REPRODUCIBLE RESEARCH: A CASE STUDY OF SAMPLING SIGNALS WITH FINITE RATE OF INNOVATION

Pina Marziliano

School of Electrical and Electronic Engineering
Nanyang Technological University
50 Nanyang Avenue
Singapore 639798

ABSTRACT

A case study on reproducible research in sampling theory of signals containing a finite rate of innovation is the topic of this paper. By building a solid research which is furthermore reproducible enables the researcher to build intuition in a research area and to progress at a much faster pace. Here, we show that the founding problem of sampling and exact reconstruction of periodic streams of Dirac pulse will be the basis of the sampling theory for signals with finite rate of innovation. The sampling theory can be extended to other signals such as piecewise polynomials, bandlimited signals with additive shotnoise and the sum of bandlimited signals with piecewise polynomial signals. It is shown that the implementation is based on the one for streams of Dirac pulses, thus making the new research reproducible as well.

Index Terms— Sampling, signals with finite rate of innovation, reproducible research, theory, applications.

1. INTRODUCTION

Reproducible research was first baptised as *literate programming*[1] by Donald Knuth in 1984, which he defined as a robust and user friendly style of writing which combines a programming language with a documentation language. In 1995, Donoho et al. [2] defined reproducible research such that all the code underlying the figures and the tables in a research publication be made available together with documentation of both the tools and the environment and finally the latter would be accessible through the internet with anonymous access. A system called ReDoc [3] was established by Claerbout in Stanford in order for all research publications and books originating from his group to be completely reproducible. The interest in reproducible research in signal processing was revitalized in a brief note by Barni and Perez [4].

The advantages of reproducible research are numerous. For instance, young researchers, novel to a research field, can reap the benefits of the findings from their peers. Reproducible research enables more people to work on the same research topic as data and experiments can be shared and therefore allowing to progress at a faster pace. Finally collaborations become easier between research groups. The flipside to reproducible research is that it may affect the creativity of a researcher. Furthermore, it may raise issues with respect to intellectual property and patents.

In this paper, reproducible research is discussed in the context of sampling theory for signals with a finite rate of innovation as described in [5]. These signals have a parametric representation and contain a finite number of degrees of freedom. The sampling theory for signals with finite rate of innovation began with the ”simple” problem of finding the transition or ”innovation” points of a bilevel signal. This was achieved by solving a system of linear equations obtained from the inner products between the bilevel signal and a box sampling kernel. When the transition points were too close apart, a hat sampling kernel was used in the inner product and the solution was obtained by solving a quadratic system of equations, and so forth. By differentiating the bilevel signal, it was noticed that the transition or innovation points were described by a stream of Dirac pulses. Therefore the problem was reduced to exactly reconstructing a stream of Dirac pulses from the set of samples obtained from the inner products. Since any piecewise polynomial signal of degree R can be obtained by integrating $R + 1$-times the stream of Dirac pulses, this would lead to a solution to the original problem.

The interest in sampling and reconstructing stream of Dirac pulses and piecewise polynomial signals lies in the fact that these are not bandlimited signals and the usual Kotelnikov-Whittaker-Shannon[6] sampling theorem does not apply. Many researchers have gained keen interest for the theory of signals with finite rate of innovation [7], [8], [9],[10] as these signals are found in a wide range of applications, for instance, in communication and biomedical systems.

This paper is organised as follows: Section 2 presents the solution to the problem on sampling and exact reconstruction of stream of Dirac pulses which sets the foundation for Section 3 on sampling and exact reconstruction of piecewise polynomial signals. The focus will be more on the practical aspects of the sampling theory, so as to accompany the previously published papers [5], [11] thus contributing to reproducible research. Section 4 will present two extensions of the theory with respective applications and finally conclusions are drawn in Section 5.

