The Shannon Sampling Theorem and Its Implications

Gilad Lerman
Notes for Math 5467

1 Formulation and First Proof

The sampling theorem of bandlimited functions, which is often named after Shannon, actually predates Shannon [2]. This is its classical formulation.

Theorem 1.1. *If $f \in L_1(\mathbb{R})$ and $\hat{f}$, the Fourier transform of f, is supported on the interval $[-B, B]$, then*

$$f(x) = \sum_{n \in \mathbb{Z}} f\left(\frac{n}{2B}\right) \operatorname{sinc}\left(2B\left(x - \frac{n}{2B}\right)\right), \tag{1}$$

where the equality holds in the L_2 sense, that is, the series in the RHS of (1) *converges to f in $L_2(\mathbb{R})$.*

In other words, the theorem says that if an absolutely integrable function contains no frequencies higher than B hertz, then it is completely determined by its samples at a uniform grid spaced at distances 1/(2B) apart via formula (1).

1.1 Terminology and Clarifications

We say that a function f (with values in either $\mathbb{R}$ or $\mathbb{C}$) is supported on a set A if it is zero on the complement of this set. The *support* of f, which we denote by $\operatorname{supp}(g)$, is the minimal closed set on which f is supported, equivalently, it is the closure of the set on which f is non-zero. We note that if $f \in L_1(\mathbb{R})$ is real-valued then the support of $\hat{f}$ is symmetric around zero (since the real part of $\hat{f}$ is even and the imaginary part is odd). A function

$f \in L_1(\mathbb{R})$ is *bandlimited* if there exists $B \in \mathbb{R}$ such that $\text{supp}(\hat{f}) \subseteq [-B, B]$. We call B a band limit for f and $\Omega := 2B$, the corresponding frequency band. The minimal value of B (such that $\text{supp}(\hat{f}) \subseteq [-B, B]$), that is, the supremum of the absolute values of all frequencies of f, is called the *Nyquist frequency* of f and its corresponding frequency band is called the *Nyquist rate*. We denote the Nyquist frequency by B_{Nyq}, so that the Nyquist rate is $2B_{\text{Nyq}}$.

Let us recall that the "general principle of the Fourier transform" states that the decay of $\hat{f}$ implies smoothness of f and vice versa. Band-limited functions have the best decay one can wish for (their Fourier transforms are zero outside an interval), they are also exceptionally smooth, that is, their extension to the complex plan is analytic, in particular, they are $C^\infty(\mathbb{R})$ functions. This fact follows from a well-known theorem in analysis, the Payley-Wiener Theorem [1]. We thus conclude that if f is a band-limited signal, then its values $\{f(k/2B)\}_{k\in\mathbb{Z}}$ are well-defined. [1]

Having a bandlimit is natural for audio signals. Human voice only occupies a small piece of the band of audible frequencies, typically between 300 Hz and 3.5 KHz (even though we can hear up to approximately 20 KHz). On the other hand, this is a problematic requirement for images. Sharp edges in natural images give rise to high frequencies and our visual system is often intolerable to thresholding these frequencies. Nevertheless, Shannon sampling theory still clarifies to some extent the distortion resulting from subsampling images and how one can weaken this distortion by initial lowpass filtering.

1.2 First Proof of the Sampling Theorem

Since $f \in L_1(\mathbb{R})$ and $\text{supp}(\hat{f}) \subseteq [-B, B]$, then $\hat{f}$ is a bounded function supported on $[-B, B]$, in particular, $\hat{f} \in L_2([-B, B])$. We can thus expand $\hat{f}$ according to its following Fourier series:

$$\hat{f}(\xi) = \sum_{n\in\mathbb{Z}} c_n e^{\frac{2\pi i n \xi}{2B}}, \tag{2}$$

[1] If f is a rather arbitrary function in $L_1(\mathbb{R})$, then by changing its value at a point, we obtain a function whose L_1 distance from f is 0. That is, in general, the values $\{f(k/2B)\}_{k\in\mathbb{Z}}$ are not uniquely determined for $f \in L_1(\mathbb{R})$, however, sufficient smoothness, e.g., continuity, implies their unique definition.

