

Using the Shannon Sampling Theorem to Design Compact Discs

Susan E. Kelly

Abstract

The Shannon Sampling Theorem states that if a signal $f(t)$ contains no frequencies greater than $\frac{W}{2}$ cycles per second, then the signal is completely determined by its values at a sequence of points spaced $\frac{1}{W}$ seconds apart. Using this theory, music can be stored on a compact disc by recording the function's amplitude at points sampled at the correct rate.

This material, including the proof provided for the theorem, is accessible to students who have done some work with Fourier series and transforms. With modifications, the same ideas could be given to Calculus students to illustrate the usefulness of series. The author has used this material with students at both of these levels.

Have you ever wondered how music is stored on a compact disc and then reproduced? Mathematically the music can be viewed as a function in time. To record the music on a disc, the function is converted to a series of numbers which represent the function's amplitude at evenly spaced points in time. These numbers are what are stored on the disc. The Shannon Sampling Theorem explains how many amplitude readings are needed and how these numbers are used as the coefficients of a series which will reproduce the music.

To understand this theorem, it is necessary to examine a small amount of Complex Number theory and some Fourier Analysis. The Shannon Sampling Theorem will then be given and we will see how it can be applied to the design of compact discs and players.

1 Background in Complex Numbers

A few basic facts from Complex numbers are needed. In this theory i is defined as the square root of -1 . The complex numbers are defined to be

the collection of all elements of the form $z = x + iy$, where x and y are real numbers.

In Calculus courses several series representations for real valued functions are studied. For example,

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \quad ,$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots \quad ,$$

and

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots \quad .$$

These series converge absolutely and uniformly if x lies in any finite interval. These formulae can be extended to complex valued functions of a complex variable. For example,

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots \quad .$$

It can be shown that the above series is uniformly convergent if $|z| \leq N$, for each $N > 0$.

Letting $z = ix$, where x is a real number, yields Euler's formula,

$$e^{ix} = \cos x + i \sin x. \tag{1}$$

From this it is easy to show that

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2} \tag{2}$$

and

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i}. \tag{3}$$

These last three formulas are all that we shall need from Complex numbers.

2 Fourier Analysis

Some background in Fourier Analysis is also needed. A Fourier sine and cosine series is based on the fact that the set $\{\frac{1}{\sqrt{2}}, \cos(\frac{nt}{W}), \sin(\frac{nt}{W})\}_{n=1}^{\infty}$ is an

orthogonal set in the interval $[-\pi W, \pi W]$. That is, for elements g_1 and g_2 of the set,

$$\int_{-\pi W}^{\pi W} g_1(t)g_2(t)dt = \begin{cases} 0 & \text{if } g_1 \neq g_2 \\ \pi W & \text{if } g_1 = g_2. \end{cases} \quad (4)$$

This can be verified using several cases. The work can be simplified using properties of even and odd functions and the complex exponential representations of sine and cosine functions given in Equations (2) and (3).

With this result we wish to represent a $2\pi W$ -periodic function in terms of a sine and cosine series,

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} [a_n \cos(\frac{nt}{W}) + b_n \sin(\frac{nt}{W})]. \quad (5)$$

If this series is valid, all that is needed are the values for the coefficients. For example, to solve for a_n , multiply both sides of Equation (5) by $\cos(\frac{nt}{W})$ and integrate. (This process uses term-by-term integration which requires special convergence conditions on the series. For example, uniform convergence would be sufficient. We will not deal with the convergence issue in this paper.) After changing the index on the summation,

$$\begin{aligned} \int_{-\pi W}^{\pi W} f(t) \cos(\frac{nt}{W})dt &= \frac{1}{2}a_0 \int_{-\pi W}^{\pi W} \cos(\frac{nt}{W})dt \\ &+ \sum_{m=1}^{\infty} [a_m \int_{-\pi W}^{\pi W} \cos(\frac{mt}{W}) \cos(\frac{nt}{W})dt \\ &+ b_m \int_{-\pi W}^{\pi W} \sin(\frac{mt}{W}) \cos(\frac{nt}{W})dt]. \end{aligned}$$

By property (4), all terms drop out of the right hand side except for the a_n term,

$$\int_{-\pi W}^{\pi W} f(t) \cos(\frac{nt}{W})dt = a_n \pi W.$$

Thus,

$$a_n = \frac{1}{\pi W} \int_{-\pi W}^{\pi W} f(t) \cos(\frac{nt}{W})dt.$$

With similar work the b_n coefficients can be found.

