QuasiRandom Sequences for Signal Sampling and Recovery

Miroslaw Pawlak, Ewaryst Rafajlowicz

Quasi-Random Sequences for Signal Sampling and Recovery

Mirosław Pawlak [(1)] and Ewaryst Rafajłowicz [(2)]

(1) Dept. of Electrical & Computer Eng., University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2
(2) Institute of Computer Eng., Control and Robotics, Wrocław University of Technology, Wroclaw, Poland
pawlak@ee.umanitoba.ca, ewaryst.rafajlowicz@pwr.wroc.pl

Abstract:

The problem of reconstruction of band-limited signals from sampled and noisy observations is studied. It is proposed to sample a signal at quasi-random points, that form a deterministic sequence with properties resembling a random variable being uniformly distributed. Such quasi-random points can be easily and efficiently generated yielding signal reconstruction algorithms with the improved accuracy. In fact, in this paper we propose a reconstruction method based on the modified orthogonal sampling formula where the sampling rate and the reconstruction rate are treated separately. This distinction is necessary to ensure consistency of the reconstruction algorithm in the presence of noise. Asymptotical properties of the algorithm are evaluated including its convergence to the true signal and the corresponding rate. It is shown that the rate of convergence is better than that for reconstructions algorithms that utilize the traditional uniform sampling. Similar results are also obtained for the case of multivariate signals.

1. Introduction

Signal sampling is an inherent part of the modern signal processing theory and as such it has attracted a great deal of research activities lately [9], [10]. In particular, the problem of signal sampling and recovery from imperfect data has been addressed in a number of recent works [1], [2], [7]. In this case, one assumes that the signal samples $\{f(k\tau)\}$ are observed with noise, i.e., we have

$$y_k = f(k\tau) + z_k,$$

where z_k is uncorrelated noise process with $E\, z_k = 0$, $\mathrm{var}(z_k) = \sigma^2 < \infty$. Throughout the paper we assume that $f(t)$ has a bounded spectrum and that $f(t)$ is a finite energy type signal. Any signal with such a property is referred to as band-limited and will denote this class of signals as $BL(\Omega)$, where Ω is the bandwidth of $f(t)$.
The celebrated Whittaker-Shannon theorem says that any band-limited signal $f(t)$ can be perfectly recovered from its discrete values $\{f(k\tau)\}$ provided that $\tau \leq \pi/\Omega$. Application of the resulting interpolation formula to noisy data would lead to the following reconstruction scheme based on $2n+1$ random samples

$$f_n(t) = \sum_{|k|\leq n} y_k \,\mathrm{sinc}\left(\pi\tau^{-1}\left(t - k\tau\right)\right), \qquad (1)$$

where $\mathrm{sinc}(t) = \sin(t)/t$, and $\tau \leq \pi/\Omega$. The fundamental question, which arises is whether $f_n(t)$ can be a consistent estimate of $f(t)$ for any $f \in BL(\Omega)$. Hence, whether $\varrho(f_n, f) \to 0$ as $n \to \infty$, in a certain probabilistic sense, for some distance measure ϱ. Since $f(t)$ is assumed to be square integrable, then the natural measure between $f_n(t)$ and $f(t)$ is the mean integrated square error

$$MISE(f_n) = \mathbf{E}\int_{-\infty}^{\infty} (f_n(t) - f(t))^2\, dt. \qquad (2)$$

It can be easily shown, see [6], that $MISE(f_n) \to \infty$ as $n \to \infty$ for any fixed $\tau \leq \pi/\Omega$. This unpleasant property of the estimate $f_n(t)$ is caused by the presence of the noise process in the observed data and the fact that $f_n(k\tau) = y_k$, i.e., $f_n(t)$ interpolates the noisy observations. It is clear that one should avoid interpolation schemes in the presence of noise since they would retain random errors. The aim of this paper is to propose a consistent estimate of $f(t)$ being a smooth correction of the naive algorithm $f_n(t)$. This task is carried out by sampling a signal at irregularly spaced quasi-random points and by carefully selecting the number of terms in the sampling series. The conditions for consistency of our estimate are established and the corresponding rate of convergence is evaluated.
The statistical aspects of signal sampling and recovery have been examined first in [5], and next in [6], [7], [2], [1]. In [2], [1] the sampling rate τ has been assume to be a fixed constant. This assumption, however, cannot lead to consistent estimates of the true signal of the band-limited type. On the other hand, in [5], [6], [7] $\tau = \tau_n$

