SCIENCE CHINA
Information Sciences

• LETTER •

October 2020, Vol. 63 209301:1–209301:3
https://doi.org/10.1007/s11432-019-2829-5

The study of signal reconstruction: an investigation independent of Shannon-Nyquist theorems

Guangjun SHAN[1,2]

[1]School of Information Engineering, East China University of Technology, Nanchang 330013, China;
[2]School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, China

Received 17 July 2019/Revised 2 September 2019/Accepted 12 March 2020/Published online 1 June 2020

Citation Shan G J. The study of signal reconstruction: an investigation independent of Shannon-Nyquist theorems. Sci China Inf Sci, 2020, 63(10): 209301, https://doi.org/10.1007/s11432-019-2829-5

Dear editor,

The sampling theorem is a keystone to modern information communication. Nyquist's early study [1] on determining the bandwidth requirements for transmitting information laid the foundations for later advances by Shannon [2]. Of course, many other brilliant scientists, e.g., Whittaker, Hartley, Kotelnikov, Gabor, also made solid contributions to modern communication theory [3, 4]. Shannon [2] summarized that if a signal function $s(t)$ contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced $\frac{1}{2}W$ seconds apart, which is well-known as the Shannon-Nyquist sampling theorem. For many years, sampling by $\frac{1}{2}W$ seconds apart leads people to believe that sampling a signal uniformly spaced within its period T is a natural law for the reconstruction of a continuous waveform signal $s(t)$. Although scientists have stated that the non-uniform sampling (NUS) is applicable to signal reconstruction, such as Yen's interpolation [5], Marvasti's iterative method [6], Jenq's periodic NUS [7], these schemes can only work conditionally as they followed Shannon-Nyquist regime. In Shannon-Nyquist regime, it is believed that non-uniform samples will cause aliasing when the frequency domain convolution is used for signal reconstruction and information recovery. While sometimes it is true, it is important that we do not stop delving into the essence of this issue.

The methodology used in Shannon-Nyquist theorem is based on the frequency convolution and sinc function in the form of $\frac{\sin 2\pi Wt}{2\pi Wt}$ [2]. Applying sinc function and frequency domain convolution is valid, but not the sole methodology for signal reconstruction. Holography is an excellent example of another means of signal reconstruction. $\frac{\sin 2\pi Wt}{2\pi Wt}$, a unity at $t = 0$ and zero at $t = \frac{n}{2W}$, passes through the sampling points separated $\frac{1}{2W}$ seconds apart, which means the sampling points can only be uniformly taken at $t = \frac{n}{2W}$ when applying $\sum_{n=-\infty}^{+\infty} x_n \frac{\sin \pi(2Wt-n)}{\pi(2Wt-n)}$ to signal reconstruction [2], where x_n denotes nth sample, $n \in Z^+$. Such a feature of sinc function leads to the uniform sampling (US) compulsorily used in Shannon-Nyquist regime. It will cause aliasing if adopting non-uniform sampling (NUS).

To address the universality of NUS, we begin the investigation using Fourier series. Unless otherwise specified, the word "signal" refers to the periodic bandwidth-limited signal in this study, and we assume there is a priori knowledge jointly possessed by source (S) and destination (D), in which the composition of frequency is known.

It is well-known that Fourier series can decompose a periodic time function $s(t)$ into a series of orthogonal frequencies that comprise it. In this study, $s(t)$ has a double meanings, which can refer

Email: acosmos@126.com

to a function $s(t)$, or a signal $s(t)$, and both are commutable with each other.

$$s(t) = a_0 + \sum_{n=1}^{\infty}(a_n \cos n\omega_1 t + b_n \sin n\omega_1 t), \quad (1)$$

where a_0 is a constant referring to a direct current component of signal $s(t)$. a_n and b_n may refer to the cosine and sine coefficients of function $s(t)$, either, the amplitude of quadrature and inphase frequency components of signal $s(t)$. ω_1 is the fundamental angular frequency of $s(t)$. Let f_1 and $f_{\max}$ be the fundamental frequency and the highest frequency of $s(t)$, and we have $f_1 = \frac{\omega_1}{2\pi}$, $f_{\max} = \frac{n\omega_1}{2\pi}$. Define R_f be $\frac{f_{\max}}{f_1}$, the signal upper-lower frequency ratio (ULFR), obviously, $R_f = n$. For a periodic band-limited periodic signal, n is no longer an infinite value, and it depends on how many frequency components $s(t)$ includes.

Signal bandwidth is usually a common prior knowledge shared by S and D for conventional communication system. Correspondingly, ω_1 and $n\omega_1$, are known to a periodic band-limited signal. Furthermore, if the values of coefficients of cosines and sines (amplitudes of quadrature and in-phase components) can be obtained, then $s(t)$ will be uniquely determined. This implies that the information in $s(t)$ is determined by these coefficients (amplitudes). All possible waveforms of $s(t)$ will not go beyond the scope of superpositions that is made up of n finite orthogonal frequency components with the specified amplitudes, as described by (1), but n becomes a finite integer.

