THRESHOLDING-BASED SAMPLING AND SIGNAL RECONSTRUCTION FROM MULTIPLE NOISY OBSERVATIONS

L. V. Nguyen and M. Pawlak

Department of Electrical and Computer Engineering
University of Manitoba,Winnipeg, Manitoba R3T 5V6, Canada
Emails: {nvluu, pawlak}@ee.umanitoba.ca

ABSTRACT

The problem of sampling and recovering bandlimited signals in the presence of noise is studied. A new signal acquisition technique, which collects multiple sets of thresholded signal's samples acquired at the Nyquist sampling rate, is proposed. The exact formula for mean integrated squared error of the proposed scheme is established. The obtained results show that the proposed technique gives better reconstruction accuracy than the common method using oversampling and post-filtering. Moreover, in most cases, the proposed scheme reduces the amount of data required for representation and reconstruction of signal. As a result, it provides the reduced data storage.

1. INTRODUCTION AND PRELIMINARIES

Let signal $x(t)$ belong to a class of $L_2(R)$ signals where its Fourier transform $X(\omega)$ vanishes outside the finite interval $(-\Omega,\Omega)$, i.e. $X(\omega)=0$ for $|\omega|>\Omega$. The finite number Ω is called the signal's bandwidth and the class of signals with this property is often referred to as the class of band-limited signals, which we will denote $BL(\Omega)$ in the subsequent discussions.

The Whittaker-Shannon (W-S) sampling theorem says that every $x(t)\in BL(\Omega)$ can be reconstructed exactly from its discrete samples $x(k\tau)$, $\quad k=0,\pm1,\pm2,\ldots\pm\infty$, by:

$$x(t)=\sum_{k=-\infty}^{\infty} x(k\tau)\cdot \mathrm{sinc}\left(\frac{t}{\tau}-k\right), \tag{1.1}$$

provided that $\tau\le\tau_q=\pi/\Omega$, where $\mathrm{sinc}(t)=\sin(\pi t)/\pi t$.

We refer to [1, 2, 3] for an extensive overview of the theory and applications of (1.1) and its extensions.

In this paper, we consider the problem of reconstructing bandlimited signals given the following observation model:

$$y_k=x(k\tau)+\varepsilon_k=x_k+\varepsilon_k, \qquad |k|\le n$$

where $\{\varepsilon_k\}$ is an additive noise process.

Our motivation comes from the fact that in many practical applications, noise is often contained in the observed samples. For instance, measurement and/or communication channel noises could be potential noise sources.

The classical W-S interpolation scheme would replace $\{x(k\tau)\}$ in (1.1) by $\{y_k\}$ yielding the following estimate:

$$\widetilde{x}(t)=\sum_{|k|\le n} y_k\cdot \mathrm{sinc}\left(\frac{t}{\tau}-k\right). \tag{1.2}$$

This work was supported in part by TR*Labs* and NSERC

Figure 1: The oversampling/post-filtering method

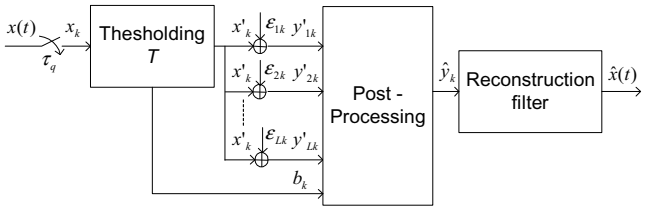

Figure 2: The proposed sampling/reconstruction scheme

It has been pointed out in [4] that the above estimate does not have noise-diminishing property since it interpolates noise, i.e., $\widetilde{x}(k\tau)=y_k$. The common method to overcome this problem is using oversampling and post-filtering. Consider the following estimate [5, 6]:

$$\breve{x}(t)=\sum_{|k|\le n} y_k\cdot \mathrm{sinc}\left(\frac{t}{\tau}-k\right)*h(t), \tag{1.3}$$

where, $\tau\le\tau_q=\pi/\Omega$ and $h(t)=\tau_q^{-1}\mathrm{sinc}(t/\tau_q)$ is the impulse response of the ideal low-pass filter with cutoff frequency Ω. The block diagram of this method is illustrated in Figure 1.