such that $\tau_n \to 0$ as $n \to \infty$ with a controlled rate. Such a choice of τ allows us to design a signal recovery algorithm for which the reconstruction error $MISE$ tends to zero with a certain rate. In this paper, we propose a nonlinear sampling scheme based on the theory of quasi-random sequences, i.e., we observe the following noisy samples

$$y_k = f(\tau_k) + z_k,$$

where $\{\tau_k\}$ is a sequence of quasi-random points. We show that a proper choice of $\{\tau_k\}$ leads to the reconstruction algorithm with the improved convergence rate.

2. Reconstruction Algorithms with Quasi-Random Points

The notion of quasi-random sequences has been originally established in the theory of numerical integration [4]. A sequence of real numbers $\{x_j\}$ is said to be a quasi-random sequence in $[0,\,1]$ if for every continuous function $b(x)$ on $[0,\,1]$ we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} b(x_j) = \int_0^1 b(x)\,dx. \qquad (3)$$

Quasi-random sequences are also called equidistributed sequences, since (3) means that the sequence $\{x_j\}$ behaves like uniformly distributed random variables. Nevertheless, an important property of quasi-random sequences is that they are more uniform than random uniform sequences which tend to clump. A consequence of this fact is that the accuracy of approximating integrals based on quasi-random sequences is superior to the accuracy obtained by random sequences. In fact, the celebrated Koksma-Hlawka inequality [4] says that for any function of bounded variation on $[0,1]$ we have

$$\left| n^{-1}\sum_{j=1}^{n} f(x_j) - \int_0^1 f(t)dt \right| \le \mathcal{V}(f) D_n^*,$$

where $\mathcal{V}(f)$ is the total variation of f on $[0,1]$, and D_n^* denotes the so-called discrepancy of the quasi-random sequence $\{x_j\}$. The discrepancy measures the strength of the sequence to approximate the uniform distribution on $[0,1]$. There are quasi-random sequences with discrepancy of order $O(log(n)/n)$ [4]. This should be contrasted with a random sequence of uniformly distributed points on $[0,1]$ that possesses the discrepancy of order $O(1/\sqrt{n})$. This basic observation plays a key role in our developments concerning the signal recovery problem from quasi-random points. Numerous quasi-random sequences have been constructed that have the aforementioned property of approximating the uniform distribution. The simplest, and sufficient for our purposes, way of generating a quasi-random sequence is the following

$$x_j = \mathrm{frac}(j\,\vartheta), \qquad (4)$$

where ϑ is an irrational number and frac(.) denotes the fractional part of a number in the parenthesis. A good choice of ϑ is $(\sqrt{5}-1)/2$, see [8] for an extensive discussion on the choice of ϑ.

Since band-limited signals are defined on the whole real line we need a rescaled version of quasi-random sequences. Thus, let us define the following sampling points on the interval $[-T,\,T]$

$$\tau_j = T\,\mathrm{sgn}(j)\,\mathrm{frac}(|j|\,\vartheta), \quad j = 0, \pm 1, \pm 2, \ldots, n, \quad (5)$$

where sgn(.) is the sign of a number. The observation horizon T must increase with n such that $T(n) \to \infty$ as $n \to \infty$. In order, however, to establish the consistency result of our reconstruction algorithm we must control the growth of $T(n)$. The approximation property of quasi-random sequences applied to the sequence defined in (5) reads now as

$$\frac{2T}{2n+1}\sum_{|j|\le n} f(\tau_j) \approx \int_{-T}^{T} f(t)\,dt. \qquad (6)$$

