1"s between samples. For example, if we had 4096 clocks between samples, the number of 1's can be counted in a 12-bit counter, allowing 12-bit quantization. Note that the value of "old level" is updated after each measurement, i.e. $old\ level = old\ level + \Delta$, if q = 1, and $old\ level = old\ level - \Delta$ if q = 0. With this updating scheme, zero difference is encoded as the alternating series 10101010....., producing a value of 2048 in the 12-bit counter mentioned above.

The method of binary-encoding deltas and counting them ("sigma", or sum), is called *delta-sigma A/D conversion*[2]. The terms "single-bit" ADC is also used.

The concept of oversampling is also used for Digital to Analog conversion (DAC). In this case it means digitally interpolating between existing samples to create a higher effective sampling rate at the DAC. A simple method is to linearly interpolate between existing samples, but there are better methods using digital filters that we shall discuss later.

A common method for DAC involves holding the value constant between samples. This method, known as the "zero-order hold", creates a blocky output waveform. The blockiness reduces the amplitude of high frequencies by an amount that varies with frequency according to the "sinc" function:

$$sinc(F) = \frac{\sin(\pi F)}{\pi F}$$

Note that sinc(0) = 1, by L'Hopital's rule. A graph of the sinc function over the range up to Nyquist is shown below in Figure 4(a). On the right, Fig. 4(b) shows the "1/sinc" correction necessary to invert this loss.

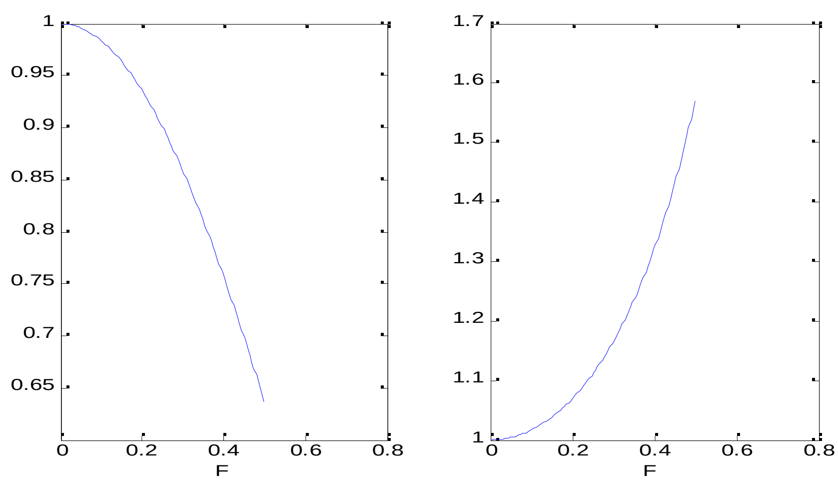

Figure 4 (a): Sinc function from DC to Nyquist; on right (b), the 1/sinc correction.

[2]Sometimes this is called "sigma-delta", which means the same thing. A nice interactive demo is at http://www.analog.com/Analog_Root/static/techSupport/designTools/interactiveTools/sdtutorial/sdtutorial.html

Chapter 3 of Smith[3] has an excellent description of the practical aspects of sampling and quantization. These notes are meant only to serve as an introduction to that chapter

Exercises

1. Sketch the spectral amplitude that would result in Figure 3 if the sampling rate was $f_S = 10$ Sa/s.

2. Why do wheels sometimes appear to go backwards in movies? A movie camera captures 24 frames in each second, and the rotation of a wheel can be modelled by a vector in the complex plane $w(t) = e^{j2\pi ft}$. Sketch values of $w(t)$ for various time samples, if $f_S = 24$ Sa/s, and the rotational speed of the wheel is $f = 10,\ 18,\ \text{or}\ 24$ cycles/second.

3. Explain what is meant by the terms: *delta-sigma A/D, 1/sinc(x) correction*, *oversampling D/A*

[3]See the book "An Engineer's and Scientist's Guide to DSP" at www.dspguide.com.