where

$$c_n = \frac{1}{2B}\int_{-B}^{B} \hat{f}(x)e^{\frac{-2\pi in\xi}{2B}}\,\mathrm{d}x = \frac{1}{2B}\int_{-\infty}^{\infty} \hat{f}(x)e^{\frac{-2\pi in\xi}{2B}}\,\mathrm{d}x = \frac{1}{2B}f\left(\frac{-n}{2B}\right). \quad (3)$$

Therefore,

$$\hat{f}(\xi) = \sum_{n\in\mathbb{Z}} \frac{1}{2B}f\left(\frac{-n}{2B}\right)e^{\frac{2\pi in\xi}{2B}} = \sum_{n\in\mathbb{Z}} \frac{1}{2B}f\left(\frac{n}{2B}\right)e^{\frac{-2\pi in\xi}{2B}}. \quad (4)$$

From (4) it is already clear that f can be completely recovered by the values $\{f(n/(2B))\}_{n\in\mathbb{Z}}$. To conclude the recovery formula we invert $\hat{f}$ as follows:

$$f(x) = \int_{-B}^{B} \hat{f}(\xi)e^{2\pi ix\xi}\,\mathrm{d}\xi = \sum_{n\in\mathbb{Z}} f\left(\frac{n}{2B}\right)\frac{1}{2B}\int_{-B}^{B} e^{2\pi i\xi\left(x-\frac{n}{2B}\right)}\,\mathrm{d}\xi \quad (5)$$

$$= \sum_{n\in\mathbb{Z}} f\left(\frac{n}{2B}\right)\frac{1}{2B}\left.\frac{e^{2\pi i\xi\left(x-\frac{n}{2B}\right)}}{2\pi i\left(x-\frac{n}{2B}\right)}\right|_{\xi=-B}^{B}$$

$$= \sum_{n\in\mathbb{Z}} f\left(\frac{n}{2B}\right)\frac{1}{2\pi B\left(x-\frac{n}{2B}\right)}\frac{e^{2\pi iB\left(x-\frac{n}{2B}\right)} - e^{-2\pi iB\left(x-\frac{n}{2B}\right)}}{2i}$$

$$= \sum_{n\in\mathbb{Z}} f\left(\frac{n}{2B}\right)\frac{\sin\left(2\pi B\left(x-\frac{n}{2B}\right)\right)}{2\pi B\left(x-\frac{n}{2B}\right)} = \sum_{n\in\mathbb{Z}} f\left(\frac{n}{2B}\right)\operatorname{sinc}\left(2B\left(x-\frac{n}{2B}\right)\right).$$

1.3 What Else Do We Want to Understand?

The above proof does not completely explain what may go wrong if we sample according to a frequency $B < B_{\mathrm{Nyq}}$; we refer to such sampling as *undersampling*. It is also not easy to see possible improvement of the theory when $B > B_{\mathrm{Nyq}}$, that is, when *oversampling*. Nevertheless, if we try to adapt the above proof to the case where $B < B_{\mathrm{Nyq}}$, then we may need to periodize $\hat{f}$ with respect to the interval $[-B, B]$ and then we can expand the periodized function according to its Fourier series. Therefore, in §2 we discuss the Fourier series expansions of such periodized functions. This is in fact the well-known Poisson's summation formula. In §3 we describe a second proof for the Shannon's sampling theorem, which is based on the Poisson's summation formula. Following the ideas of this proof, §4 explains the distortion obtained by the recovery formula (1) when sampling with frequency rates lower than Nyquist; it also clarifies how to improve such signal degradation by initial lowpass filtering. Section 5 explains how to obtain better recovery formulas when the sampling frequency is higher than Nyquist. At last, we discuss in §6 further implications of these basic principles, in particular, analytic interpretation of the Cooley-Tukey FFT.

