16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

ADAPTIVE WINDOW FOR LOCAL POLYNOMIAL REGRESSION FROM NOISY NONUNIFORM SAMPLES

A. Sreenivasa Murthy[1] *and T. V. Sreenivas*[2]

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore, India
Emails: {[1]asmuvce,[2]tvsree}@ece.iisc.ernet.in

ABSTRACT

We consider the problem of local polynomial regression of noisy nonuniform samples of a time-varying signal in the presence of observation noise. We formulate the problem in the time domain and use the pointwise minimum mean square error (MMSE) as the cost function. The choice of the window length for local regression introduces a bias-variance tradeoff which we solve by using the *intersection-of-confidence-intervals* (ICI) technique. This results in an adaptive pointwise MMSE-optimal window length. The performance of the adaptive window technique is superior to the conventional fixed window approaches. Simulation results show that the improvement in reconstruction accuracy can be as much as 9 dB for 3 dB input signal-to-noise ratio (SNR).

1. INTRODUCTION

Sampling is a fundamental operation in digital signal processing. Nearly six decades of signal processing advances have been made possible thanks to the Shannon-Whittaker-Kotelnikov uniform sampling theorem [1]. However, uniform sampling comes with its own limitations. In many cases, nonuniform sampling is unavoidable or natural in the acquisition process. Consider for example signals sampled at instants where they cross a certain threshold (level-crossing sampling). Such event-based sampling schemes have certain advantages which may be useful in some applications [2, 3]. Sometimes, nonuniform sampling is deliberately incorporated to overcome the limitations of uniform sampling. For example, in parameter estimation of phase signals, nonuniform sampling overcomes the parameter aliasing problem that is often encountered in uniform sampling [3, 4]. Yet another scenario in which nonuniform sampling is natural is kernel density estimation [7].

Unlike uniform sampling, there are no generalized results for signal reconstruction from nonuniform samples. Reconstruction in the nonuniform case is mostly done by using suitable basis functions and iterative/noniterative algorithms are used with a certain regularization constraint. The basis functions can be localized or nonlocalized depending on the properties of the underlying signal. In [5], the authors present a multilevel algorithm for regularized reconstruction of a signal from noisy nonuniform samples using trigonometric polynomials as basis functions. In [9], the authors propose a polynomial filtering technique but the solution is not associated with a mean square error (MSE) criterion. The objective of this paper is to overcome this drawback by formulating a MSE cost function.

In many practical applications, the nonuniform samples are also corrupted by measurement noise. In such a scenario, one is interested in estimating the underlying signal by the optimization of a suitable cost function. In this paper, we propose the use of a pointwise minimum mean square error (PMMSE) cost function which is also suitable for time-varying signals. We choose the simplest of basis functions, namely, the polynomial, and address the reconstruction problem in the framework of local polynomial regression. Local polynomial regression is a well established area of research and has found many applications [7, 8]. One can also choose a different set of basis functions, but we prefer to use polynomials because of their simplicity.

We show that the optimization of the pointwise mean square error cost function leads to a bias-variance tradeoff, a problem that is well known in the statistical signal processing literature [10, 11]. We show that the tradeoff can be solved in a near-optimal fashion by using the *intersection-of-confidence-intervals*(ICI) technique. In doing so, we show that local polynomial regression together with the PMMSE criterion adapts to the signal local behavior.

2. PROBLEM FORMULATION

Consider the practical sampling scenario shown in Fig. 1, in which the output of an analog source $s(t)$ is corrupted by additive noise $w(t)$. The samples of the process

$$y(t) = s(t) + w(t) \tag{1}$$

are taken at ordered nonuniform instants $\{t_n, n \in \mathbb{Z}\}$. The noise may be inherent in the recording system or it may be due to the channel. We assume that $w(t)$ is white Gaussian in nature, zero-mean and variance σ_w^2. This scenario occurs in many practical cases such as observation of astronomical data, data measured in a moving vehicle for oceanographic applications, magnetic/gravitational field measurements in geophysics, jitter in sampling, data loss in images or audio signals due to channel erasures etc. The observations are given by

$$y(t_n) = s(t_n) + w(t_n). \tag{2}$$

The sampling instants are distinct and strictly ordered in time i.e., $t_n < t_{n+1}$. Sampling can be random or deterministic; in this paper, we assume the latter.