It has been known since the work of Hardy [3] that the cardinal expansion can be viewed as the orthogonal expansion in $BL(\Omega)$. Using this fact and then the reasoning as in [5] we can define the following estimate of $f(t)$

$$\tilde{f}_n(t) = \sum_{|k|\le N} \tilde{c}_k s_k(t)\ , \qquad (7)$$

$$\tilde{c}_k = \frac{2T}{(2n+1)\,h}\sum_{|j|\le n} y_j\, s_k(\tau_j)\ , \qquad (8)$$

where $\{\tau_j\}$ is the quasi-random sequence defined in (5). Here $\{s_k(t) = \mathrm{sinc}(\pi h^{-1}(t-kh)), k = 0, \pm 1, \ldots\}$ forms the orthogonal and complete system in $BL(\Omega)$ provided that $h \le \pi/\Omega$. The corresponding Fourier coefficient is $c_k = h^{-1}\int_{-\infty}^{\infty} f(t)s_k(t)dt$. It is also clear that for $f \in BL(\Omega)$ we have $c_k = f(kh)$. The parameter h is called the reconstruction rate. In (7) the parameter N defines the number of terms in the expansion which are taken into account and $2n+1$ is the sample size. The truncation parameter plays important role in our asymptotic analysis, i.e., N depends on n such that $N(n) \to \infty$ with the controlled rate. It is also worth noting that the sampling rate is nonuniform (defined by the discrepancy of the quasi-random sequence in (5)) and different than the reconstruction rate h. We assume that h is constant and not greater than π/Ω.

Throughout the paper we use the worst localized base system utilizing the $sinc$ function. The methodology presented in this paper can be extended to the windowed version of the estimate $\tilde{f}_n(t)$ of the form

$$\hat{f}_n(t) = \sum_{|k|\le n} w_k \tilde{c}_k s_k(t),$$

where $\{w_k, |k| \le n\}$ is a sequence of numbers such that $0 \le w_k \le 1$. The proper choice of this window sequence yields an estimate with better time-localized properties and consequently better convergence rates. The case when $w_k = 1$ for $|k| \le N$ and $w_k = 0$ otherwise corresponds to the estimate $\tilde{f}_n(t)$.

3. The MISE Consistency and Rate

In this section we summarize the results concerning the convergence of MISE($\tilde{f}_n$) to zero as $n \to \infty$ for any signal $f \in BL(\Omega)$. Also the rate of convergence is established.

Due to Pareseval's formula we can decompose the $MISE(\tilde{f}_n)$ as follows:

$$MISE(\tilde{f}_n) = h \sum_{|k|\le N} var(\tilde{c}_k) + h \sum_{|k|\le N} (E\tilde{c}_k - c_k)^2$$

$$+h \sum_{|k|\ge N} c_k^2.$$

The first term of the decomposition controls the stochastic part of the estimate, whereas the the remaining term describe the systematic error (bias) of the estimate. A careful examination of these terms lead to the following result on the consistency of our estimate.

Theorem 1 *Let $f \in BL(\Omega)$ and let the reconstruction rate h be constant such that $h \le \pi/\Omega$. Suppose that $N(n) < T(n)/h$. Assume $T(n) \to \infty$, $N(n) \to \infty$ such that $T(n)$ does not grow faster than $\sqrt{n}/log(n)$. Let, moreover,*

$$\frac{N(n)T(n)}{n} \to 0.$$

Then

$$MISE(\tilde{f}_n) \to 0$$

as $n \to \infty$.

The conditions required on the parameters $T(n)$ and $N(n)$ in Theorem 1 impose some general restrictions on their growth. In order further see how to choose $T(n)$ and $N(n)$ let us assume the following condition on the decay of band-limited signals.

(F) There exists $r \ge 0$ and a constant $C_f > 0$ such that for $|t|$ sufficiently large we have $|f(t)| \le C_f/|t|^{r+1}$.

This assumption can be also expressed in the frequency domain by requiring that the Fourier transform of $f(t)$ has r derivatives on $[-\Omega, \Omega]$. A further analysis of the reconstruction error leads to the following bound

$$\begin{aligned} MISE(\tilde{f}_n) \quad \le \quad & (2N+1)\Big(C_1 T^{-(2r+1)} \\ & + \frac{C_2 T^3 \log^2(n)}{n^2} + \frac{C_3 T}{n}\Big) \quad (9) \\ & + C_4 N^{-(2r+1)}, \end{aligned}$$

for some constants C_1, C_2, C_3, C_4. By optimizing the above bound we can obtain the following asymptotically optimal choice of $T(n)$ and $N(n)$.