2. THE FOUNDATION: SAMPLING AND EXACT RECONSTRUCTION OF PERIODIC STREAMS OF DIRAC PULSES

Let us begin by defining the τ-periodic signal, $x_{\text{D}}(t)$, which is a stream of K weighted Dirac pulses

$$\begin{aligned} x_{\text{D}}(t) &= \sum_{n\in\mathbb{Z}} c_n\,\delta(t-t_n) && (1)\\ &\quad \text{where } t_{n+K}=t_n+\tau \text{ and } c_{n+K}=c_n \\ &= \sum_{k=0}^{K-1} c_k \sum_{n\in\mathbb{Z}} \delta(t-t_k-n\tau) \\ &= \sum_{k=0}^{K-1} c_k \frac{1}{\tau}\sum_{m\in\mathbb{Z}} e^{i(2\pi m(t-t_k)/\tau)} && (2)\\ &\quad \text{from Poisson's summation Formula} \\ &= \sum_{m\in\mathbb{Z}} X_{\text{D}}[m]\, e^{i2\pi mt/\tau} && (3) \end{aligned}$$

where

$$X_{\text{D}}[m] = \frac{1}{\tau}\int_0^{\tau} x_{\text{D}}(t)e^{-i\frac{2\pi mt}{\tau}}\,dt = \frac{1}{\tau}\sum_{k=0}^{K-1} c_k\, e^{-i\frac{2\pi m t_k}{\tau}},\ m\in\mathbb{Z} \quad (4)$$

are the corresponding continuous-time Fourier series (CTFS) coefficients which completely define the periodic signal $x_{\text{D}}(t)$. From Eq. (4), the number of degrees of freedom, in one period τ, is K from the locations $\{t_k\}_{k=0,\dots,K-1}$ and K from the weights $\{c_k\}_{k=0,\dots,K-1}$ therefore the rate of innovation of $x_{\text{D}}(t)$ is $\rho = 2K/\tau$.

Theorem 1 *Consider a $\tau-$periodic stream of K weighted Dirac pulses $x_D(t)$ as defined in Eq. (1) with rate of innovation $\rho = \frac{2K}{\tau}$. Consider a sinc[1] sampling kernel $h_B(t) = B\,\text{sinc}(Bt)$ with bandwidth $2B\pi$ where B is greater than or equal to the rate of innovation ρ, $B \geq \rho$. If the lowpass filtered signal, $y(t) = (h_B * x_D)(t)$ is sampled at N uniform locations $t = nT, n = 0,\dots,N-1$, where $T = \frac{\tau}{N}$, $N \geq 2M+1$ and $M = \lfloor\frac{B\tau}{2}\rfloor$, then the samples of the uniform set*

$$y(nT) = y_n =< h_B(t-nT), x_D(t) >, \quad n = 0,\dots,N-1 \quad (5)$$

where $< \cdot,\cdot >$ denotes the inner product[2], are sufficient to perfectly reconstruct $x_D(t)$.

[1] The sinc definition used here is $\text{sinc}(t) = sin(\pi t)/\pi t$.

[2] Note that the inner product is defined by $< f(t), g(t) >= \int_{-\infty}^{\infty} f(t)g^*(t)\,dt$.

The proof of the theorem is given in [11]. Here, instead, we will show the step by step implementation and the Matlab code can be found in [12]. There are three steps involved: first we need to generate the signal, second we need to implement the sampling procedure and third the reconstruction method.

2.1. Generation of a periodic stream of Dirac pulses

For some research problems where data is generated randomly, the initial parameters as well as the generated data must be saved in a data file in order to reproduce the exact figures as in the published paper. Therefore even if the programming code is written in a user friendly manner it does not necessarily imply that a research paper is reproducible which is one of the problems faced when attempting to reproduce the figures in [5]. Reproducible research is feasible if one starts off with the correct frame of mind.

First the period and the number of Dirac pulses need to be fixed. The period of the signal is assumed to be a positive integer for discrete-time signals and a positive real value which may be normalised and thus vary between 0 and 1 for continuous-time signals. Then the locations and the weights of the Dirac pulses are generated according to a probability distribution. Thus after the initialisation and generation step we have the following parameters:

1. τ and K are fixed integers and represent the period and the number of Dirac pulses, respectively.
2. $t_k, k = 0,\dots,K-1$, represent the locations of the Diracs and are randomly generated following a unifom distribution such that $0 \leq t_{k-1} < t_k \leq \tau$.
3. $c_k, k = 0,\dots,K-1$ are the weights of the Dirac pulses located at t_k and are randomly generated following a Normal(0,1) distribution.
4. $\rho = \frac{2K}{\tau}$ is the rate of innovation of the stream of Dirac pulses.