Since the estimate in (1.3) relies on the use of high sampling rate for noise reduction, its realization requires costly high-speed electronic circuits. There also exists the potential problem of increasing the correlation between successive samples as the sampling rate increases, which leads to unstable reconstruction.

In this paper, instead of oversampling, we propose a new technique that acquires multiple sets of the thresholded signal's samples at the Nyquist sampling rate. The proposed technique is an extension of the thresholding-based sampling scheme introduced in [7].

The paper is organized as follows. In Section 2, we introduce the proposed sampling and reconstruction scheme. In Section 3, we present the error analysis of the proposed scheme. Section 4 provides some simulation results. Section 5 concludes the paper.

2. THE PROPOSED SAMPLING AND RECONSTRUCTION SCHEME

The block diagram of the proposed scheme is illustrated in Figure 2. It can be divided into three parts (data acquisition, observation model and reconstruction) as follows.

2.1 Data Acquisition

The data acquisition process consists of two steps:

Step 1: Uniformly sampling the signal $x(t)$ at the Nyquist sampling rate to obtain the sequence of signal's samples $\{x_k\}$. Note that the length of $\{x_k\}$ is $2n_q+1$, where n_q relates to n in (1.3) by $(2n_q+1)\tau_q = (2n+1)\tau$. Here, we assume that the duration of observation is fixed.

Step 2: Passing $\{x_k\}$ through a threshold device with the threshold level T. Only those samples of $\{x_k\}$ for which $|x_k| > T$ are kept. The other samples are destroyed. Therefore, in Figure 2, the sequence $\{x'_k\}$ relates to the sequence $\{x_k\}$ by:

$$\{x'_k\} = \{x_k : |x_k| > T\}, \tag{2.1}$$

where, T is a threshold level. Its value is to be determined later.

Instead of using the single set $\{x'_k\}$ as in [7], in this paper, L copies of the set $\{x'_k\}$ are collected.

For signal reconstruction purpose, we need to know the location at which a particular sample is kept or destroyed. This information is made available at the receiver by using the sequence of binary digits $\{b_k\}$, which is formed as follows.

$$b_k = \begin{cases} 1 & \text{if} \quad |x_k| > T \\ 0 & \text{if} \quad |x_k| \le T. \end{cases} \tag{2.2}$$

Note that the length of sequence $\{b_k\}$ is $2n_q+1$, whereas the length of sequence $\{x'_k\}$ is $\eta \le 2n_q+1$. Since the length of sequence $\{x'_k\}$ is always shorter than that of sequence $\{x_k\}$, the thresholding unit acts as a compressor. At the receiver side, the expander (embedded within post-processing unit) inserts zeros into the noisy sequence $\{\overline{y'}_k\}$ defined in (2.4), at the locations for which $b_k = 0$. Notice that the compressor (also called decimator) and the expander are the well-known building blocks in digital signal processing theory.

2.2 Observation Model

Since $\{b_k\}$ is the digital sequence, it is resistant to noise or can be protected by forward error codes [8]. Therefore, we assume that $\{b_k\}$ is available for signal reconstruction. However, $\{x'_k\}$ are corrupted by noise, i.e.,

$$y'_{lk} = x'_k + \varepsilon_{lk}, \qquad \begin{array}{l} k = 1,...,\eta \\ l = 1,2,\cdots,L \end{array} \tag{2.3}$$

where $\{\varepsilon_{lk}\}$ is an additive noise process and η is the length of $\{x'_k\}$.

2.3 Reconstruction

The signal reconstruction process consists of 3 steps as follows.

Step 1: The post-processing unit calculates the sample means

$$\overline{y'}_k = \frac{1}{L}\sum_{l=1}^{L} y'_{lk}. \tag{2.4}$$

Step 2: From two sequence $\{b_k\}$ and $\{\overline{y'}_k\}$, the sequence $\{\widehat{y}_k\}$ is formed according to the procedure described below.