Given a definite instant of time, for instance, t_1, all harmonics will get their respective values, such as $\cos\omega_1 t_1, \sin\omega_1 t_1, \cos 2\omega_1 t_1, \sin 2\omega_1 t_1, \ldots, \cos n\omega_1 t_1, \sin n\omega_1 t_1$. Additionally, the amplitude value of $s(t_1)$ can be obtained by measuring (sampling) the signal $s(t)$ at time t_1. As the result, a sample is formed, the equation is as follows:

$$\begin{aligned} s(t_1) = {} & a_0 + a_1 \cos\omega_1 t_1 + b_1 \sin\omega_1 t_1 + b_2 \sin 2\omega_1 t_1 \\ & + \cdots + a_n \cos n\omega_1 t_1 + b_n \sin n\omega_1 t_1. \end{aligned} \quad (2)$$

Eq. (2) is a linear equation with $2n+1$ unknown coefficients: a_0, a_1, a_2, $b_2, \ldots, a_n$, b_n. Obviously, it is impossible to solve for the $2n+1$ unknown constant coefficients by a sole equation. Thus, $s(t)$ should be sampled $2n+1$ times at different time points, marked as $s(t_1)$, $s(t_2)$, $s(t_3), \ldots, s(t_{2n+1})$, which leads to $2n+1$ linear equations, listed in the form of a matrix as shown by (3). However, one question arises: Is there a unique solution regarding the $2n+1$ coefficients, regardless of the choice of the $2n+1$ sampling points? If the answer is positive, we can implement the sampling at arbitrary time points, regardless of how these samples are distributed along with the time axis.

$$\begin{bmatrix} a_0 \cdot 1 + a_1 \cos\omega_1 t_1 + b_1 \sin\omega_1 t_1 + \cdots + a_n \cos n\omega_1 t_1 + b_n \sin n\omega_1 t_1 \\ a_0 \cdot 1 + a_1 \cos\omega_1 t_2 + b_1 \sin\omega_1 t_2 + \cdots + a_n \cos n\omega_1 t_2 + b_n \sin n\omega_1 t_2 \\ \vdots \\ a_0 \cdot 1 + a_1 \cos\omega_1 t_{2n+1} + b_1 \sin\omega_1 t_{2n+1} + \cdots + a_n \cos n\omega_1 t_{2n+1} + b_n \sin n\omega_1 t_{2n+1} \end{bmatrix} = \begin{bmatrix} s(t_1) \\ s(t_2) \\ \vdots \\ s(t_{2n+1}) \end{bmatrix}, \quad (3)$$

$$A = \begin{bmatrix} 1 & \cos\omega_1 t_1 & \sin\omega_1 t_1 & \cdots & \cos n\omega_1 t_1 & \sin n\omega_1 t_1 \\ 1 & \cos\omega_1 t_2 & \sin\omega_1 t_2 & \cdots & \cos n\omega_1 t_2 & \sin n\omega_1 t_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos\omega_1 t_{2n+1} & \sin\omega_1 t_{2n+1} & \cdots & \cos n\omega_1 t_{2n+1} & \sin n\omega_1 t_{2n+1} \end{bmatrix}. \quad (4)$$

Proof. Express (3) by the form of matrix equation $AX = b$, where A is a $(2n+1) \times (2n+1)$ square matrix as shown by (4), $X = [a_0 \ \ a_1 \ \ b_1 \ \ \cdots \ \ a_n \ \ b_n]^{\mathrm{T}}$, and $b = [s(t_1) \ \ s(t_2) \ \ \cdots \ \ s(t_{2n+1})]^{\mathrm{T}}$. As A is a square matrix in which the row vectors consist of a series of orthogonal trigonometric functions, it can be concluded that A is a full rank and invertible matrix. Thus, $AX = b$ has a unique solution for X. Additionally, a given Fourier series $s(t)$ can be regarded as a point or a vector in the Hilbert space L^2 whose basis consists of a set of complete orthonormal basis $\{1, \cos\omega t, \sin\omega t, \cos 2\omega t, \sin 2\omega t, \ldots, \cos n\omega t, \sin n\omega t\}$. Consequently, the coordinate X of L^2 would not change, regardless of how the samples of $s(t)$ distribute along with the time axis.

Thus, the following conclusion can be made.

Theorem 1. Given a signal $s(t)$ whose period T and the highest frequency $f_{\max}$ are known, $R_f = n$, provided that $2n+1$ samples are collected within T, both US and NUS are capable of reconstructing the original signal $s(t)$.