A $2\pi W$ -periodic function $f(t)$ can now be represented by an infinite series. The sine and cosine form of the **Fourier Series** for $f(t)$ is

$$f(t) \sim \frac{1}{2}a_0 + \sum_{n=1}^{\infty} [a_n \cos(\frac{nt}{W}) + b_n \sin(\frac{nt}{W})]. \quad (6)$$

where the Fourier coefficients are defined to be

$$\begin{aligned} a_n &= \frac{1}{\pi W} \int_{-\pi W}^{\pi W} f(t) \cos\left(\frac{nt}{W}\right) dt & \text{for } n = 0, 1, 2, \dots \\ b_n &= \frac{1}{\pi W} \int_{-\pi W}^{\pi W} f(t) \sin\left(\frac{nt}{W}\right) dt & \text{for } n = 1, 2, \dots \end{aligned}$$

Dirichlet's Theorem states that if a function has a finite number of extrema and discontinuities on the interval $[-\pi W, \pi W]$ and if $\int_{-\pi W}^{\pi W} |f(t)| dt < \infty$, then the Fourier series in (6) converges to $f(t)$ at all points where the function is continuous. The series converges to the midpoint of the jumps at points of discontinuity. This pointwise convergence result is what is meant by the \sim symbol used in Equation (6).

Similar work can be done with the orthogonal set $\{e^{int/W}\}_{n \in \mathbf{Z}}$. The complex exponential form of the Fourier series for a $2\pi W$ -periodic function is

$$f(t) \sim \sum_{n=-\infty}^{\infty} c_n e^{int/W}, \quad (7)$$

where

$$c_n = \frac{1}{2\pi W} \int_{-\pi W}^{\pi W} f(t) e^{-int/W} dt.$$

Note that this series could also represent a nonperiodic function on the interval $[-\pi W, \pi W]$.

This last idea can be extended to piecewise smooth $L^1(\mathbf{R})$ function, $\int_{-\infty}^{\infty} |f(t)| dt < \infty$. The following approach illustrates the idea, but is not rigorous. A formal proof would require additional conditions on the function. Define f_W to be the $2\pi W$ -periodic function equal to f on $(-\pi W, \pi W)$. Then from Equation (7)

$$f_W(t) \sim \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\pi W} \int_{-\pi W}^{\pi W} f(x) e^{-inx/W} dx \right) e^{int/W}.$$

Letting $\xi_n = n/W$ and $\Delta\xi_n = \xi_{n+1} - \xi_n = 1/W$,

$$\begin{aligned} f_W(t) &\sim \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\pi W} \int_{-\pi W}^{\pi W} f(x) e^{-i\xi_n x} dx \right) e^{i\xi_n t} W \Delta\xi_n \\ &= \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\pi W}^{\pi W} f(x) e^{-i\xi_n x} dx \right) e^{i\xi_n t} \Delta\xi_n. \end{aligned}$$

Now, as $W \rightarrow \infty$, $f_W \rightarrow f$ and

$$f(t) \sim \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\xi x} dx \right) e^{i\xi t} d\xi.$$

This motivates the definition for the Fourier transform.

Definition 2.1 For $f \in L^1(\mathbf{R})$, the **Fourier transform** of f is

$$\hat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\xi t} dt$$

and the **inverse Fourier transform** is

$$\check{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(\xi) e^{i\xi t} d\xi.$$

We can think of $f(t)$ as a function in the time domain and $\hat{f}(\xi)$ as the functions frequency content. The Fourier Inverse Theorem states that if f and \hat{f} are in $L^1(\mathbf{R})$, then $(\hat{f})^\check{ } = (\check{f})^\hat{ } = f$ almost everywhere. This equality is true everywhere if f is continuous.

