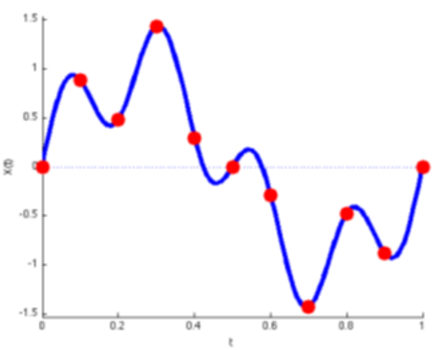
Lectures 9-10: Sampling Theorem

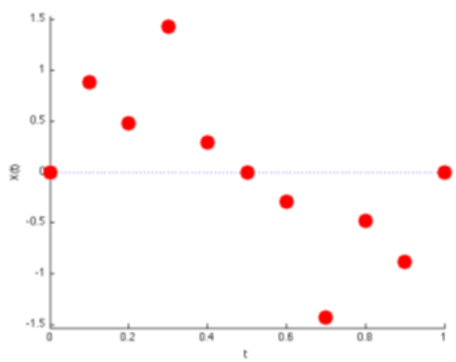
ENGR 76 lecture notes — May 2, 2024
Ayfer Ozgur, Stanford University

1 From continuous to discrete: The Sampling Theorem

Understanding the Fourier transform and the spectrum is the foundation for turning analog information into digital information. The question is: How accurately can a discrete set of sampled values of a continuous function represent the function at other values? That is, if we know the values of a function at a discrete set of points, how well can we interpolate the values in between those points?

This hardly seems reasonable – it even seems kind of crazy. A function, a signal, can jump all over the place, so how on earth would you expect to be able to say anything about *all* of its values by only knowing *some* of its values? Consider the following two pictures. The first is a sum of sine curves. The second is a bunch of points selected from the first. How do you propose to reconstruct the curve in the first picture from the bunch of dots in the second? The craziest thing is that this is not crazy. It's the representation of the frequency domain that leads to an answer.





Consider the continuous-time signal $s(t)$ shown in the figure above. Assume that this is a finite-length signal over the interval $[0, T]$, or equivalently a periodic function of period T. In any case, we have the Fourier series representation:

$$X(t) = \frac{b_0}{2} + \sum_{k=1}^{\infty} a_k \cos(2\pi \frac{k}{T} t) + b_k \sin(2\pi \frac{k}{T} t).$$

Let's assume the spectrum of this signal is limited to the band $[0, B]$, i.e. the largest harmonic in the above representation of the signal has frequency B. Equivalently, the summation runs from $k = 1$ to $k = BT$:

$$X(t) = \frac{b_0}{2} + \sum_{k=1}^{BT} a_k \cos(2\pi \frac{k}{T} t) + b_k \sin(2\pi \frac{k}{T} t).$$

This implies that the signal $X(t)$ can be completely described in terms of the $2BT + 1$ coefficients in its representation. In other words we have $2BT + 1$ degrees of freedom to describe the signal. Now assume we take samples of this signal every T_s seconds in the interval from 0 to T, so that the resultant samples are given by $X(0), X(T_s), X(2T_s), X(3T_s) \ldots$. Note that each sample of $X(t)$ gives us an equation of the form

$$X(mT_s) = \frac{b_0}{2} + \sum_{k=1}^{2BT} a_k \cos(2\pi \frac{k}{T} mT_s) + b_k \sin(2\pi \frac{k}{T} mT_s),$$

for some integer m. This defines an equation in terms of the $2BT+1$ Fourier coefficients. Hence, if we sample the signal so that we have $2BT+1$ samples, and therefore $2BT+1$ equations in its Fourier coefficients, we can compute the $2BT+1$ Fourier coefficients and recover the continuous-time signal. To obtain $2BT+1$ samples over the duration $[0,T]$ of the signal, we need

$$T_s < \frac{T}{2BT} = \frac{1}{2B}.$$

Equivalently, the sampling rate (or frequency) $f_s = \frac{1}{T_s} > 2B$. This is the number of samples we need to take per second. We arrive at the following sampling theorem:

Sampling Theorem: Suppose a signal is bandlimited. Let B be the maximum frequency in its frequency spectrum. If the signal is sampled at rate $f_s > 2B$, then it can be reconstructed *exactly* from its samples. $2B$ is called the Nyquist rate and the condition $f_s > 2B$ required for reconstruction is called the Nyquist condition.