The objective is to compute an estimate of $s(t)$ from the noisy measurements, $\{y(t_n), n \in \mathbb{Z}\}$ such that the mean square error is least. We do not assume that $s(t)$

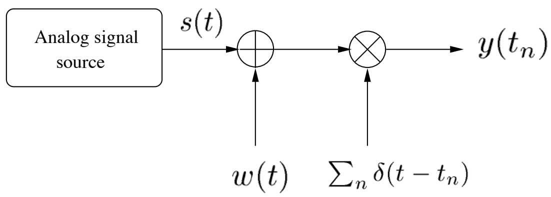

Figure 1: Schematic of a nonuniform sampling system.

is bandlimited; infact, many natural signals are not bandlimited. We assume that the signal is continuous and smooth so as to permit a local polynomial representation to a desired accuracy (Weierstrass's theorem [6]).

3. POINTWISE MINIMUM MEAN SQUARE ERROR

Let t be a point where the estimate of $s(t)$ is computed by a local p^{th} order polynomial regression to the signal $s(t)$. The least squares approximation error is given by

$$\mathcal{C}(t,\mathbf{a}) = \sum_n \left[y(t_n) - \sum_{k=0}^{p} a_k(t) t_n^k \right]^2 h(t-t_n) \tag{3}$$

where $\mathbf{a} = [a_0(t) \;\; a_1(t) \;\; a_2(t) \;\; \ldots \;\; a_p(t)]$ are the time-varying polynomial coefficients. The polynomial approximation is localized by a window function $h(t)$ which is chosen to be positive-valued, symmetric and satisfies the properties:

$$\int_{-\infty}^{+\infty} h(t)dt = 1 \;\text{ and } \int_{-\infty}^{+\infty} h^2(t)dt < \infty. \tag{4}$$

The coefficients, $\{a_k(t),\ 0 \le k \le p\}$ are functions of time and are estimated as

$$\widehat{a}_\ell(t) = \arg\min_{a_\ell} \mathcal{C}(t,\mathbf{a}). \tag{5}$$

By using the polynomial coefficient estimates, we can compute an estimate $\widehat{s}(t)$ of the signal $s(t)$. Usually, one is interested in the behaviour of the estimator as the number of samples within the observation window tends to infinity. This limiting case leads to asymptotic bias and variance expressions, which in the present case are given by

$$\text{Bias}\left(\widehat{s}(t)\right) = \frac{(-1)^{p+1} s^{p+1}(t)}{(p+1)!} \frac{\int \beta^{p+1} h(\beta) d\beta}{\int h(\beta) d\beta} \tag{6}$$

and

$$\text{Var}\left(\widehat{s}(t)\right) = \sigma_w^2 \frac{\int h^2(\beta) d\beta}{\left[\int h(\beta) d\beta\right]^2}. \tag{7}$$

The derivations for the above expressions can be found in [7]. Although the context in [7] is one of kernel density estimation, the expressions are valid for our problem too since the underlying theme is the same namely, local polynomial regression.