The above presents all the Fourier analysis needed for our work.

3 Shannon Sampling Theorem

We are now ready to look at the Shannon Sampling Theorem. This result was known to Borel by 1897 and Shannon introduced it to engineering literature in the 1940's. Independent research on this result was done by others in Russia, Japan and England.

Theorem 3.1 (Shannon Sampling Theorem) *If a continuous function $f(t) \in L^1(\mathbf{R})$ contains no frequencies higher than $\frac{W}{2}$ cycles per second, that is $\hat{f}(\xi)$ is only non-zero on the interval $[-\pi W, \pi W]$, then $f(t)$ is completely determined by a sequence of points spaced $\frac{1}{W}$ seconds apart. Specifically,*

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{W}\right) \text{sinc}(Wt - n),$$

where $\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$. The slowest possible sampling rate, W samples per seconds, is called the **Nyquist rate**.

REMARK. In general an infinite number of curves can be found that will pass through our sample data points $(\frac{n}{W}, f(\frac{n}{W}))$. The unique function comes from the fact that our function's frequency is bounded. If our sample values are sufficiently close in relation to the highest allowed frequency, the function is controlled as far as how fast it can change and thus only one function will pass through the give data points.

PROOF. One of the conditions of our theorem states that $\hat{f}(\xi)$ is only non-zero on the interval $[-\pi W, \pi W]$. From the remark after Equation (7) we can represent $\hat{f}(\xi)$ on the region $[-\pi W, \pi W]$ by a Fourier series,

$$\hat{f}(\xi) = \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\pi W} \int_{-\pi W}^{\pi W} \hat{f}(y) e^{-iny/W} dy \right) e^{in\xi/W} \quad \text{for } \xi \in [-\pi W, \pi W].$$

This formula can be extended to all values of ξ by introducing a characteristic function,

$$\chi_{[-\pi W, \pi W]}(\xi) = \begin{cases} 0 & \text{if } |\xi| > \pi W \\ 1 & \text{if } |\xi| \leq \pi W. \end{cases}$$

Since \hat{f} is zero off of $[-\pi W, \pi W]$,

$$\hat{f}(\xi) = \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\pi W} \int_{-\pi W}^{\pi W} \hat{f}(y) e^{-iny/W} dy \right) e^{in\xi/W} \chi_{[-\pi W, \pi W]}(\xi) \quad \text{for all } \xi .$$

As above, since \hat{f} is zero off of $[-\pi W, \pi W]$,

$$\begin{aligned} \hat{f}(\xi) &= \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\pi W} \int_{-\infty}^{\infty} \hat{f}(y) e^{-iny/W} dy \right) e^{in\xi/W} \chi_{[-\pi W, \pi W]}(\xi) \\ &= \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi W}} (\hat{f})\left(-\frac{n}{W}\right) e^{in\xi/W} \chi_{[-\pi W, \pi W]}(\xi) \\ &= \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi W}} f\left(-\frac{n}{W}\right) e^{in\xi/W} \chi_{[-\pi W, \pi W]}(\xi). \end{aligned}$$

By the Fourier Inverse Theorem, the last equation is true since $f(t)$ is continuous. With the substitution of $-n$ in place of n ,

$$\hat{f}(\xi) = \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi W}} f\left(\frac{n}{W}\right) e^{-in\xi/W} \chi_{[-\pi W, \pi W]}(\xi).$$

