RECONSTRUCTION OF BANDLIMITED SIGNALS FROM NOISY DATA BY THRESHOLDING

Vu-Luu 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 recovering bandlimited signals from discrete and noisy data is studied. A non-linear signal processing algorithm based on thresholding noisy data prior to reconstruction is proposed for improving the accuracy and robustness of reconstruction. The upper bound and the exact formula for mean integrated squared error of the proposed reconstruction scheme is established. The performance of the proposed reconstruction scheme is compared to that of the classical Whittaker-Shannon interpolation scheme.

1. INTRODUCTION AND PRELIMINARIES

The Whittaker-Shannon (W-S) sampling theorem plays a fundamental role in representing signals and images in the discrete domain. The result may be briefly stated as follows. 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 bandlimited signals, which we will denote $BL(\Omega)$ in the subsequent discussion.

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

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

provided that $\tau \leq \frac{\pi}{\Omega}$, where $\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$.

A number of properties and extensions of (1.1) have been given in literature. In particular, truncation, aliasing, location (jitter), amplitude errors of the W-S interpolation series (1.1) have been examined. Furthermore, generalizations to multiple dimensions, random signals, not necessarily bandlimited signals, missing data, wavelet subspaces and irregular sampling have been proposed. We refer to [1],[2],[3],[4] for an extensive overview of the theory and applications of (1.1) and its extensions.

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

Only recently, the statistical aspects of the W-S sampling theorem, i.e. the statistical analysis of (1.1) when only finite record of noisy data is available, have been throughly investigated [3],[5],[6],[7], [8]. In these studies, the following observation model was considered:

$$y_k = x(k\tau) + \varepsilon_k = x_k + \varepsilon_k, \qquad |k| \leq n \qquad (1.2)$$

where $\{\varepsilon_k\}$ is an additive noise process. We shall assume in this paper that $\{\varepsilon_k\}$ is a zero-mean white Gaussian noise process, independent of $\{x_k\}$ and having the variance σ^2.

The naive reconstruction algorithm would replaces $\{x(k\tau)\}$ in (1.1) by $\{y_k\}$ yielding the following estimate:

$$\widetilde{x}_n(t) = \sum_{|k|\leq n} y_k \mathrm{sinc}\left(\frac{t}{\tau} - k\right). \qquad (1.3)$$

In the subsequent discussions, we shall refer to this reconstruction scheme as the Whittaker-Shannon interpolation.

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

$$MISE(\widetilde{x}_n) = E\left[\int_{-\infty}^{\infty} (\widetilde{x}_n(t) - x(t))^2 dt\right]. \qquad (1.4)$$

Due to Paseval's formula, the error can be expressed as follows, see [7]:

$$\begin{aligned} MISE(\widetilde{x}_n) &= \tau \sum_{|k|\leq n} E[(y_k - x_k)^2] + \tau \sum_{|k|>n} x_k^2 \\ &= \tau\sigma^2(2n+1) + \tau \sum_{|k|>n} x_k^2. \qquad (1.5) \end{aligned}$$

The first term in the right-hand side of (1.5) represents the error due to the noise, whereas the second term represents the truncation error. For any finite τ, the first term approaches infinity, whereas the second term approaches zero as $n \to \infty$. Hence, $MISE(\widetilde{x}_n) \to \infty$ as $n \to \infty$.

Therefore, the reconstruction scheme based on (1.3) yields a diverged reconstruction.

The interpretation of the above is that the signal reconstruction based on the Whittaker-Shannon interpolation in (1.3) does not have noise-diminishing property. This is due to the fact that this reconstruction scheme interpolates noise.

Several the reconstruction schemes have been proposed in order to overcome this problem. For instance, moving average filtering and exponential weighting algorithms have been proposed for smoothing the noisy data prior to reconstruction [5],[6],[8]. However, since these reconstruction schemes require very high sampling rate, their applicability is somehow limited.

In this paper, we propose a reconstruction scheme for reconstruction of bandlimited signals from noisy data when oversampling is not the option, i.e. when $\tau = \tau_q = \pi/\Omega$. The proposed scheme does not require oversampling but effectively reduce the contribution of noise in the reconstructed signal, especially when signal-to-noise ratio (SNR) is low. It consists of two steps as follows:

Step 1: Thresholding noisy samples to obtain:

$$\widehat{y}_k = \begin{cases} y_k & \text{if } |y_k| > T \\ 0 & \text{if } |y_k| \le T. \end{cases} \tag{1.6}$$

Step 2: Using $\{\widehat{y}_k\}$ founded above to form the following estimate:

$$\widehat{x}_n(t) = \sum_{|k|\le n} \widehat{y}_k \mathrm{sinc}\left(\frac{t}{\tau_q} - k\right). \tag{1.7}$$

The idea of using thresholding for removal of noise is well-known in wavelet literature [9],[10]. Unlike wavelet thresholding procedures where wavelet coefficients are thresholded, in our proposed reconstruction scheme, thresholding is applied directly to the plain data, which is a sequence of signal's samples.