From (6) and (7), we note that the bias and variance depend on $h(t)$. We now study the effect of $h(t)$, its shape and size. In this paper, we confine ourselves to two commonly-encountered window functions:

1. *Rectangular window*: The rectangular window of compact time support σ_h is defined as

$$h(t) = \begin{cases} \frac{1}{\sigma_h} & \text{for } |t| < -\frac{\sigma_h}{2} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The corresponding bias is given by

$$\text{Bias}\left(\widehat{s}(t)\right) = \begin{cases} \dfrac{(-1)^{p+1} s^{p+1}(t)}{2^{p+1}(p+2)!} \sigma_h^{p+1} & p \text{ odd} \\ 0 & p \text{ even}, \end{cases} \tag{9}$$

and the variance is given by

$$\text{Var}\left(\widehat{s}(t)\right) = \frac{\sigma_w^2}{\sigma_h}. \tag{10}$$

2. *Gaussian window*: The infinite-support centered-Gaussian window of variance σ_h^2 given by

$$h(t) = \frac{1}{\sqrt{2\pi\sigma_h^2}} e^{-\frac{t^2}{2\sigma_h^2}}. \tag{11}$$

yields
Bias $\left(\widehat{s}(t)\right) =$

$$\begin{cases} \dfrac{(-1)^{p+1} s^{p+1}(t)}{(p+1)!} 1.3.5...p\ \sigma_h^{p+1} & p \text{ odd} \\ 0 & p \text{ even} \end{cases} \tag{12}$$

and

$$\text{Var}\left(\widehat{s}(t)\right) = \frac{\sigma_w^2}{2\sqrt{\pi}\sigma_h}. \tag{13}$$

With respect to the window parameter σ_h, the bias and variance of the estimator show complementary characteristics. As $\sigma_h \to 0$, $\text{Bias}\left(\widehat{s}(0)\right) \to 0$ and $\text{Var}\left(\widehat{s}(0)\right) \to \infty$. On the other hand, as $\sigma_h \to \infty$, $\text{Bias}\left(\widehat{s}(0)\right) \to \infty$ and $\text{Var}\left(\widehat{s}(0)\right) \to 0$.

For the same value of σ_h, the Gaussian window function yields an estimate with a variance that is $2\sqrt{\pi}$ times less than that obtained by the rectangular window. The bias is appreciably smaller with the rectangular window than the Gaussian window. For example, with $p = 3$, the bias with the rectangular window is 80/3 times smaller than that obtained with the Gaussian window. This reduction in bias is more than the reduction in variance that the Gaussian window offers. Also, in a practical application, due to the finite data availability, the rectangular window is a better choice.

Since the bias and variance have complementary characteristics with respect to σ_h, it is not possible to minimize both. Instead, one can minimize the mean square error $\text{MSE}(\widehat{s}(t)) = \mathcal{E}\left\{[s(t) - \widehat{s}(t)]^2\right\}$. The MSE is equal to the sum of the squared bias and the variance and is given by:

$$\text{MSE}(\widehat{s}(t)) = \left[\text{Bias}(\widehat{s}(t))\right]^2 + \text{Var}(\widehat{s}(t)) \tag{14}$$

$$= \left(\frac{s^{p+1}(0)}{2^{p+1}(p+2)!}\right)^2 \sigma_h^{2p+2} + \frac{\sigma_w^2}{\sigma_h}. \tag{15}$$

The optimum MMSE choice of σ_h is given by

$$\begin{aligned}\sigma_{h_{\text{opt}}} &= \arg\min_{\sigma_h} \text{MSE}(\widehat{s}(t)) \\ &= \left[\frac{\sigma_w^2 2^{2p+2}((p+2)!)^2}{(2p+2)(s^{p+1}(t))^2}\right]^{\frac{1}{2p+3}} \qquad (16)\end{aligned}$$

Since the MSE is a function of time (pointwise MSE), we refer to the above optimization as the pointwise MMSE criterion. The optimal value of $\sigma_{h_{\text{opt}}}$ requires apriori knowledge of the higher-order derivatives of $s(t)$ which may not be available in practice. However, by exploiting the complementary bias-variance characteristics, one can obtain a near-optimal solution by using the ICI technique which we present next.

4. NEAR-OPTIMAL PMMSE SOLUTION BY THE ICI TECHNIQUE

The ICI technique, in its present form, proposed by Stankovic and Katkovnik [12] has been successfully applied to a variety of problems to solve the bias-variance tradeoff. Some applications of the ICI technique include instantaneous frequency estimation by using time-frequency distributions [12, 13] and zero-crossings [14]. In this paper, we promote the application of the ICI technique for the problem of reconstruction from noisy nonuniform samples.