Again, since $f(t)$ is continuous

$$\begin{aligned}
f(t) &= (\hat{f})^\vee(t) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi W}} f\left(\frac{n}{W}\right) e^{-in\xi/W} \chi_{[-\pi W, \pi W]}(\xi) \right] e^{i\xi t} d\xi \\
&= \sum_{n=-\infty}^{\infty} f\left(\frac{n}{W}\right) \frac{1}{2\pi W} \int_{-\pi W}^{\pi W} e^{i(t-n/W)\xi} d\xi \\
&= \sum_{n=-\infty}^{\infty} f\left(\frac{n}{W}\right) \frac{1}{\pi(Wt-n)} \frac{e^{i(Wt-n)\pi} - e^{-i(Wt-n)\pi}}{2i} \\
&= \sum_{n=-\infty}^{\infty} f\left(\frac{n}{W}\right) \frac{\sin[\pi(Wt-n)]}{\pi(Wt-n)} \\
&= \sum_{n=-\infty}^{\infty} f\left(\frac{n}{W}\right) \text{sinc}(Wt-n).
\end{aligned}$$

Note that term-by-term integration was once again used in the above work. □

This theorem tells us that if for a continuous function whose Fourier transform has finite support, zero off of $[-\pi W, \pi W]$, all information about the function is known from its value at points spaced $\frac{1}{W}$ seconds apart. In the next section this theorem will be applied to the design of compact discs.

4 Storing Information on a Compact Disc

Let us begin with a continuous analog signal that we wish to record on a compact disc. Since the human ear can only hear in a range up to about 20 kHz, our function can be considered to have bounded frequency. With these assumptions, the necessary conditions for the Shannon Sampling Theorem are met.

To make a compact disc, our analog signal is sampled at a rate of 44.1 kHz. This rate is greater than the Nyquist rate necessary for reproducing the signal. These sampled values are now recorded on the compact disc as 32 bit binary codes, 16 for the right speaker and 16 for the left. These codes represent the amplitude of each sampled point $f\left(\frac{n}{44100}\right)$. This binary number is entered on the disc using a pattern of lands and pits (see Fig. 1). Changes from lands to pits or pits to lands represent a 1. Maintaining the same height represents a 0.

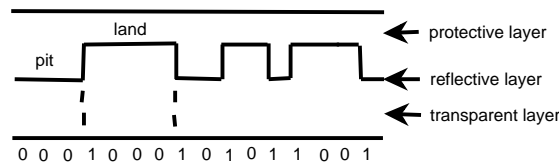


Figure 1: Cross Section of Compact Disc

To play our compact disc (see Fig. 2), a laser light is shown on a half silvered mirror which allows part of the light to pass through and reflects the remaining light towards photo diodes. The light which passes through the mirror then passes through two lenses. The first lens focuses the light and the second lens makes divergent laser beams parallel. After this, the light reaches our disc. The light passes through the protective transparent layer of the disc and is reflected back to the mirror when it hits the reflective layer. The pits, shown in Figure 1, are about one quarter wavelength deep. After this light is reflected back to the mirror, it is sent towards the photo diodes. The light that hit the disc now mixes with the light originally reflected off the mirror. The depth of the pits causes the light which hit these spots to be out of phase with the originally reflected light. This interference means little light reaches the diodes. Mathematically this is seen by adding two sine signals whose phase is off by half a wavelength. On the other hand, light which hit the lands will be in phase with the reflected light and much more light will hit the diodes for these locations. In this way, the samples recorded on the disc are read.

The compact disc player now has the sampled values. Jumps between the amplitudes sampled would produce high frequencies. These frequencies are above the audio range but the power contained in them can overload the player amplifier. Thus, the frequencies above the audio range must be reduced to a lower power level.

To shorten these jumps and thus reduce the power contained in these frequencies, a digital transversal filter (DTF) over samples our signal by a factor of four and interpolates. This interpolation can be understood best if we think of the filter as holding 96 elements at any time. Each of the 96 cells contains a value considered to be $\frac{1}{4(44100)}$ seconds apart. For the 24 cells which line up with the actual sampled data, the sampled values are recorded in the cells. To fill in the remaining cells, three between each pair of actual sampled data, the filter uses the pattern of the 24 known