It has long been customary to believe that US is an indispensable prerequisite for signal reconstruction. However, our proof shows it is a misunderstanding. NUS is unequivocally valid for signal reconstruction, akin to its US counterpart. Unlike the conventional signal reconstruction using convolution and sinc operations, our methodology accomplishes the signal reconstruction by solving linear equations, which naturally supports NUS. To distinguish our signal reconstruction methodology from Shannon sampling theorem, we name it linear equation reconstruction (LER).

For simplicity, assume $a_0 = 0$, thus only $2n$ samplings are required. Following the proof of LER, a reasonable presumption rises: because $s(t)$ can be reconstructed from its $2n$ samples regardless of the distribution of sampling points, it makes sense to reconstruct $s(t)$ by a portion itself on which the $2n$ samples are concentrated. To confirm this presumption, we make the following hypothesis.

(1) $2n$ samples are completed within time τ, $\tau < T$.

(2) The recognition accuracy with regard to samples meets the requirement of signal reconstruction.

Let $\sigma(\tau)$ denote a portion, or few fragments of $s(t)$, where τ is the persistent time of $\sigma(\tau)$, we have then the following conclusion.

Theorem 2. $\sigma(\tau)$ is capable of reconstructing the entire original $s(t)$, as long as it can accommodate $2n$ (or $2n+1$ if $a_0 \neq 0$) samples.

It is not difficult to understand Theorem 2 as Theorem 1 shows that NUS is qualified to recover the original signal as long as $2n$ samples are provided. Of course these samples may concentrate on a portion or a few separated fragments of $s(t)$. $s(t)$ consists of a series of orthogonal frequency components that are simultaneously superposed and entangled together, thus $\sigma(\tau)$ may include all information of $s(t)$. The remaining part of $s(t)$ appears to be trivial for reconstructing the original signal, it would not affect the signal reconstruction even if the remaining part of $s(t)$ has not yet been received.

Suppose the amplitude (a_n, b_n) of each frequency component can be stratified to m hierarchical levels, thus, each pair of them (a_n, b_n) could represent $2\log_2 m$ bits. Let C_s denote the information transmission rate of a signal, because $R_f = n$, then,

$$C_s = \frac{2R_f \log_2 m}{T} \text{ (bps)}. \tag{5}$$

Considering $f_1 = 1/T$, and $R_f = f_{\max}/f_1$, thus Eq. (5) can be written as

$$C_s = 2nf_1 \log_2 m = 2f_{\max} \log_2 m \text{ (bps)}. \tag{6}$$

Eq. (6) shows that under LER methodology, signal's information transmission rate is only determined by twice the highest frequency it contains. Assume $2n$ samplings are completed within time τ, where $\tau = \frac{T}{k}$, and $k > 1$, $\frac{T}{k}$ can be used to replace the T in (5), as the result, C_s will rise to

$$C_s = 2kf_{\max} \log_2 m \text{ (bps)}, \tag{7}$$

where k is only limited by the recognition accuracy of sample. Eq. (7) shows that the information transmission rate can be effectively improved by completing all the required samplings within a shorter duration τ than the signal period T. Sometimes, a signal perhaps has an extremely long duration, it makes sense to "predict" the future signal even if only a portion of this signal is received. Because we may plot the unreceived part by sampling the received signal, providing our prior knowledge regarding this signal's spectrum is correct. This reminds us of holography: when a hologram is broken, each piece of broken hologram still can be used to reconstruct the entire image of the subject photographed by holography. Considering the similarity with holography, we call the signal reconstruction phenomenon described by Theorem 2 as holographic sampling reconstruction (HSR).

This study aims to investigate the signal reconstruction issue, the findings can be used in conjunction with compressed sensing theorem [8]. We proved that NUS is inevitably feasible for signal reconstruction as its US counterpart by proposed LER methodology. Our study shows that one can still recover the entire signal from a fraction of the original. Consequently, it is unnecessary to occupy the whole signal period for the sampling process, which led to the HSR scheme proposed.

References

1 Nyquist H. Certain topics in telegraph transmission theory. Trans Am Inst Electr Eng, 1928, 47: 617–644
2 Shannon C E. A mathematical theory of communication. Bell Syst Tech J, 1948, 27: 379–423
3 Hartley R. The transmission of information. Bell Syst Tech J, 1928, 3: 535–564
4 Gabor D. Theory of communication. J Inst Electr Eng, 1946, 26: 429–457
5 Yen J L. On nonuniform sampling of bandwidth-limited signals. IRE Trans Circ Theor, 1956, 3: 251–257
6 Marvasti F, Analoui M, Gamshadzahi M. Recovery of signals from nonuniform samples using iterative methods. IEEE Trans Signal Process, 1991, 39: 872–878
7 Jenq Y C. Digital spectra of nonuniformly sampled signals: fundamentals and high-speed waveform digitizers. IEEE Trans Instrum Meas, 1988, 37: 245–251
8 Zhao J, Bai X, Tao R. Improved RIP-based performance guarantees for multipath matching pursuit. Sci China Inf Sci, 2018, 61: 102303