The paper is organized as follows. In section 2, we give the upper bound as well as the exact formula for computing MISE of the proposed reconstruction scheme. In section 3, we investigate the influence of the choice of thresholding value T on the reconstruction accuracy. Section 4 provides some simulation results. Section 5 concludes the paper.

2. ERROR ANALYSIS

The exact formula and upper bound for MISE of the proposed reconstruction scheme are presented without proof in the subsequent theorems.

Theorem 2.1. *Mean integrated squared error of the estimate $\widehat{x}_n(t)$ is found to be:*

$$\begin{aligned} MISE(\widehat{x}_n) &= \tau_q\sigma^2(2n+1) - \tau_q\sigma^2 \cdot \frac{1}{\sqrt{\pi}} \Bigg\{ \\ &\sum_{|k|\le n; |x_k|<T} \left[\gamma\left(\frac{3}{2}, \frac{(x_k-T)^2}{2\sigma^2}\right) + \gamma\left(\frac{3}{2}, \frac{(x_k+T)^2}{2\sigma^2}\right)\right] \\ + &\sum_{|k|\le n; |x_k|\ge T} \left[\gamma\left(\frac{3}{2}, \frac{(x_k-T)^2}{2\sigma^2}\right) - \gamma\left(\frac{3}{2}, \frac{(x_k+T)^2}{2\sigma^2}\right)\right] \\ - &\frac{\sqrt{\pi}}{2} \sum_{|k|\le n} \frac{x_k^2}{\sigma^2} \left[\mathrm{erf}\left(\frac{x_k+T}{\sigma\sqrt{2}}\right) - \mathrm{erf}\left(\frac{x_k-T}{\sigma\sqrt{2}}\right)\right] \Bigg\} \\ + &\tau_q \sum_{|k|>n} x_k^2. \end{aligned} \tag{2.1}$$

where, $\gamma(\alpha, x)$ and $\mathrm{erf}(x)$ are incomplete gamma function and error function, respectively [11].

Theorem 2.2. *Mean integrated squared error of $\widehat{x}_n(t)$ is upper bounded by:*

$$\begin{aligned} &MISE(\widehat{x}_n) \le \tau_q\sigma^2(2n+1) \\ &- \tau_q\sigma^2(2n+1)\left[\frac{2}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{T^2}{2\sigma^2}\right)\right. \\ &- \left.\frac{\sum_{|k|\le n} x_k^2}{\sigma^2(2n+1)} \mathrm{erf}\left(\frac{T}{\sigma\sqrt{2}}\right)\right] + \tau_q \sum_{|k|>n} x_k^2. \end{aligned} \tag{2.2}$$

3. CHOICE OF THRESHOLD T

Let

$$f(G_n, T) = \frac{2}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{T^2}{2\sigma^2}\right) - G_n \mathrm{erf}\left(\frac{T}{\sigma\sqrt{2}}\right). \tag{3.1}$$

where, $G_n = \frac{\sum_{|k|\le n} x_k^2}{\sigma^2(2n+1)}$.

It should be noted that (2.2) can be re-written as follows:

$$\begin{aligned} \Delta &= MISE(\widetilde{x}_n) - MISE(\widehat{x}_n) \\ &\ge \tau_q\sigma^2(2n+1) f(G_n, T). \end{aligned} \tag{3.2}$$

Consequently, to maximize Δ, we have to choose T such that $f(G_n, T)$ is maximum. Figure 1 plots $f(G_n, a)$ as the function of a for various G_n, where a relates to T by $T = \sigma\sqrt{a}$. Here, parameter a is used to control the reconstruction accuracy. The following conclusions can be drawn from the plot:

- For $G_n \ge 1$, one should select $T = 0$. In this case, the Whittaker-Shannon interpolation scheme yields better reconstruction accuracy compared to the proposed reconstruction scheme.