2 Poisson's Summation Formula

The following theorem is a formulation of Poisson summation formula with additional frequency B (so that it fits well with the sampling formula). It uses the function $\sum_{n=-\infty}^{\infty} f(x + n/(2B))$, which is a periodic function of period $1/(2B)$ (indeed, it is obtained by shifting the function $f(x)$ at distances $n/(2B)$ for all $n \in \mathbb{Z}$ and adding up all this shifts). It is common to formulate Poisson's summation formula with only $B = 1/2$.

Theorem 2.1. *If $f \in L_2(\mathbb{R})$, $B > 0$ and*

$$\sum_{n=-\infty}^{\infty} f\left(x + \frac{n}{2B}\right) \in L_2\left(\left[0, \frac{1}{2B}\right]\right), \tag{6}$$

then

$$\sum_{n=-\infty}^{\infty} f\left(x + \frac{n}{2B}\right) = \sum_{n=-\infty}^{\infty} 2B\hat{f}(2Bn)\, e^{2\pi i n x \cdot 2B}. \tag{7}$$

Proof. Since $\sum_{n=-\infty}^{\infty} f\left(x + \frac{n}{2B}\right)$ is in $L_2([0, 1/(2B)])$, we can expand it by its Fourier series as follows:

$$\sum_{n=-\infty}^{\infty} f\left(x + \frac{n}{2B}\right) = \sum_{m=-\infty}^{\infty} c_m\, e^{2\pi i m x \cdot 2B}, \tag{8}$$

where

$$c_m = 2B \int_0^{\frac{1}{2B}} \sum_{n=-\infty}^{\infty} f\left(x + \frac{n}{2B}\right) e^{-2\pi i m x \cdot 2B}\, \mathrm{d}x \tag{9}$$

$$= \sum_{n=-\infty}^{\infty} 2B \int_0^{\frac{1}{2B}} f\left(x + \frac{n}{2B}\right) e^{-2\pi i m x \cdot 2B}\, \mathrm{d}x.$$

Applying the change of variables $y = x + n/(2B)$ into the RHS of (9) and then using the fact that $e^{-2\pi inm} = 1$, we conclude the theorem as follows

$$\begin{aligned} c_m &= \sum_{n=-\infty}^{\infty} 2B \int_{\frac{n}{2B}}^{\frac{n+1}{2B}} f(y)e^{-2\pi imy\cdot 2B}\,\mathrm{d}y \\ &= 2B \int_{-\infty}^{\infty} f(y)e^{-2\pi imy\cdot 2B} = 2B\,\hat{f}(2Bm). \end{aligned} \tag{10}$$

□

We can reformulate Theorem 2.1 as follows:

Theorem 2.2. *If* $f \in L_2(\mathbb{R})$*,* $B > 0$ *and* $\sum_{n=-\infty}^{\infty} \hat{f}(\xi + 2Bn) \in L_2([0, 2B])$*, then*

$$\sum_{n=-\infty}^{\infty} \hat{f}(\xi + 2Bn) = \frac{1}{2B} \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2B}\right) e^{\frac{-2\pi in\xi}{2B}}. \tag{11}$$

Theorems 2.1 and 2.2 are clearly equivalent. For example, Theorem 2.2 follows from Theorem 2.1 by replacing f with $\hat{f}$ (and thus $\hat{f}(\xi)$ with $f(-x)$, which equals $\hat{\hat{f}}(x)$) and replacing $2B$ with $1/2B$ as well as changing variables in the RHS of (7) from n to $-n$. We use the period $2B$ for the function on the LHS of (11) and $1/(2B)$ for the function on the LHS of (11) due to the transformation of scale between the spatial and frequency domains (however, the choice of scale can be arbitrary).

We note that Theorem 2.2 implies that given the samples $\{f(n/(2B))\}_{n\in\mathbb{Z}}$, we can then recover the periodic summation of $\hat{f}$ with period $2B$, that is, the function

$$P_{2B}(\hat{f}) := \sum_{n=-\infty}^{\infty} \hat{f}(\xi + 2Bn). \tag{12}$$

This observation is the essence of our second proof of the sampling theorem.