- Setting $\{\widehat{y}_k\}$ to the all-zero sequence. Its length is $2n_q+1$.
- Denoting $k_1, k_2, \ldots, k_\eta$, $(k_1 < k_2 <, \ldots, < k_\eta)$ are the locations at which $b_k = 1$. Setting $\widehat{y}_{k_i} = \overline{y'}_i$, $i = 1,2,\ldots,\eta$.

For example, if $\{b_k\} = 1\ 0\ 1\ 1\ 0\ 1\ 1\ 0\ 0\ 1\ 0$ and $\{\overline{y'}_k\} = \overline{y'}_6\overline{y'}_5\overline{y'}_4\overline{y'}_3\overline{y'}_2\overline{y'}_1$, then $\{\widehat{y}_k\} = \overline{y'}_6\ 0\ \overline{y'}_5\ \overline{y'}_4\ 0\ \overline{y'}_3\ \overline{y'}_2\ 0\ 0\ \overline{y'}_1\ 0$. Mathematically, we have:

$$\widehat{y}_k = \begin{cases} \overline{y'}_k & \text{for} \quad k : |x_k| > T \\ 0 & \text{for} \quad k : |x_k| \le T. \end{cases} \tag{2.5}$$

Step 3: The original signal is reconstructed by:

$$\widehat{x}(t) = \sum_{|k|\le n_q} \widehat{y}_k \cdot \text{sinc}\left(\frac{t}{\tau_q} - k\right). \tag{2.6}$$

2.4 Data Storage Reduction Ratio

Let assume that a m-bit uniform quantizer is used to obtain binary representation of the sampled values. The oversampling/post-filtering reconstruction scheme in (1.3) requires $B_{OS} = (2n+1)\cdot m$ bits for representation of the original signal. On other hand, our proposed sampling and reconstruction scheme requires $B_T = L\eta m + (2n_q+1)$ bits. If we denote $D_R = B_T/B_{OS}$ as the data storage reduction ratio (DSRR), then it can be shown that:

$$D_R = \frac{L\eta}{2n+1} + \frac{\tau}{\tau_q}\cdot\frac{1}{m}. \tag{2.7}$$

For comparison with the oversampling/post-filtering method in (1.3), consider the following assumption.

Assumption 2.1. *Assume that the number of copies of the set $\{x'_k\}$ in our proposed technique equals to the oversampling factor (OSF) in the oversampling/post-filtering method, i.e., $L = (2n+1)/(2n_q+1) = \tau_q/\tau$.*

In this case, we have:

$$D_R = \frac{\eta}{2n_q+1} + \frac{1}{Lm}. \tag{2.8}$$

Several conclusions can be drawn from (2.8):

- D_R decreases as m increases. In other word, the finer the quantizer, the higher degree of data storage reduction can be achieved.
- D_R decreases as η decreases, or equivalently, T increases. However, T can not be set arbitrary large since this will effect the reconstruction accuracy.
- For the fixed T, D_R depends on the shape of signal. If the total amount of time that the signal's amplitude falls below threshold level T is large, then D_R is small. In this case, we are benefited from using the proposed technique.
- D_R decreases as L increases, or equivalently, τ_q/τ increases. In other word, if very accurate signal reconstruction is expected, then using the proposed technique requires smaller amount of bits for representation and reconstruction of the original signal than using the oversampling/post-filtering method.

3. ERROR ANALYSIS

Let us consider mean integrated squared error (MISE) as a measure of performance of $\widehat{x}(t)$ and other estimates examined in the paper.

$$MISE(\widehat{x}) \triangleq E\int_{-\infty}^{\infty}(\widehat{x}(t)-x(t))^2dt.$$

The following theorem gives the exact formula for the MISE of the proposed sampling and reconstruction scheme.

Theorem 3.1. *Consider the observation model in (2.3), where ε_{lk} are independent random variables with zero mean and variance σ^2. The MISE of the estimate $\widehat{x}(t)$ in (2.6) is given by:*

$$MISE(\widehat{x}) = \tau_q\sigma_1^2\eta + \tau_q\sum_{k\in I_d}x_k^2 + \tau_q\sum_{|k|>n_q}x_k^2, \tag{3.1}$$

where, $\sigma_1^2 = \sigma^2/L$; $I_p = \{k : |x_k| > T\}$ and $I_d = \{k : |x_k| \le T\}$ are two index sets; η is the length of I_p.