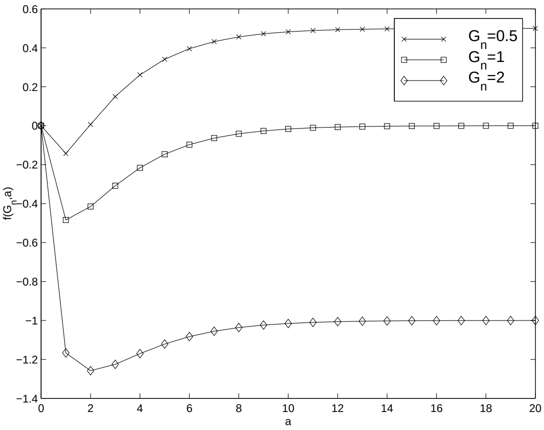

Fig. 1. $f(G_n, a)$ versus a

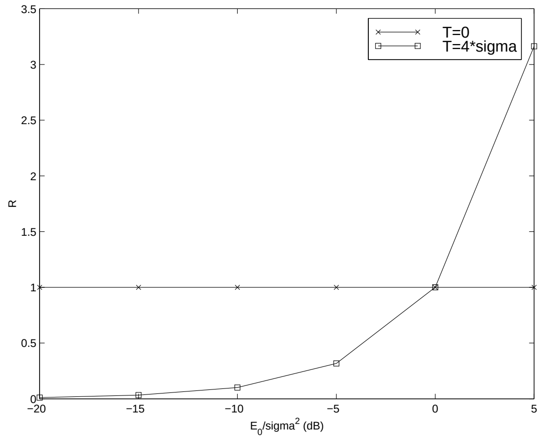

Fig. 2. Upper bound of R versus E_0/σ^2

- For $G_n < 1$, $T = 0$ is no longer the best choice. In fact, one should select $T \geq 4\sigma$. In this case, the proposed reconstruction scheme yields better reconstruction accuracy compared to Whittaker-Shannon interpolation scheme.

Taking the above observation into account, we further propose the following joint detection/estimation scheme for reconstruction of bandlimited signals from noisy data:

Step 1: Estimating $P = \sum_{|k| \leq n} x_k^2$ from data. Computing G_n using the estimated P

Step 2: If $G_n \geq 1$, setting $T = 0$. Otherwise, setting $T = 4\sigma$.

Step 3: Passing data through the thresholding device, and then, the reconstruction filter to reconstruct the original signal.

Remark 3.1. For sufficiently large n and/or for fast decay functions, we do not need to estimate P in order to determine G_n. In this case we have $\sum_{|k| \leq n} x_k^2 \approx E_0/\tau_q$, which implies $G_n \approx E_0/(\tau_q \sigma^2 (2n+1))$. We assume that the signal-to-noise ratio E_0/σ^2 is known in advance.

Remark 3.2. For n sufficiently large and/or for fast decay functions, we have $MISE(\widetilde{x}_n) \approx \tau_q \sigma^2 (2n+1)$. Let $R = MISE(\widehat{x}_n)/MISE(\widetilde{x}_n)$. We can re-write (3.2) as follows:

$$R \leq 1 - f(G_n, T) = 1 - f\left(\frac{E_0}{\tau_q \sigma^2 (2n+1)}, T\right). \quad (3.3)$$

Figure 2 plots the right-hand side of (3.3) as the function of E_0/σ^2 when the duration of measurement $\tau_q(2n+1) = 1$ second, $T = 0$ and $T = 4\sigma$, respectively. The plot clearly shows the advantage of using thresholding-based reconstruction scheme for the region of low SNR

The behavior of Δ and R as $n \to \infty$ is stated in the following theorem:

Theorem 3.1. *Selecting $T = 4\sigma$, then $\Delta \to \infty$ and $R \to 0$ as $n \to \infty$*

4. SIMULATION RESULTS

The signal we selected for simulation is

$$x(t) = \sqrt{2 f_{max}} \mathrm{sinc}(2 f_{max} t).$$

with $f_{max} = 3900$ Hz. Therefore, $x(t)$ is a unit energy $BL(\Omega)$ signal, where $\Omega = 2\pi f_{max}$ is the corresponding radian frequency.

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

$$EMISE(\widehat{x}_n) = \frac{\bar{\tau}}{M} \sum_{j=1}^{M} \sum_{k=-n}^{n} [\widehat{x}(k\bar{\tau}) - x(k\bar{\tau})]^2. \quad (4.1)$$

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

Figure 3 plots $EMISE(\widehat{x}_n)$ as a function of E_0/σ^2 when $T = 4\sigma$ was chosen. In this simulation, $2n = 100$ and $\tau_q = 1.25 \cdot 10^{-4}$ seconds (sampling frequency is just slightly larger than the Nyquist rate) were selected. For comparison, $EMISE(\widetilde{x}_n)$ and $EMISE$ of reconstruction schemes based on moving average (MA) filtering and median filtering are also plotted. For MA-based and median-based signal reconstructions, the over-sampling factor of 2, length-3 moving average filter and length-3 median filter were used. The plot clearly shows the advantage of using the proposed reconstruction scheme over the others in the region of low SNR. The figure also indicates that the signal reconstruction scheme based on thresholding is very robust again noise variance, while the Whittaker-Shannon interpolation scheme is not.