3 A Second Proof of the Sampling Theorem

Since $\mathrm{supp}(\hat{f}) \subseteq [-B, B]$, $P_{2B}(\hat{f})$ satisfies the identity

$$\hat{f}(\xi) = \chi_{[-B,B]}(\xi) P_{2B}(\hat{f})(\xi) \quad \text{for all} \quad \xi \neq -B, B. \tag{13}$$

We remark that if $\hat{f}(B) \neq 0$ (and consequently also $\hat{f}(-B) \neq 0$), then (13) does not hold for $\xi = -B$ (and consequently also for $\xi = B$), but $P_{2B}(\hat{f})(\xi) = \hat{f}(B)+\hat{f}(-B)$. Nevertheless, the LHS and RHS of (13) are equal in $L_2([-B,B])$.

Combining (11)-(13), we conclude the following equality in L_2:

$$\hat{f}(\xi) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2B}\right) \frac{1}{2B}\, \chi_{[-B,B]}(\xi) e^{\frac{-2\pi i n \xi}{2B}}. \tag{14}$$

We recall that $\frac{1}{2B}\chi_{[-B,B]}(\xi)$ is the Fourier transform of $\mathrm{sinc}(2Bx)$ and thus $\frac{1}{2B}\chi_{[-B,B]}(\xi)e^{\frac{-2\pi i n\xi}{2B}}$ is the Fourier transform of $\mathrm{sinc}\left(2B\left(x - \frac{n}{2B}\right)\right)$. At last, we use this observation and invert the Fourier transforms in both sides of (14). We consequently obtain the sampling formula (as an equality in L_2):

$$f(x) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2B}\right) \mathrm{sinc}\left(2B\left(x - \frac{n}{2B}\right)\right). \tag{15}$$

4 On Aliasing and Anti-aliasing

Assume that f is a band-limited function in L_1 and B is lower than its Nyquist frequency. Suppose that we sample f at $\{n/2B\}_{n\in\mathbb{Z}}$ and try to recover f by its samples. It is interesting to know how well we can approximate f this way. The second proof of the sampling theorem provides a good answer. Indeed, since $P_{2B}(\hat{f})$ is in L_2 (this is because $\hat{f}$ is bounded and supported in $[-B,B]$), then we can replace (14) with

$$\chi_{[-B,B]}(\xi)P_{2B}(\hat{f})(\xi) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2B}\right) \frac{1}{2B}\chi_{[-B,B]}(\xi)e^{\frac{-2\pi i n\xi}{2B}}. \tag{16}$$

Let g be the inverse Fourier transform of $\chi_{[-B,B]}(\xi)P_{2B}(\hat{f})(\xi)$, then similarly to deriving (15), we obtain that

$$g(x) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2B}\right) \mathrm{sinc}\left(2B\left(x - \frac{n}{2B}\right)\right). \tag{17}$$

That is, the sampling formula recovers g, which we refer to as an alias of f.

Example 4.1. *We arbitrarily fix $\epsilon > 0$ and $B_0 > 0$ and let*

$$f(x) = h_\epsilon(x)\, 2B_0\, \mathrm{sinc}(2B_0 x), \tag{18}$$

where $\operatorname{supp}(\hat{h}_\epsilon) = [-\epsilon, \epsilon]$ *and* $h_\epsilon(0) = 0$ *(equivalently* $\int \hat{h}_\epsilon(\xi)\,\mathrm{d}\xi = 0$ *and thus* $\hat{h}_\epsilon$ *oscillates). We note that*

$$\hat{f}(\xi) = \hat{h}_\epsilon(\xi) * \chi_{[-B_0,B_0]}(\xi) = \int_{-B_0}^{B_0} \hat{h}_\epsilon(\xi - \eta)\,\mathrm{d}\eta = \int_{-\xi-B_0}^{-\xi+B_0} \hat{h}_\epsilon(\zeta)\,\mathrm{d}\zeta. \quad (19)$$

In particular, $\hat{f}$ *is supported on the interval* $[-B_0 - \epsilon, B_0 + \epsilon]$*, that is,* $B_{\mathrm{Nyq}} = B_0 + \epsilon$.

