Nyquist–Shannon sampling theorem

Emiel Por, Maaike van Kooten & Vanja Sarkovic

May 2019

1 Theory

1.1 The Nyquist-Shannon sampling theorem

The Nyquist theorem describes how to sample a signal or waveform in such a way as to not lose information. Suppose that we have a bandlimited signal $X(t)$. Bandlimited means that if we were to take the Fourier transform of this signal, $\hat{X}(f) = \mathcal{F}\{X(t)\}$, there would be a certain f_{max} for which

$$|\hat{X}(f)| = 0 \quad \forall \; |f| > f_{\mathrm{max}}, \tag{1}$$

so that there is no power in the signal beyond the maximum frequency f_{max}. The Nyquist theorem then states that if we were to sample this signal we would need samples with a frequency larger than twice the maximum frequency contained in the signal, that is

$$f_{\mathrm{sample}} \geq 2 f_{\mathrm{max}}. \tag{2}$$

If this is the case, we have not lost any information during the sampling process and we could theoretically reconstruct the original signal from the sampled signal.

Alternatively we can define a Nyquist frequency based on a certain sampling frequency:

$$f_{\mathrm{Nyquist}} = \tfrac{1}{2} f_{\mathrm{sample}}. \tag{3}$$

Any signals that contain frequencies higher than this Nyquist frequency cannot be perfectly reconstructed from the sampled signal, and are called *undersampled.* If our signal only contains frequencies smaller than the Nyquist frequency, we can perfectly reconstruct the original signal given the sampled signal, and we are *oversampled.* When our signal is bandlimited to a frequency equal to the Nyquist frequency, we are *critically sampled.*

It is sometimes useful to state the Nyquist theorem in a different way. Suppose we have a certain bandlimited signal with a maximum frequency f_{max}. The period of this maximum frequency is $\Delta t_{\mathrm{min}} = 1/f_{\mathrm{max}}$. If we want to correctly sample this signal, we would need to sample with a period of

$$\Delta t_{\mathrm{sample}} \leq \tfrac{1}{2} \Delta t_{\mathrm{min}}. \tag{4}$$

If we were to sample our signal slower, with a longer interval between samples, we would be undersampling our signal.

1.2 Aliasing

So what happens when a signal is undersampled? Aliasing occurs. This is best shown this visually; in Figure 1 we show the process of sampling two different signals (in yellow). Both signals are sampled with the same sampling frequency at points in red. The top signal is oversampled, i.e., its frequency is lower than half the sampling frequency, so we have more than two samples per period of this sinusoid. Here we can perfectly reconstruct the original signal. The bottom signal, however, is undersampled. We have less than two samples per period of this sinusoid and when we try to reconstruct the signal (blue line), we are not reconstructing the original signal, but rather a much lower frequency. This effect is called aliasing. If we are undersampled, the frequencies that are higher than the Nyquist frequency are reconstructed at lower frequencies and will add noise to the actual signal at those lower frequencies.

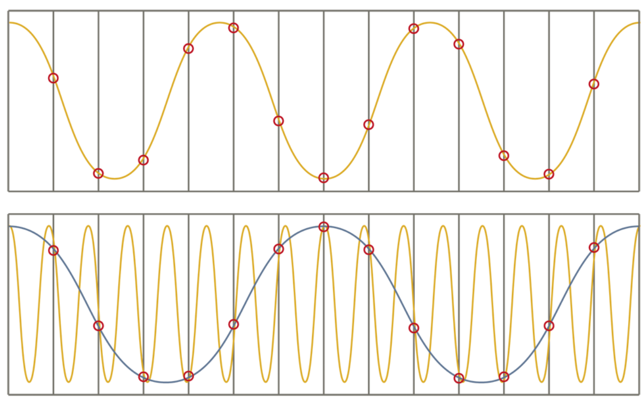

Figure 1: Aliasing. The top signal (in yellow) is oversampled (samples in red), while the bottom signal is undersampled. The reconstructed signal (in blue)from the sampled data yields a much lower frequency than the original signal. This is called aliasing.

In practice we want to avoid being undersampled (to avoid aliasing). At the same time we also want to avoid being too much oversampled, as this typically increases the noise per pixel and is less efficient.