The summary of the ICI technique is given below. Define $\mathcal{H} = \left\{\sigma_h^{(i)} \middle| \sigma_h^{(i)} = 2^i \sigma_h^{(0)}, \quad i = 0, 1, 2, ..., i_{max}\right\}$. i_{max} is chosen such that $2\sigma_h^{(i_{max})}$ is the largest number that is less than the number of samples observed.

1. *Initialization*: $i = 0$; $\sigma_h^{(0)}$ is chosen as the window length about t encompassing $(p+1)$ data samples. The signal estimate, $\widehat{s}(t)$ is obtained by local polynomial regression to the centered-data, with $\sigma_h^{(0)}$ as the window parameter. Denote the estimate by $\widehat{s}^{(0)}(t)$. We compute the 2κ confidence interval $\mathcal{J}_0$ as

$$\mathcal{J}_0 = \left[\widehat{s}^{(0)}(t) - \kappa\sigma, \widehat{s}^{(0)}(t) + \kappa\sigma\right]. \qquad (17)$$

2. *Iteration*: With the window parameter $\sigma_h^{(i+1)}$, we compute the estimate, $\widehat{s}^{(i+1)}(t)$ and the associated confidence interval:

$$\mathcal{J}_{i+1} = \left[\widehat{s}^{(i+1)}(t) - \kappa\sigma, \widehat{s}^{(i+1)}(t) + \kappa\sigma\right]. \qquad (18)$$

3. *Stopping Condition*: If $\mathcal{J}_i \cap \mathcal{J}_{i+1} = \emptyset$ (empty set), then $\sigma_{h_{opt}} = \sigma_i$, else $i \leftarrow i+1$ and go to step 2 above.

The confidence interval computation requires the knowledge of σ_w but it can be estimated. Apriori knowledge of the bias is not needed wherein lies the advantage of the ICI technique. The only parameter that needs to be determined is the confidence interval parameter κ. κ should be as small as possible and still keep the probability of choosing a wrong window length below a certain threshold. A reliability analysis conducted in [12] showed that $\kappa = 2.5$ is a reasonable choice. The corresponding confidence interval coverage probability is approximately 0.99. Earlier results on local polynomial regression [7] showed that third-order polynomials are optimal for yielding smooth estimates; therefore, we use $p = 3$.

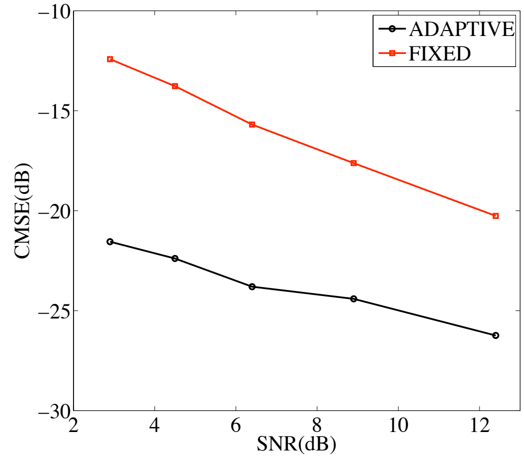

Figure 2: Cumulative mean square error performance for fixed vs. adaptive window techniques as a function of the SNR.