Assume that we want to sample f *with frequency* B_0*. Equation* (19) *implies that*

$$\chi_{[-B_0,B_0]}(\xi) P_{2B_0}(\hat{f})(\xi) = \chi_{[-B_0,B_0]}(\xi) \int_{-\infty}^{\infty} \hat{h}_\epsilon(\zeta)\,\mathrm{d}\zeta = \chi_{[-B_0,B_0]} h_\epsilon(0) = 0. \quad (20)$$

That is, the alias function g *is the zero function. We also note that by the definitions of* f *and* h_ϵ*,* $f(n/(2B_0)) = 0$ *for any* $n \in \mathbb{Z}$*. This observation also confirms that the alias recovered by the sampling formula (i.e.,* g*) is the zero function.*

Example 4.2. *Let* $f(x) = \operatorname{sinc}^2(x)$ *so that* $\hat{f}(\xi) = (1 - |\xi|)\chi_{[-1,1]}(\xi)$ *(this follows from the solution to the second part of problem 4 in Homework 3 and the fact that* f *is real and even). The Nyquist frequency of* f *is* $B_{\mathrm{Nyq}} = 1$*. Assume that we sample* f *with frequency* $B = 0.5$ *and then try to recover* f *by the sampling formula with this* B*. By direct calculation (or plotting all intersecting shifts and adding them up) we obtain that*

$$\chi_{[-0.5,0.5]}(\xi) P_1(\hat{f})(\xi) = \chi_{[-0.5,0.5]}(\xi)(1 - |\xi| + |\xi|) = \chi_{[-0.5,0.5]}(\xi). \quad (21)$$

This is the Fourier transform of the alias $g(x) = \operatorname{sinc}(x)$*. Clearly* f *and its alias* g *are rather different (*$f = g^2$*).*

While we cannot recover f *from its samples at the integer values, we can get a better approximation to it. Indeed, let us zero out its Fourier transform outside the interval* $[-0.5, 0.5]$*. That is, we define* $\tilde{f}$ *to be the function such that* $\hat{\tilde{f}} = \hat{f}\chi_{[-0.5,0.5]}$*. We note that*

$$\begin{aligned} \hat{\tilde{f}}(\xi) &= (1 - |\xi|)\chi_{[-0.5,0.5]}(\xi) = \frac{1}{2}\chi_{[-0.5,0.5]}(\xi) + \frac{1}{2}(1 - 2|\xi|)\chi_{[-0.5,0.5]}(\xi) \\ &= \frac{1}{2}\chi_{[-0.5,0.5]}(\xi) + \frac{1}{4} \cdot 2\hat{f}(2\xi). \end{aligned} \quad (22)$$

Therefore (applying the solutions of the first two parts of problem 4 in Homework 3)

$$\tilde{f}(x) = \frac{1}{2}\operatorname{sinc}(x) + \frac{1}{4}\operatorname{sinc}^2(x/2) = \frac{\operatorname{sinc}(x)}{2} + \frac{\operatorname{sinc}^2(x)}{2(1+\cos(x))}. \tag{23}$$

The function $\tilde{f}$ is closer to f than the alias g. We clarify this claim in a more general setting next.

We claim that $\tilde{f}$ of Example 4.2 is the best L_2 approximation to f that coincides with f at the undersampled values and reproduced by the sampling formula. The requirement of reproducing by the sampling formula with spacing $1/(2B)$ can be replaced with the requirement of band limit $2B$. We even claim the following more general statement.

Proposition 4.1. *If $f \in L_1(\mathbb{R})$ and $B > 0$, then among all functions $h \in L_1$ of band limit B the function $\tilde{f}$, whose Fourier transform is $\hat{\tilde{f}} = \hat{f}\chi_{[-B,B]}$, obtains the shortest L_2 distance to f.*

Proof. If $h \in L_1$ and its Fourier transform is supported in $[-B, B]$, then

$$\|f-h\|_2^2 = \|\hat{f}-\hat{h}\|_2^2 = \int_{\xi\leq B} |\hat{f}(\xi)-\hat{h}(\xi)|^2\,\mathrm{d}\xi + \int_{\xi>B} |\hat{f}(\xi)|^2\,\mathrm{d}\xi. \tag{24}$$