Proof. Due to Paseval's formula, $MISE(\widehat{x})$ can be expressed as follows.

$$MISE(\widehat{x}) = \tau_q\sum_{|k|\le n_q}E(\widehat{y}_k - x_k)^2 + \tau_q\sum_{|k|>n_q}x_k^2.$$

Taking (2.5) into account, we have:

$$\begin{aligned} MISE(\widehat{x}) &= \tau_q\sum_{k\in I_p}E(\overline{y'}_k - x'_k)^2 + \tau_q\sum_{k\in I_d}x_k^2 + \tau_q\sum_{|k|>n_q}x_k^2 \\ &= \tau_q\sum_{k\in I_p}\mathrm{var}(\overline{y'}_k) + \tau_q\sum_{k\in I_d}x_k^2 + \tau_q\sum_{|k|>n_q}x_k^2 \\ &= \tau_q\sigma_1^2\eta + \tau_q\sum_{k\in I_d}x_k^2 + \tau_q\sum_{|k|>n_q}x_k^2. \end{aligned}$$

This completes the proof. □

Theorem 3.2. *For a given signal $x(t)$ and noise variance σ^2, there always exists the optimum threshold level T_{opt} so that $MISE(\widehat{x})$ is minimum.*

Proof. $MISE(\widehat{x})$ can be decomposed into the integrated variance (IVAR) and integrated bias (IBIAS) components as follows.

$$MISE(\widehat{x}) = IVAR(\widehat{x}) + IBIAS(\widehat{x})$$

$$IBIAS(\widehat{x}) \triangleq \int_{-\infty}^{\infty}(E\widehat{x}(t) - x(t))^2dt$$

$$IVAR(\widehat{x}) \triangleq E\int_{-\infty}^{\infty}(\widehat{x}(t) - E\widehat{x}(t))^2dt.$$

The term $IBIAS(\widehat{x})$ describes the truncation error and the error due to threshoding operation. On other hand the term $IVAR(\widehat{x})$ reflects the existence of random errors in the observed samples.

Parseval's formula yields the following decompositions of the errors [4]:

$$IBIAS(\widehat{x}) = \tau_q\sum_{|k|\le n_q}(E\widehat{y}_k - x_k)^2 + \tau_q\sum_{|k|>n_q}x_k^2$$

$$IVAR(\widehat{x}) = \tau_q\sum_{|k|\le n_q}\mathrm{var}(\widehat{y}_k).$$

Taking (2.5) into account, we have:

$$\begin{aligned} IBIAS(\widehat{x}) &= \tau_q\sum_{k\in I_p}(E\overline{y'}_k - x'_k)^2 + \tau_q\sum_{k\in I_d}x_k^2 + \tau_q\sum_{|k|>n_q}x_k^2 \\ &= \tau_q\sum_{k\in I_d}x_k^2 + \tau_q\sum_{|k|>n_q}x_k^2 \end{aligned} \tag{3.2}$$

$$IVAR(\widehat{x}) = \tau_q\sum_{k\in I_p}\mathrm{var}(\overline{y'}_k) = \tau_q\sigma_1^2\eta. \tag{3.3}$$

As T increases, η decreases and $IVAR(\widehat{x})$ decreases accordingly. In contrast, since the first term in (3.2) increases as T increases, $IBIAS(\widehat{x})$ increases as T increases. Since $IVAR(\widehat{x})$ is a monotonically decreasing function of T, whereas $IBIAS(\widehat{x})$ is a monotonically increasing function of T, the sum $MISE(\widehat{x}) = IVAR(\widehat{x}) + IBIAS(\widehat{x})$ has the minimum at the value of T, for which $IVAR(\widehat{x}) = IBIAS(\widehat{x})$. This value of T is called the optimal threshold level T_{opt}. The proof is complete. □

For a given noise variance, $MISE(\widehat{x})$ can be pre-computed during data acquisition process using (3.1) and the optimal threshold level T_{opt} can be pre-determined accordingly. T_{opt} is considered as the side information and can be made available for signal reconstruction by various methods.