5. EXPERIMENTAL RESULTS

5.1 Experiment-1

First, we conduct an experiment using $s(t) = e^{-\alpha t}\sin(\omega t)$, $\alpha = 0.005, \omega = 0.05$rad/s. The noise is white Gaussian with a variance σ_w^2. An arbitrary sampling sequence is generated at the beginning of the experiment by using a pseudorandom generator associated with a uniform distribution. We fix the sampling sequence for all the statistical Monte-Carlo realizations of the noisy signal since we have assumed a deterministic sampling sequence. In the experiment, the noisy signal is sampled at 256 locations in $[0, 255]$ as determined by the sampling sequence. The signal $s(t)$ is estimated at 256 uniform locations $[0, 1, 2, ..., 255]$ using the adaptive window technique. The initial window length is chosen as the interval encompassing $(p+1)$ samples. At each iteration of the window parameter estimation technique, the increment is chosen as the window length encompassing $(p+1)$ additional samples. We repeat the experiment 100 times with a different random realization of the noise sequence at each repetition. Since the signal $s(t)$ is known to us in this simulated experiment, we can compute the cumulative mean square error (CMSE) to quantify the estimation accuracy. The CMSE is given by

$$\text{CMSE} = \frac{1}{RN}\sum_{r=1}^{R}\sum_{n=1}^{N}\left[\widehat{s}^{(r)}(n) - s(n)\right]^2, \qquad (19)$$

where $s^{(r)}(n)$ is the estimate of $s(n)$ in the r^{th} realization (R: number of realizations), and where N is the number of points at which the estimates are obtained. The CMSE accumulates the MSE at each point in the observation interval and gives an objective measure for

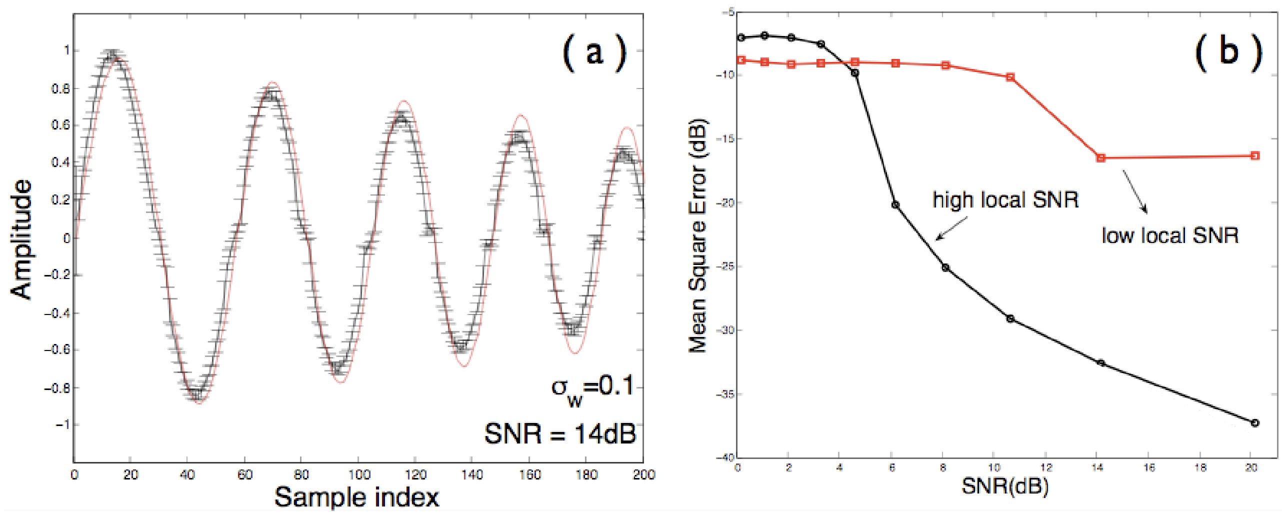

Figure 3: (a) The estimated signal with the 2σ-confidence interval and the actual signal; (b) Mean square error performance as a function of the SNR.

performance comparison. To show that the adaptive window technique is superior to a fixed window technique, we repeat the experiment with a fixed window of k samples; in our experiments, we chose $k = 7$. The CMSE corresponding to the fixed and adaptive window lengths are shown in Fig. 2. Note that the adaptive technique is consistently superior to the fixed window technique. The CMSE of the adaptive window regression is lower than that of the fixed window by about 9 dB for 3 dB input SNR. As SNR increases, the CMSE of both techniques decreases but the accuracy advantage of the adaptive technique still exists.