2.2. Sampling a periodic stream of Dirac pulses

In the sampling step, we need to first determine the sampling parameters according to the conditions in Theorem 1 and then calculate the uniform sample set of values:

1. Take $B = \rho = \frac{2K}{\tau}$ the bandwidth parameter of the sinc sampling kernel, $h_B(t)$ and calculate $M = \lfloor\frac{B\tau}{2}\rfloor = K$.
2. Take the number of samples $N \geq 2M+1$ such that the sampling interval $T = \tau/N$ is an integer.
3. Calculate the sample values from the inner products $y_n =< h_B(t-nT), x_{\text{D}}(t) >, n = 0,\dots,N-1$.

2.3. Reconstruction of a periodic stream of Dirac pulses

There are two main parts in the reconstruction method. First, from the N samples y_n we need to determine at least $2M$ spectral values $X_{\text{D}}[m], m = -M, \ldots, M$. The second step consists in determining the annihilating filter coefficients $A[m]$ where calculating the roots of the $Z-$transform will lead to the locations of the Dirac pulses. Finally, the locations together with the M spectral coefficients will determine the weights of the Dirac pulses. The reconstruction procedure is as follows:

1. Determine $X_{\text{D}}[m], m = -M, \ldots, M$ by solving the following system of equations $y_n = \sum_{m=-M}^{M} X_{\text{D}}[m]\, e^{i\frac{2\pi mn}{N}}, n = 0, \ldots, N-1.$

2. Formulate a Toeplitz matrix with the spectral coefficients $X_{\text{D}}[m], m = -M, \ldots, M$
$$\mathbf{X}_{\text{Toep}} = \begin{bmatrix} X_{\text{D}}[0] & X_{\text{D}}[-1] & \cdots & X_{\text{D}}[-M] \\ X_{\text{D}}[1] & X_{\text{D}}[0] & \cdots & X_{\text{D}}[-(M-1)] \\ & & \ddots & \\ X_{\text{D}}[M] & X_{\text{D}}[M-1] & \cdots & X_{\text{D}}[0] \end{bmatrix}.$$

3. Find the filter coefficients $\mathbf{A} = [A[0], \ldots, A[M]]^T$ that will annihilate the spectral coefficients, that is, by solving the following system of equations $\mathbf{X}_{\text{Toep}} \cdot \mathbf{A} = 0$.

4. Find the roots of the polynomial made up of the annihilating filter coefficients $\mathbf{A} = [A[0], \ldots, A[M]]^T$. These correspond to $u_k = e^{-i\frac{2\pi t_k}{\tau}}, k = 0, \ldots, M-1$ which leads to the locations $t_k, k = 0, \ldots, M-1$ of the Diracs.

5. Find the weights $c_k, k = 0, \ldots, M-1$ of the Dirac pulses by solving the following system of M equations $X_{\text{D}}[m] = \frac{1}{\tau}\sum_{k=0}^{M-1} c_k\, u_k^m, m = 0, \ldots, M-1$

For the continuous-time case the errors involved in the root finding will be more permissible than those for the discrete-time case.

3. BUILDING ON THE FOUNDATION: SAMPLING AND EXACT RECONSTRUCTION OF PERIODIC PIECEWISE POLYNOMIAL SIGNALS

Reproducible research for streams of Dirac pulses will facilitate the development of the programming code for the sampling and exact reconstruction of piecewise polynomial signals.

Consider a periodic piecewise polynomial signal $x(t)$ with K pieces of degree R which is obtained by integrating $R+1$ times a periodic stream of Dirac pulses $x_{\text{D}}(t)$. By differentiating $R+1$ times the given piecewise polynomial signal we obtain a stream of $(R+1)K$ Diracs, that is, $x_{\text{D}}(t) = x^{(R+1)}(t)$. Therefore the degrees of freedom are $(R+1)K$ from the locations and $(R+1)K$ from the weights, thus the rate of innovation is $\rho = \frac{2(R+1)K}{\tau}$.