Corollary 3.1. *If $T \le \sigma_1$, then*

$$R = \frac{MISE(\widehat{x})|_{T\le\sigma_1}}{MISE(\widehat{x})|_{T=0}} \le 1, \quad \textit{for all} \quad \tau, n, \sigma^2 \tag{3.4}$$

Proof. If $T \le \sigma_1$, then for any $k \in I_d$ we have $x_k^2 \le T^2 \le \sigma_1^2$. Consequently, $\sum_{k\in I_d}x_k^2 \le \sigma_1^2|I_d| = \sigma_1^2[(2n_q+1)-\eta]$. Here, $|I_d|$ denotes the volume (size) of the index set I_d.

Therefore, from (3.1) we have:

$$MISE(\widehat{x})|_{T\le\sigma_1} \le \tau_q\sigma_1^2(2n_q+1) + \tau_q\sum_{|k|>n_q}x_k^2 = MISE(\widehat{x})|_{T=0}.$$

This completes the proof. □

Remark 3.1. Corollary 3.1 does not imply that the optimal threshold level must satisfy $T_{opt} \le \sigma_1$. However, it did suggest a sub-optimal criteria for setting threshold level, i.e., $T = \sigma_1$. Here, we assume that noise variance is known. Using this criteria will eliminate the need to deliver the side information (i.e. the optimal threshold level) to the receiver for signal reconstruction.

With $T = \sigma_1$, one can observe that R decreases if $|I_d|$ increases, or equivalently, σ^2 increases. This indicates the advantage of using the proposed technique when SNR is low.

4. SIMULATION RESULTS

The following signal is used as the testing signal in simulations:

$$x(t) = \frac{1}{\sqrt{\pi\Omega}}\cdot\frac{\sin\Omega t}{t} \tag{4.1}$$

The above signal is the unit energy $BL(\Omega)$ signals, where $\Omega = 2\pi f_{max}$ is the radian frequency. In our simulation, we set $f_{max} = 3900$ Hz. This signal has the slowest rate of decay as $t \to \infty$ among all bandlimited signals having the same bandwidth. Since the proposed scheme works especially well for

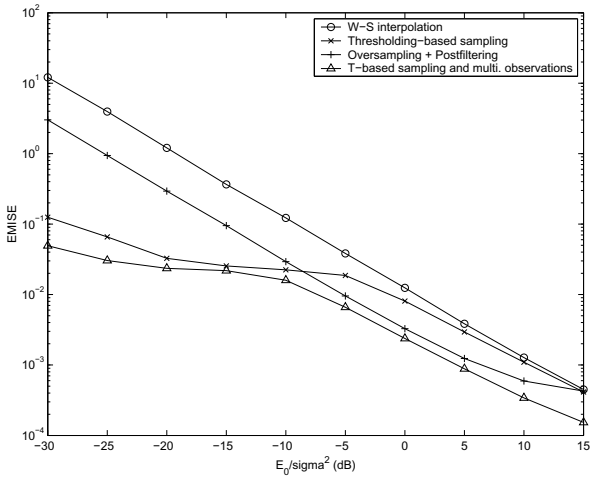

Figure 3: $EMISE(\widehat{x})$ versus E_0/σ^2 (L=4)

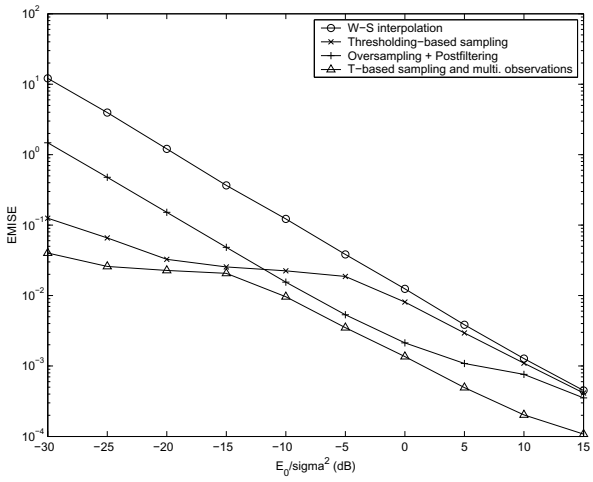

Figure 4: $EMISE(\widehat{x})$ versus E_0/σ^2 (L=8)

fast-decay signals, our use of this signal as the testing signal is justified.