5.2 Experiment-2

We conduct another experiment with a linear frequency-modulated chirp with an exponentially decaying envelope. This can also be seen as a synthesized isolated formant in a speech signal. We chose the envelope to decay by a factor of $\frac{1}{e}$ within the observation window. We chose the frequency to increase linearly from 0.05 to 0.1 (these are normalized frequencies with 0.5 corresponding to the Nyquist sampling frequency). The estimated signal and the mean square error performance at the 111^{th} position (arbitrarily chosen point corresponding to high local SNR) and 200^{th} position (corresponding to low local SNR) in the window is shown in Fig. 3. Note that the new technique offers accurate estimation for moderate to high SNRs even for time-varying signals. The signal estimation accuracy is higher in regions of high local SNR than in regions of poor local SNR, which is intuitive and acceptable.

6. CONCLUSIONS

We addressed the problem of signal reconstruction from noisy nonuniform samples within the framework of local polynomial regression, using a pointwise minimum-mean square error criterion. This involves a bias-variance tradeoff, which we solved by using the intersection-of-confidence-intervals technique. The technique enables one to tradeoff adaptability to signal variations (bias term) and robustness to noise (variance term). The performance of the technique is superior to the fixed window-based approach.

Software

MATLAB software for the techniques reported in this paper can be requested from the authors by email.

REFERENCES

[1] M. Unser, “Sampling50 Years After Shannon,” *Proc. IEEE*, Vol. 88, No. 4, pp. 569-587, April 2000.

[2] F. Marvasti, *Nonuniform sampling: Theory and practice*, Kluwer Academic Publishers, New York, 2001.

[3] S. C. Sekhar and T. V. Sreenivas, “Auditory motivated level-crossing approach to instantaneous frequency estimation” *IEEE Trans. on Sig. Proc.*, Vol. 53, Issue 4, pp. 1450-1462, Apr 2005.

[4] J. A. Legg and D. A. Gray, “Performance bounds for polynomial phase parameter estimation with nonuniform and random sampling schemes” *IEEE Trans. on Sig. Proc.*, Vol. 48, Issue 2, pp. 331 - 337, Feb 2000.

[5] M. Rauth and T. Strohmer, “Smooth approximation of potential fields from noisy scattered data”, *Geophysics*, Vol. 63, No. 1, pp. 85-94, 1998.

[6] S. K. Sen and E. V. Krishnamurthy, Numerical Algorithms: Computations in Science and Engineering, *Affiliated East-West Press*, 2001.

[7] J. Fan and I. Gijbels, *Local polynomial modeling and its application*, Chapman and Hall, London, 1996.

[8] W. Hardle, *Applied nonparametric regression*, Cambridge University Press, Cambridge, 1990.

[9] T.I. Laakso, A. Tarczynksi, N.P. Murphy and V. Valimaki, "Polynomial filtering approach to reconstruction and noise reduction of nonuniformly sampled signals", *Signal Processing*, pp. 567-575, 2000.

[10] S. M. Kay, Modern spectral estimation: Theory and applications, *Prentice Hall*, 1988.

[11] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Upper Saddle River, New Jersey, USA: *Prentice-Hall*, 1993.

[12] L. Stankovic and V. Katkovnik, "Algorithm for the instantaneous frequency estimation using time frequency distributions with adaptive window length", *IEEE Sig. Proc. Lett.*,, vol. 5, No.9, pp. 224-227, Sept 1998.

[13] S. C. Sekhar and T. V. Sreenivas, "Adaptive spectrogram vs. adaptive pseudo-Wigner-Ville distribution for instantaneous frequency estimation", *Signal Proc.*, vol. 83 no.7, pp. 1529-1543, July 2003.

[14] S. C. Sekhar and T. V. Sreenivas, "Adaptive Window Zero-Crossing-Based Instantaneous Frequency Estimation," *EURASIP J. on Applied Sig. Proc.*, vol. 2004, no. 12, pp. 1791-1806, 2004.