Since the second term in the RHS of (24) is independent of h, $h = \tilde{f}$ is a minimizer of $\|f-h\|_2$. □

The replacement of f with $\tilde{f}$ before subsampling is referred to as *anti-aliasing* (we slightly generalize this definition below). That is, if we need to sample a signal f at $\{n/2B\}_{n\in\mathbb{Z}}$, where $B < B_{\mathrm{Nyq}}$, we can anti-alias it by first zeroing out the Fourier transform of f outside the interval $[-B, B]$ and then subsample. In other words, we apply a lowpass filter before subsampling. However, we may prefer a low-pass filter, which is more localized in the spatial domain, equivalently, smoother in the Fourier domain. We will thus refer to any such application of low-pass filter (with support of size comparable to $2B$, but not smaller than $2B$) before subsampling at equidistances $1/(2B)$ as anti-aliasing.

Examples of aliasing and anti-aliasing for one-dimensional signals, images and videos are further demonstrated in the class slides.

5 Recovery Formulas for Oversampled Signals

The second proof of Theorem 1.1 also suggests better formulas for signal recovery when f is oversampled. Indeed assume that $B > B_{\mathrm{Nyq}}$, then (13) can be modified as follows:

$$\hat{f}(\xi) = 2B\hat{h}(\xi/B)P_{2B}(\hat{f})(\xi) \quad \text{for all} \quad \xi \in \mathbb{R}, \tag{25}$$

where $h(x)$ is any arbitrary L_1 function such that $\hat{h}(\xi/B) \in C^{\infty}(\mathbb{R})$, supported on $[-B, B]$ and satisfying $\hat{h}(\xi) = 1$ for all $\xi \in [-B_{\mathrm{Nyq}}, B_{\mathrm{Nyq}}]$. Similarly, to deriving (15), we conclude that

$$f(x) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2B}\right) h\left(2B\left(x - \frac{n}{2B}\right)\right). \tag{26}$$

Since $\hat{h}$ is a C^{∞} function, h is localized (it cannot have compact support since $\hat{h}$ is compactly supported, but it will decay fast).

6 Further Implications

The basic ideas of the Shannon sampling theorem and its proof will be fundamental in the next material of image pyramids, subband coding and wavelet transforms. They are all based on subdividing the frequency band into low and high frequencies (ideally in each part we have exactly half the original band) and then subsampling the highpass and lowpass parts by a factor of 2 , so that the frequency rate and sampling rate in the subsampled signals are the same as those of the original ones.

The Cooley-Tukey FFT algorithm can also be interpreted in view of the sampling theorem. We recall that this algorithm recursively divides the DFT signal into its odd and even parts, while proceeding from top to bottom. It then continues from bottom to top recovering DFT of subsignals from the DFTs of their odd and even parts. A DFT of a signal (or subsignal) of length $2L$, denoted by $\{\hat{X}_{2L}(n)\}_{n=0}^{2L-1}$, is recovered by its even and odd parts of length L, denoted by $\{\hat{X}_L^{\mathrm{even}}(n)\}_{n=0}^{L-1}$ and $\{\hat{X}_L^{\mathrm{even}}(n)\}_{n=0}^{L-1}$, by applying the following formulas for $n = 0, \ldots, L-1$:

$$\hat{X}_{2L}(n) = \hat{X}_L^{\mathrm{even}}(n) + \hat{X}_L^{\mathrm{odd}}(n)e^{\frac{-2\pi i n}{2L}}, \tag{27}$$

and

$$\hat{X}_{2L}(n+L) = \hat{X}_L^{\text{even}}(n) + \hat{X}_L^{\text{odd}}(n) e^{\frac{-2\pi i(n+L)}{2L}}. \tag{28}$$

That is, the two subsampled signals are combined, while shifting the frequencies of the odd signals.

References

[1] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[2] A. Zayed. *Advances in Shannon's Sampling Theory*. Taylor & Francis, 1993.