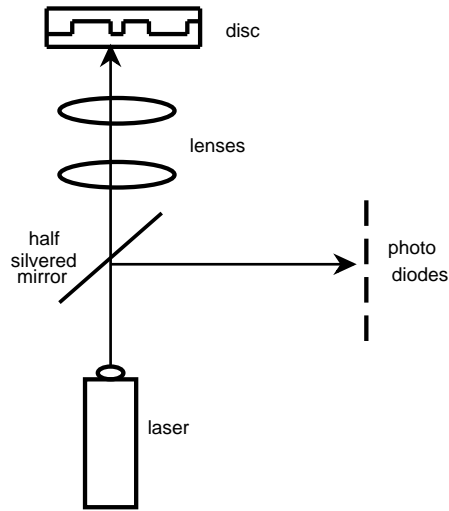


Figure 2: Reading the Compact Disc

cells to compute the new values. In other words, the filter finds a curve, with desirable properties which fits the 24 sampled values and uses this curve to compute the remaining needed values. In other words, the filter produces three extra readings between each pair of original sampled data. This amounts to producing data with smaller jumps between the values.

This new data is now sent to a digital-to-analog converter (DAC). The DAC takes each data value and generates an electric current of that magnitude until the next code arrives $\frac{1}{4(44100)}$ seconds later. This produces a staircase function which approximates the original analog signal.

To further smooth the staircase function before it is sent to the speakers the signal is passed through a lowpass filter which further reduces the high frequency content to a level which will not harm the player. The function of this lowpass filter can be explained mathematically. The basic idea is to cut off frequencies outside the audio range. This is like multiplying the Fourier transform of the staircase function by a characteristic function on $[-\pi W, \pi W]$, that is a function which is one on the interval $[-\pi W, \pi W]$, and 0 elsewhere. This multiplication takes place in the frequency domain. In terms of the time domain, this is equivalent to multiplying our sample values by *sinc* functions.

These last ideas go deeper into Fourier Analysis. The interested reader should look at the definition of a convolution integral, the Convolution The-

orem and properties of the Dirac Delta function. For a more of the engineering details on lowpass filters the reader can refer to pages 514-526 of [6]. For the necessary background mathematics for the lowpass filter the reader can refer to pages 82-87 and 249-250 of [7] or material in a similar Fourier text.

In practice there are a few ideas which work in theory but which are slightly altered in actual circuits. References [2], [3] and [6] give more details. The central idea behind reproducing our signal is done by applying the result of the Shannon Sampling Theorem.

5 Conclusion

In this paper some Complex Analysis and Fourier Analysis were used to understand the Shannon Sampling Theorem. Series have many applications in current technology. Series are used to store information on computers, to send signals from satellites, to record fingerprints digitally (forgive the pun) and to gather medical images with MRI's. This paper can introduce students to one application. A list of some possible further readings on this topic are listed below. References [2] and [3] give more details to the design of compact discs. The other references listed give overviews of much of the work in Sampling Theory.

References

- [1] P. Butzer, W. Splettstößer, L. Stens, *The Sampling Theorem and Linear Prediction in Signal Analysis*, Jahresbericht der Deutschen Mathematiker-Vereinigung, **90** (1988) 1-70
- [2] M. Carasso, J. Peek, J. Sinjou, *The Compact Disc Digital Audio system*, Phillips Technical Review, **40** (1982), 151-156.
- [3] D. Goedhart, R. van dan de Plassche and E. Stikvoot, *Digital-to-analog conversion in playing a Compact Disc*, Phillips Technical Review, **40** (1982), 174-180.
- [4] J. Higgins, *Five Short Stories About the Cardinal Series*, Bulletin of the AMS, **12** (1985), 45-89.

- [5] A. Jerri, *The Shannon Sampling Theorem - Its Various Extensions and Applications: A Tutorial Review*, Proceedings of the IEEE, **65**, (1977), 1565-1596.
- [6] A. Oppenheim, A. Willsky and I. Young, *Signals and Systems*, Prentice-Hall,(1983).
- [7] J. Walker, *Fourier Analysis*, Oxford University Press, (1988).