2 Examples

2.1 Temporal

Understanding the Nyquist sampling theorem is important in dealing with time series analysis. It also provides insight into the limitations of your temporal data set. It is important when you need to determine a continuous function from a discrete measurements.

Some examples of aliasing in the temporal domain occurs for:

1. *Helicopter blades illusion* (`https://youtu.be/qgvuQGY946g`). You can find videos on YouTube where the helicopter is flying through the air but the blades do not move. There are many other examples of so called wagon-wheel effect (for example `https://youtu.be/6XwgbHjRo30`).

2. *Source variability in astronomy such as pulsars, stellar vibrations, etc...* For these sources there is a minimum time interval by which the source can vary, given by the transit time of the light across the object (ie. D/c where D is the diameter of the object). Trying to sample the light from an object faster than this doesn't improve your analysis, and might actually harm your data-reduction as you are reducing your signal-to-noise per image. If you sample slower though, you might miss some features that you would have seen otherwise, or you see them at their aliased frequencies.

2.2 Spatial

Until now we were using time and frequency, but the Nyquist theorem applies equally well to space and spatial frequency (see Equation 4).

Some examples of aliasing in the spatial domain occurs for:

1. *Small text on computer screen.* Small text on a computer screen is often hard to read. Modern operating systems use anti-aliasing to suppress these aliasing effects to make text much easier to read.

2. *Medical imaging techniques.* For example MRI imaging suffers from aliasing which results in wrap around artifacts where a part of the body is seen on the other side of the final image. This is a specific cause of aliasing that occurs when the object being imaged is larger than the FOV of the MIR machine.

3. *Astronomical imaging.* For imaging we are sampling the intensity image with pixels. This means that to preserve all details in the image, the pixel size should be smaller than half of the minimum period of the image, typically given by the resolution of your telescope (ie. λ/D or λ/r_0 when we are diffraction-limited or seeing-limited respectively). (So we need at least two pixels per resolution element.) If we have larger pixels, this results in a loss in resolution and might introduce aliasing. Using smaller pixels increases noise originating from your camera electronics.

4. *Spectrographs.* Similarly to imaging, we require at least two pixels per resolution element for our spectrograph camera. So each pixel must be smaller than $\frac{1}{2}R\lambda$ in wavelength.

Figure 2: An example of aliasing (and anti-aliasing) on text for computer screens. Anti-aliasing uses a low-pass filter on the text so that aliasing effects are suppressed, making it much easier to read on low-resolution displays.

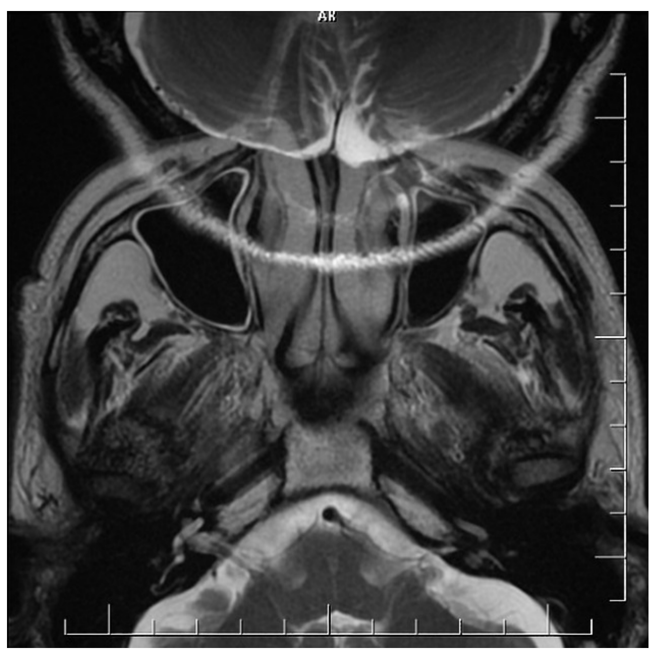

Figure 3: An example of an MRI image with aliasing artefacts. The aliasing causes the image to be reflected by the edges of the screen, making it hard for the operator to determine signal from artefact.