Figure 4 plots $EMISE(\widehat{x}_n)$ as a function of n when $T = 4\sigma$ was chosen. For comparison, $EMISE(\widetilde{x}_n)$ is also plotted. In this simulation, $E_0/\sigma^2 = -5$ dB and $\tau_q = 1.25 \cdot 10^{-4}$ seconds were selected. The plot clearly

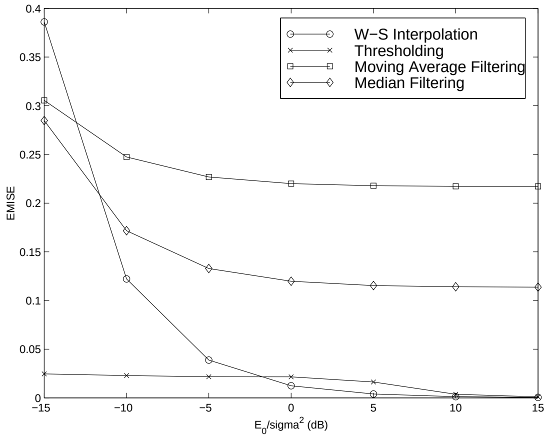

Fig. 3. $EMISE(\widehat{x}_n)$ versus E_0/σ^2

shows that the $EMISE$ of the proposed reconstruction scheme is less dependent on n compared to that of Whittaker-Shannon interpolation scheme. It also indicates that the performance gap between two reconstruction schemes becomes wider as n increases.

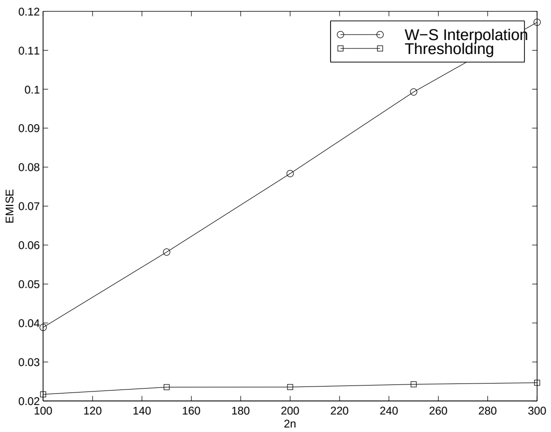

Fig. 4. $EMISE(\widehat{x}_n)$ versus n

5. CONCLUSIONS

A signal reconstruction scheme based on thresholding was proposed for recovering band-limited signals from noisy data. It was shown that in the region of low SNR, the proposed reconstruction scheme gives better reconstruction accuracy compared to the classical reconstruction scheme based on Whittaker-Shannon interpolation. This improvement can be achieved without requiring oversampling. In fact, sampling at the Nyquist rate was considered in this paper. In contrast, in the region of high SNR, it is always better to use the Whittaker-Shannon interpolation for signal reconstruction.

In addition, by adaptively tuning the threshold value T with noise variance, the robust signal reconstruction can be obtained. The MISE of the proposed reconstruction scheme increases as n increases, but at much slower rate compared to that of the classical Whittaker-Shannon interpolation scheme.

6. REFERENCES

[1] P.L. Butzer and R.L. Steins, "Sampling theory for not necessary band-limited functions: A historical overview," *SIAM Review*, vol. 34, no. 1, pp. 40–53, Mar. 1992.

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

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

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

[5] A. Krzyzak, E. Rafajlowicz, and M. Pawlak, "Moving average restoration of bandlimited signals from noisy observations," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2967–2976, Dec. 1997.

[6] M. Pawlak, E. Rafajlowicz, and A. Krzyzak, "Exponential weighting algorithms for reconstruction of bandlimited signals," *IEEE Transactions on Signal Processing*, vol. 44, pp. 538–545, Mar. 1996.

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

[8] 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.

[9] D. L. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425– 455, Dec. 1994.

[10] M. Jansen, *Noise Reduction by Wavelet Thresholding*, Lecture Notes in Statistics. Springer, New York, NY, 2001.

[11] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, 4th edition, 1980.