How can we reconstruct the continious-time signal from its samples? The above argument suggests to first use the samples of the signal to compute the Fourier coefficients and then use the Fourier representation to construct the signal $X(t)$. In practice, we often want to be able to reconstruct the signal by interpolating its samples. Interpolation of the samples can be done in the following generic form,

$$\hat{X}(t) = \sum_{m\in\mathbb{Z}} X(mT_s) F\left(\frac{t-mT_s}{T_s}\right),$$

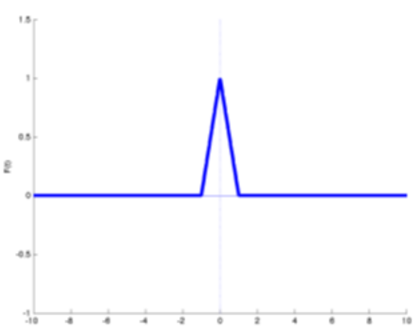
where $\hat{X}(t)$ denotes the reconstructed time-domain signal and $F(t):\mathbb{R}\to\mathbb{R}$ is an interpolation function s.t. $F(0)=1$ and $F(k)=0$ for all integer $k\neq 0$. Note that this property of F ensures that

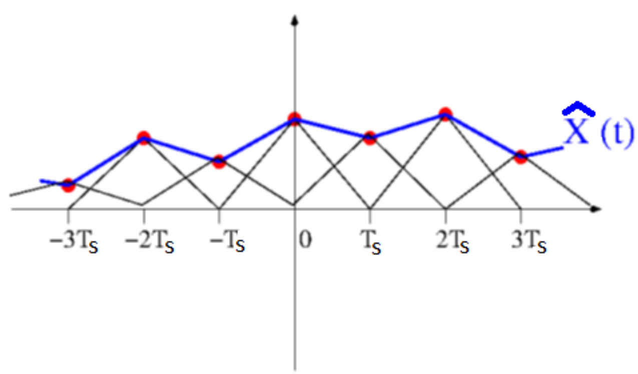
$$\hat{X}(kT_s) = X(kT_s),$$

i.e., the reconstruction $\hat{X}(t)$ matches the original function $X(t)$ at the sampling instants. For example, F can be chosen as

$$F(t) = \begin{cases} 1-|t|, & \text{if } -1\leq t\leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Note that this function takes the value 0 at 1, and is equal to 0 at all other integer time instants. This particular function leads to linear interpolation between the samples. See the picture below.





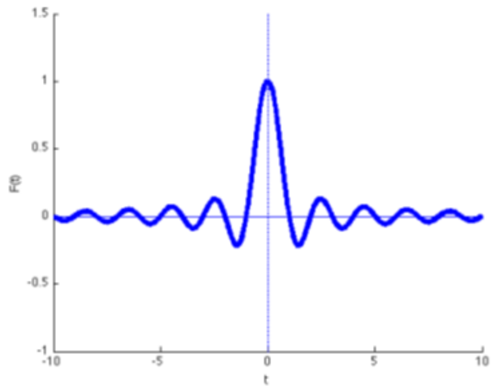
Linear interpolation is usually not desirable as it leads to sharp edges and a function that is not differentiable at the sampling instants. Instead, we can look for smoother functions F that can result in a smoother interpolation of the samples. For example, we can design F to be a polynomial with roots at integers $k\neq 0$ of the form

$$F(t) = (1-t)(1+t)(1-\frac{t}{2})(1+\frac{t}{2})(1-\frac{t}{3})(1+\frac{t}{3})\ldots$$