$$T^*(n) = an^{\frac{1}{2r+3}} \quad N^*(n) = bn^{\frac{1}{2r+3}},$$

subject to the condition $a > bh$. Plugging these values of $T(n)$ and $N(n)$ back into the bound for $MISE(\tilde{f}_n)$ we obtain the following rate

$$MISE(\tilde{f}_n) = O(n^{-\frac{2r+1}{2r+3}}).$$

It is worth noting that under Assumption (F) the best possible rate obtained for the reconstruction algorithms discussed in [6] and [7] is of order $O(n^{-\frac{r}{r+1}})$. This is clearly a slower rate than the one obtained in this paper.

4. Concluding Remarks

In this paper we have proposed an algorithm for recovering a band-limited signal observed under noise. Assuming that the signal is a square integrable function the sufficient conditions for the convergence of the mean integrated square error have been established. The distinguishing feature of the proposed approach is its utilization of nonuniform samples taken at quasi-random points. When quasi-random sequences are applied to the problem of numerical evaluation of integrals they reveal the approximation rate $O(log(n)/n)$ for a class of bounded variation functions. This rate is superior to the rate $O(1/\sqrt{n})$ that characterizes usual numerical algorithms and classical Monte Carlo methods. This advantage of quasi-random sequences seems to be carried out to the problem of signal sampling and recovery. In our consistency results we assume that the reconstruction rate h is constant and could be chosen as large as π/Ω. One could also consider the case when $h = h(n)$ and $h(n) \to 0$ as $n \to \infty$. The estimates with variable h would be needed for the problem of recovering not necessarily band-limited signals. Finally, let us mention that the results of this paper can be extended to the d-dimensional case, where the orthogonal system can be obtained in the form of the product of sinc functions, i.e., $\mathbf{s_k}(\mathbf{t}) = \prod_{i=1}^{d} s_{k_i}(t_i)$, where $\mathbf{k} = (k_1, k_2, \ldots, k_d)$,

$\mathbf{t} = (t_1, t_2, \ldots, t_d)$. We should mention that multidimensional quasi-random sequences can be generated in a relatively straightforward way. Moreover, they exhibit the favorite discrepancy of order $O(n^{-1}(log(n))^d)$ for any d. This fact may have important consequences for sampling problems of two-dimensional objects like images.

5. Acknowledgements

The work of E. Rafajłowicz was supported by the Research and Development Grant from the Ministry of Science and Higher Education of Poland.

References:

[1] A. Aldroubi, C. Leonetti, and Q. Sun. Error analysis of frame reconstruction from noisy samples. *IEEE Trans. Signal Processing*, 56:2311–2315, 2008.

[2] Y.C. Eldar and M. Unser. Non-ideal sampling and interpolation from noisy observations in shift-invariant spaces. *IEEE Trans. Signal Processing*, 54:2636–2651, 2006.

[3] G.H. Hardy. Notes on special systems of orthogonal functions (iv): the orthogonal functions of Whittaker's cardinal series. *Proc. Camb. Phil. Soc.*, 37:331–348, 1941.

[4] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. Wiley, New York, 1974.

[5] M. Pawlak and E. Rafajłowicz. On restoration of band-limited signals. *IEEE Trans. Information Theory*, 40:1490–1503, 1994.

[6] M. Pawlak, E. Rafajłowicz, and A. Krzyżak. Post-filtering versus prefiltering for signal recovery from noisy samples. *IEEE Trans. Information Theory*, 49:3195–3212, 2003.

[7] M. Pawlak and U. Stadtmüller. Signal sampling and recovery under dependent noise. *IEEE Trans. Information Theory*, 53:2526–2541, 2007.

[8] E. Rafajłowicz and R. Schwabe. Equidistributed designes in nonparametric regression. *Statistica Sinica*, 13:129–142, 2003.

[9] M. Unser. Sampling – 50 years after Shannon. *Proceedings of the IEEE*, 88:569–587, 2000.

[10] P.P. Vaidyanathan. Generalizations of the sampling theorems: seven decades after Nyquist. *IEEE Trans. on Ciruits and Systems – I : Fundamental Theory and Applications*, 48:1094–1109, 2001.