The sampling stage remains the same except for the value of $B = \rho = \frac{2(R+1)K}{\tau}$ and thus accordingly $M = \lfloor \frac{B\tau}{2} \rfloor = (R+1)K$. Next the samples are obtained by taking the innerproducts between the sampling kernel and the piecewise polynomial signal, that is, $y_n =< h_B(t-nT), x(t) >, n = 0, \ldots, N-1$, where $N \geq 2M+1$ such that $T = \tau/N$ is an integer value.

In the reconstruction stage, Step 1 gives the $2M+1$ spectral components $X[m]$ of the piecewise polynomial signal, from which we compute the $2M+1$ spectral components of the periodic stream of Diracs according to the following expression $X^{(R+1)}[m] = X_{\text{D}}[m] = (i2\pi m/\tau)^{R+1}\, X[m], m \in [-M, M]$. The next steps are the same till Step 5 which will give the stream of Diracs and thus by integrating $(R+1)$ times the stream of Dirac pulses, the piecewise polynomial signal is recovered.

4. EXTENDED THEORY AND APPLICATIONS

By adopting a reproducible research frame of mind, the above sampling theory can be extended to signals that are the sums of bandlimited signals with stream of Dirac pulses [11] and the sums of bandlimited signals with piecewise polynomials [13].

4.1. Bandlimited signals with additive shotnoise

Bandlimited signals with additive shot noise are modeled as the sum of a bandlimited signal with a stream of Dirac pulses. In Section 2 it was shown that a periodic stream of K Dirac pulses can be perfectly reconstructed from $2K$ contiguous CTFS coefficients that were obtained from a uniform set of samples of the lowpass filtered approximation of the stream of Dirac pulses. Therefore based on the foundation, a sampling and exact reconstruction method is developed for bandlimited signals with additive shot noise.

Consider a τ-periodic signal, $x(t)$, defined as the sum of a τ-periodic L-bandlimited signal, $x_{\text{BL}}(t)$, with a τ-periodic stream of K Dirac pulses $x_{\text{D}}(t)$, that is, $x(t) = x_{\text{BL}}(t) + x_{\text{D}}(t)$, where an L-bandlimited signal $x_{\text{BL}}(t)$ is such that its CTFS $X_{\text{BL}}[m] = 0, \forall m \notin [-L, L]$ and the stream of K Dirac pulses, $x_{\text{D}}(t)$, is defined in Eq. (1). The rate of innovation is given by $\rho = \frac{2L+1+2K}{\tau}$ where $(2L+1)/\tau$ and $2K/\tau$ are the number of degrees of freedom of the bandlimited signal and the stream of K Dirac pulses, respectively.

In the sampling step, we take $B = \frac{2(L+2K)}{\tau} \geq \rho$ thus $M = \lfloor \frac{B\tau}{2} \rfloor = L + 2K$ and the samples are $y_n =< h_B(t-nT), x(t) >, n = 0, \ldots, N-1$, where $N \geq 2M+1$ such that $T = \tau/N$ is an integer value.

As for the reconstruction, similar to Section 2 the bandlimited signal with additive shot noise, $x(t)$, will be reconstructed from a contiguous set of its CTFS coefficients, $X[m], m \in [-M, M]$ with $M = L + 2K$ defined by

$$X[m] = \begin{cases} X_{\text{BL}}[m] + X_{\text{D}}[m] & \text{if } m \in [-L, L] \\ X_{\text{D}}[m] & \text{if } m \notin [-L, L] \end{cases} . \quad (6)$$

Recall that the periodic stream of K Dirac pulses $x_{\text{D}}(t)$ is perfectly recovered from any $2K$ contiguous frequency values $X_{\text{D}}[m]$, thus from Equation (6) take $X_{\text{D}}[m] = X[m], m \in [L + 1, L + 2K]$ and reconstruct the stream of Diracs following the steps in Section 2.3. Finally, the bandlimited signal is obtained by the CTFS coefficients $X_{\text{BL}}[m] = X[m] - X_{\text{D}}[m], m \in [-L, L]$ and then calculating the inverse CTFS.