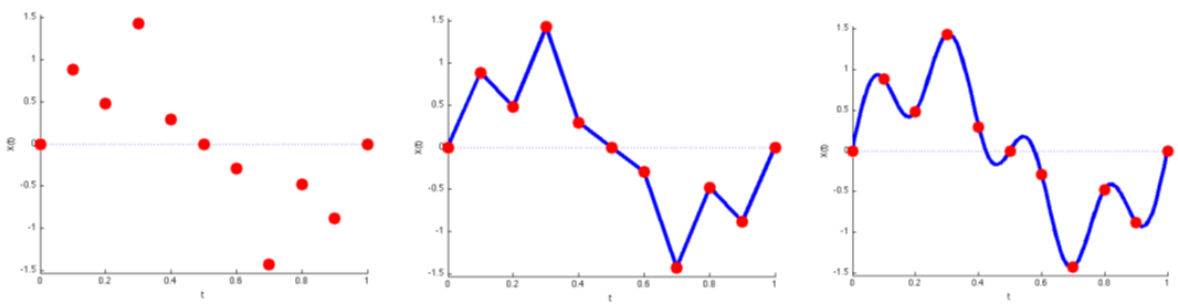
This function is known as the *sinc* function and can be expressed in the following equivalent form:

$$F(t) = \text{sinc}(\pi t) = \frac{\sin(\pi t)}{pit}. \tag{2}$$

Here is how this function looks like:



and here is a comparison of the two reconstructions of the signal by using the linear interpolation function in (1) and the sinc function in (2):



It can be shown that the interpolation with the sinc function will recover exactly the original $X(t)$ if sampling was done at the rate required by the sampling theorem, i.e. $f_s > 2B$. This leads to the following refined version of the sampling theorem. (The previous version is called the Nyquist sampling theorem, and the version below is usually credited to Shannon.)

Sampling Theorem: Suppose a signal is bandlimited. Let B be the maximum frequency in its frequency spectrum. If the signal is sampled at rate $f_s > 2B$, then

$$X(t) = \sum_{m \in \mathbb{Z}} X(mT_s)\text{sinc}\left(\frac{t - mT_s}{T_s}\right).$$

Look carefully at what we have done here. On the right hand side, we have discrete values of the function, the values $X(mT_s)$ at the sample points mT_s, and the formula says that via a sum of shifted sinc funcitons we can interpolate *any* value of the function $X(t)$ if we know its values only at the sample points. You can get it all back. Remember, however, that there is the assumption that the signal is band-limited. That's not a trivial assumption, but it's also not a very restrictive one in practice. For example, people hear in the range from about 20 Hz to 20,000 Hz. Thus, an audio signal is limited to bandwidth 20 kHz and can be sampled at 40 kHz to allow reconstruction without any loss of information. Indeed, audio is often sampled at a rate of 44.1 kHz. In practice, we also perform **quantization**, which stores each of the real samples at some finite precision using a finite number of bits; quantization, unlike sampling, does result in some loss of information.

2 The Stroboscopic Effect

What does the reconstruction

$$\hat{X}(t) = \sum_{m \in \mathbb{Z}} X(mT_s)\text{sinc}\left(\frac{t - mT_s}{T_s}\right)$$

give us when the signal is sampled below the Nyquist rate, i.e. $f_s < 2B$ or equivalently $T_s > 1/2B$. In this case, $X(t) \neq \hat{X}(t)$ but the reconstruction $\hat{X}(t)$ still matches the original signal $X(t)$ at the sampling instants, i.e. $X(mT_s) = \hat{X}(mT_s)$ for integer m. This is simply due to the fact that the sinc function is an interpolation function, i.e. $F(0) = 1$ and $F(k) = 0$ for all integer $k \neq 0$. It can be shown that the constructed signal $\hat{X}(t)$ will always have spectrum limited to the interval $[0, \frac{1}{2T_s} = [0, \frac{f_s}{2}]$. Note that $f_s/2 < B$ when sampling is done below the Nyquist rate, so the reconstructed signal has smaller bandwidth than the original signal. In other words when $f_s/2 < B$, the formula above constructs a signal $\hat{X}(t)$ which matches the original signal $X(t)$ at the sampling instants and for which the sampling rate f_s would satisfy the Nyquist condition. This is called the stroboscopic effect. It accounts for the "wagon-wheel effect", so-called because in video, spoked wheels (such as on horse-drawn wagons) sometimes appear to be turning backwards at a slower rate. Check the Lecture slides 9 for other examples of the stroboscopic effect.