DISCRETE DATA SMOOTHING TECHNIQUE FOR NOISE REDUCTION

Masoud Ghoreishi Madiseh
e-mail: mghoreishi@iust.ac.ir

Vahid Tabataba Vakily
e-mail: vakily@iust.ac.ir

Iran University of Science and Technology, Department of Electrical Engineering, 16844, Tehran, Iran

Key words: data smoothing, sampling, estimation, Mean Square Error (MSE), colored noise

ABSTRACT

In the abovementioned paper, we expand a new algorithm for reduction the noise effect with respect to Mean Square Error Criterion for discrete signals. This technique had been used for continues signals and we could see the effect of time duration and shape of impulse response of filter on MSE and the optimum filter had been designed. In this paper, we settle the digital filter coefficients-- length, shape— to decrease the MSE error and better elimination of noise. The simulation results prove that we can control the interval of MSE by changing parameters of deterministic function in this model and increasing the input sample number.

I. INTRODUCTION

Suppose that we wish to estimate an unknown signal f(t) in terms of the observed value of sum x(t) = f(t) + v(t). We assume that noise v(t) is white with know autocorrelation $R(t_1,t_2) = q(t_1)\delta(t_1 - t_2)$. We use from smoothing and our estimator is the impulse response ***y(t)***of the filter ***h(t)***:

$$y(t) = \int_{-\infty}^{+\infty} x(t-\tau)h(\tau)d\tau \qquad (1)$$

Estimator is biased and biased error is:

$$b = y_f(t) - f(t) = \int_{-\infty}^{+\infty} f(t-\tau)h(\tau)d\tau - f(t) \qquad (2)$$

And variance is:

$$\sigma^2 = E[y_v^2(t)] = \int_{-\infty}^{+\infty} q(t-\tau)h^2(\tau)d\tau \qquad (3)$$

We should find ***h(t)*** so as to minimize the MS error:

$$e = E\{[y(t) - f(t)]^2\} = b^2 + \sigma^2 \qquad (4)$$

We suppose the following assumptions for ***h(t)***. It is for simplification:

$$h(-t) = h(t)\,;\ \int_{-T}^{T} h(t)dt = 1\,;\ h(-t) > 0 \qquad (5)$$

T is ***h(t)*** duration and We could adapt the MS error with this parameter. If ***T*** is small the value $y_f(t) \cong f(t)$, hence the bias error is small; however the variance is large. When ***T*** increases, the variance decreases but the bias error increases. In general determination of the optimum ***T*** is complicated but here we suppose some assumption for developing a simple solution. We suppose that ***f(t)***, ***q(t)*** are smooth in the sense that ***f(t)*** can be approximated by a parabola and ***q(t)*** by a constant in any interval of length ***2T***. From this assumption with respect to Taylor expansion it follows that:

$$f(t-\tau) \cong f(t) - \tau f'(t) + \frac{\tau^2}{2} f''(t)\,,$$

$$q(t-\tau) \cong q(t) \qquad (6)$$

We conclude that:

$$b \cong \frac{f''(t)}{2}\int_{-T}^{T}\tau^2 h(\tau)d\tau\ ,\ \sigma^2 \cong q(t)\int_{-T}^{T} h^2(\tau)d\tau \qquad (7)$$

With attention to (5), the resulting MS error equals [3]:

$$e \cong \frac{1}{4}M^2[f''(t)]^2 + Eq(t) \qquad (8)$$

Where $M = \int_{-T}^{T}\tau^2 h(\tau)d\tau$ and $E = \int_{-T}^{T} h^2(\tau)d\tau$.