The experiment is repeated $M = 100$ times for various realization of random errors $\{\varepsilon_{lk}\}$. The empirical counterpart of MISE, further called EMISE, is calculated according to the following formula:

$$EMISE(\widehat{x}) = \frac{\bar{\tau}}{M} \sum_{j=1}^{M} \sum_{k=-n}^{n} [\widehat{x}(k\bar{\tau}) - x(k\bar{\tau})]^2,$$

where $\bar{\tau} \ll \tau_q$ is the simulation sampling period.

Figure 3 and Figure 4 plot the EMISE of the proposed sampling and reconstruction scheme when $L = 4$ and $L = 8$, respectively. In this simulation, $2n_q = 100$ and $\tau_q = 1.25 \cdot 10^{-4}$ seconds (sampling frequency is just slightly larger than the Nyquist rate). Threshold level is set to $T = \sigma_1 = \sigma/\sqrt{L}$. For comparison, the EMISE of the thresholding-based sampling scheme proposed in [7] and the EMISE of the oversampling/post-filtering reconstruction scheme in (1.3) are also plotted. As being indicated in the plots, the proposed scheme gives better reconstruction accuracy than the others for every values of SNR. However, the significant improvement in the reconstruction accuracy is observed at the region of low SNR.

DSRR: For the testing signal considered; $L = 4$; $T = \sigma_1$; $2n_q + 1 = 101$ and $E_0/\sigma^2 = -5$ dB, we have $\eta = 55$. Therefore, if 3-bit uniform quantizer is used then according to (2.8) the DSRR is $D_R = 55/101 + (1/4) \cdot (1/3) \approx 1/5.9$. Table 1 shows the computed D_R for various values of E_0/σ^2.

Table 1: The Data Storage Reduction Ratio

E_0/σ^2 (dB)	-10	-5	0	5
D_R (Signal 1)	1/2.2	1/1.6	1/1.4	1/1.1

5. CONCLUSIONS

A new method for acquisition and reconstruction of bandlimited signals in the presence of noise was proposed. The proposed scheme acquires multiple sets of the thresholded signal's samples at the Nyquist sampling rate. We showed that threshold level can be optimally chosen to yield the minimum reconstruction error. We also suggested a suboptimal criteria for setting threshold level, which eliminates the need for delivering the side information (i.e. the optimal threshold level) to the receiver. The proposed technique not only offers better reconstruction accuracy but requires smaller amount of bits for representation and reconstruction of the original signal than the oversampling/filtering method. Since the proposed technique does not use oversampling, it can be used to acquire ultra wideband signals. It also has the application in multi-sensor information systems.

REFERENCES

[1] A.J. Jerri, "The Shannon sampling theorem - its various extensions and applications: A tutorial review," *Proceedings of the IEEE*, vol. 65, pp. 1565–1596, Nov. 1977.

[2] R.J. Marks II, *Introduction to Shannon Sampling and Interpolation Theory*, Springer-Verlag, New York, NY, 1991.

[3] M. Unser, "Sampling-50 years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[4] M. Pawlak and E. Rafajlowicz, "On restoring bandlimited signals," *IEEE Transactions on Information Theory*, vol. 40, pp. 1490–1503, Sep. 1994.

[5] M. Pawlak, E. Rafajlowicz, and A. Krzyzak, "Postfiltering versus prefiltering for signal recovering from noisy samples," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3195–3212, Dec. 2003.

[6] M. Pawlak and U. Stadmuller, "Statistical aspects of sampling for noisy and grouped data," *In J. Benedetto and P. Ferreira, editors, Advances in Shannon Sampling Theory: Mathematics and Applications*, Boston, TX, pp. 317-342, 2001.

[7] Vu-Luu Nguyen and M. Pawlak, "Thresholding-based non-uniform sampling and signal reconstruction from noisy data," *International Conferences on Sampling Theory and Its Applications*, May 2003.

[8] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, 3rd edition, 1995.