Removing shotnoise from an old record is an example application of the sampling and exact reconstruction of bandlimited signals with additive shot noise. A reproducible research webpage containing the programming codes can be downloaded from [12].

4.2. Sum of Bandlimited signals with piecewise polynomial

By combining the methods in Section 3 and Section 4.1 the sampling and reconstruction of the sum of a bandlimited signal with a piecewise polynomial signal was developed. These types of signals can be used to model electrocardiogram signals which typically contain three parts: P wave, QRS complex, and T wave. In order to preserve the diagnostic information of the ECG signal, the QRS complex has to be well preserved and is modelled as a piecewise linear signal. By subtracting the QRS complex from the original signal, the remaining part can be modelled as a bandlimited signal. Sampling theory on ECG signals was investigated in [13] in the context of a compression application via reproducible research.

5. CONCLUSIONS

Reproducible research is not about not reinventing the wheel but about showing how the wheel was built and inspire the researcher to invent a better and stronger wheel. Reproducible research can go beyond providing code which enables to generate the figures found in a research paper. For instance by creating a Graphic User Interface, as in [14], it enables the researcher to build intuition for the novel research area.

ACKNOWLEDGMENTS

The author would like to thank Mukesh Rao Engla Syam and Sravanthi Vullikanti for developing the second version of the Graphic User Interface on sampling signals with finite rate of innovation which include nonuniform splines, piecewise polynomial signals, bandlimited with additive shotnoise, as well as the sum of bandlimited signals with piecewise polynomial signals.

6. REFERENCES

[1] D. E. Knuth, “Literate programming,” *Computer Journal*, vol. 27, no. 2, pp. 97–111, 1984.

[2] J. Buckheit and D. Donoho, *WaveLab and Reproducible Research*, Springer and Verlag, 1995.

[3] M. Schwab, N. Karrenbach, and J. Claerbout, “Making scientific computations reproducible,” *Computing in Science and Engineering*, vol. 2, pp. 61–67, Nov–Dec 2000.

[4] M. Barni and F. Perez-Gonzalez, “Pushing science into signal processing,” *IEEE Signal Processing Magazine*, vol. 22, pp. 120–119, 2005.

[5] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Transactions Signal Processing*, pp. 1417–1428, June 2002.

[6] J. J. Benedetto and P. J. S. G. Ferreira, Eds., *Modern Sampling Theory: Mathematics and Applications*, Birkhäuser, Boston, 2001.

[7] J. Kusuma, I. Maravic, and M. Vetterli, “Sampling With Finite Rate of Innovation: Channel and Time Estimation for UWB and GPS,” *Proc. ICC*, vol. 5, pp. 3540–3544, 2003.

[8] I. Maravic and M. Vetterli, “Sampling and Reconstruction of Signals with Finite Rate of Innovation in the Presence of Noise,” *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2788–2805, 2005.

[9] P. Dragotti, M. Vetterli, and T. Blu, “Exact sampling results for signals with finite rate of innovation using strang-fix conditions and local kernels,” *Proc. IEEE ICASSP*, March 2005.

[10] N. Gehrig and P. L. Dragotti, “Distributed sampling and compression of scenes with finite rate of innovation in camera sensor networks,” *Proc. Data Compression Conference*, March 2006.

[11] P. Marziliano, M. Vetterli, and T. Blu, “Sampling and exact reconstruction of bandlimited signals with additive shot noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2230–2233, May 2006.

[12] “http://www.ntu.edu.sg/home/epina/Research,” .

[13] Y. Hao, P. Marziliano, M. Vetterli, and T. Blu, “Compression of ECG as signal with finite rate of innovation,” *IEEE 27th Annual International Conference on Engineering, Medicine and Biology*, pp. 7564–7567, Sept. 2005.

[14] “http://lcavwww.epfl.ch/software/,” .