We introduce normalized filter in this form:

$$w(t) = Th(Tt) \qquad (9)$$

And define the parameters of ***w(t)***:

$$M_w = \int_{-1}^{1}\tau^2 w(\tau)d\tau\ ,\ E_w = \int_{-1}^{1} w^2(\tau)d\tau \qquad (10)$$

It follows from (5) and (8) that

$$b \cong \frac{T^2}{2}M_w f''(t)\ ,\ \sigma^2 \cong \frac{E_w}{T}q(t) \qquad (11)$$

$$e \cong \frac{1}{4}T^2 M_w^2[f''(t)]^2 + \frac{E_w}{T}q(t) \qquad (12)$$

Thus "***e***" depends on the shape of ***w(t)*** and it's duration, ***T*** [3] (Fig. 1). We assume first that ***w(t)*** is specified. In Fig.1 we plot the bias ***b***, the variance σ^2, and the MSE error ***e*** as functions of ***T***. As ***T*** increases, ***b*** increases, and σ^2 decreases. Their sum ***e*** is minimum for:

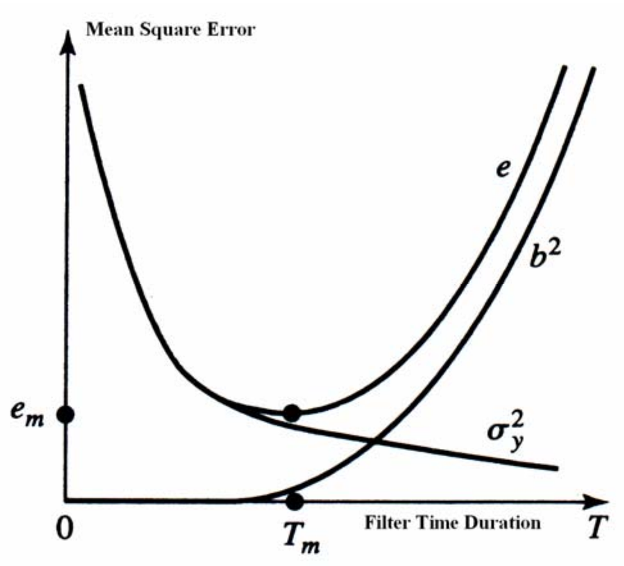

Fig. 1. Mean Square Error with respect to different time duration of smoothing filter.

$$T_m = \left(\frac{E_w q(t)}{M_w^2 [f''(t)]^2}\right)^{\frac{1}{5}} \quad (13)$$

And we conclude that: $\sigma = 2b$

And we can see in [4,2] the shape of ***w(t)*** to minimizing (12) Is determined and it's parameters are as follows:

$$w(t) = \begin{cases} 0.75(1-t^2) & ; \quad |t| \le 1 \\ 0 & ; \quad else \end{cases} \quad (14)$$

And $E_w = \frac{3}{5}$, $M_w = \frac{1}{5}$ [1,4].

II. SAMPLING

Assume that signal ***x(t)*** will be sampling with frequency f_s, which $f_s \ge Max(2f_1, 2f_2)$ [5,7] is Nyquist condition and f_1, f_2 are maximum frequency of ***f(t)*** and ***q(t)***. Now we want to study about effects of filter time duration, Number of data samples and type of ***f(t)***,***q(t)*** on estimation error of discrete data which is input of smoothing filter [5,6]. It is obviously clear when ***f(t)*** and ***q(t)*** are smooth in interval **[*0,T*]** then they are also smooth in interval **[*0,Δt*]** where ***Δt=T/M*** and ***M*** is the maximum number of samples.

In this method, we suppose the time duration of filter ***w(t)*** equal $1/f_s$ and try to sample from continues signal with frequency more than Nyquist frequency. The sake of this selection is descent of MSE with respect to number of samples. It means that with newer samples we make better estimation and error will be decreased. At last we determine total MSE of all input samples with simple summation of each sample error:

$$e_t = \sqrt{\sum_{n=0}^{M} \frac{e_n^2}{M}} \quad (15)$$

Where e_n is the MS error for ***n***th sample.

III. ESTIMATION ERROR CALCULATION

As mentioned above, we suppose filter time duration is $T = 1/f_s$ and the sample time is **Δ*t=T/M***. The equation (15) shows that the total estimation error if function of ***M*** (sample numbers) and e_n is variable with respect to ***f(t)*** and ***q(t)*** then e_t is variable with different ***f(t)*** and ***q(t)*** too. In this algorithm we suppose for example $f(t) = f_1(t)$ and $q(t) = q_1(t)$ then find some point (almost 5 points) from ***f(t)*** in the interval **[0,T]** and try to fit a parabola on this points with software assistance. When we find specified parabola, put its second derivative $f''(t)$ which is constant ***a***, in equation (12). For ***q(t)*** we find the middle point of interval and because of smooth property of ***q(t)*** we select $t = (n+0.5)\Delta t$ and put it in the equation (12) as constant value. Then we have:

$$x(n) = x(t = (n+0.5)\Delta t))$$

$$e_n = \frac{1}{4}(\Delta t)^2 M_w^2 [\underbrace{f_1''(t)}_{a}]^2 + \frac{E_w}{\Delta t} q_1(t)$$

$$E_w = \frac{3}{5}\ , \ M_w = \frac{1}{5}\ , \ \Delta t = \frac{T}{M} \quad (16)$$

$$e_n = \frac{1}{100}\left(\frac{T}{M}\right)^2 [a]^2 + \frac{3M}{5T} q_1\left(\frac{(n+0.5)T}{M}\right)$$

Now we substitute e_n from (16) in equation (15) and find the total error with respect to *M* number of samples.

IV. SIMULATION RESULTS

In multi-level digital communication systems function ***f(t)*** usually is a single carrier with different amplitude then we suppose that ***f(t)=A+Bsin(wt)*** and ***q(t)*** could be the simplest non-linear function which could be supposed $e^{b(t-c)}$ which ***c*** is the delay time. We try to depict the MSE for ***M*=100** and the result for below values of ***f(t)*** and ***q(t)*** has been draw in Fig. 2.

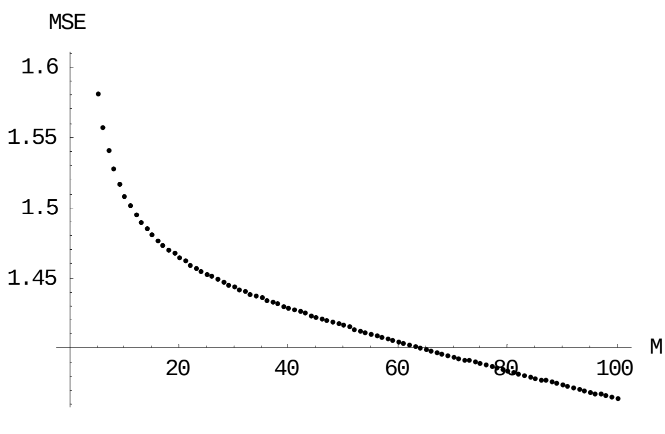

Fig. 2. $f_1(t) = 1 + 2\sin(2\pi t)$, $q_1(t) = 1 - e^{-0.003(t-4)}$

In Fig. 2 the MSE is decreased from 2.04 to 1.36 for 100 first samples. Now if we increased ***f(t)*** frequency 10 time and the ***q(t)*** be the same as before, the result has been shown in Fig. 3.

It is clear that the MSE is increased with increase of ***f(t)*** frequency but the shape of MSE decreases with respect to *M*.

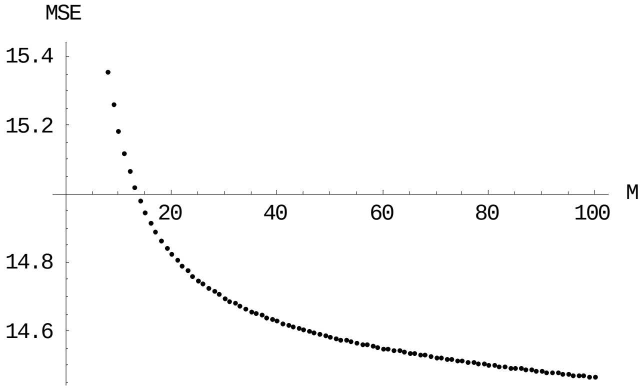

Fig. 3. $f_2(t) = 1 + 2\sin(20\pi t)$, $q_2(t) = 1 - e^{-0.003(t-4)}$

Now suppose that ***f(t)*** is the same as Fig. 3 but we change ***q(t)*** parameters ***b*** and ***c*** in a way to decrease MSE. We set new values for these parameters and the MSE error decrease between 0.99 and 1.52 which has been shown in Fig. 4. The graph is decreasing when ***M*** increases.

Again we increase ***f(t)*** frequency 10 time and draw the MSE as Fig. 5.

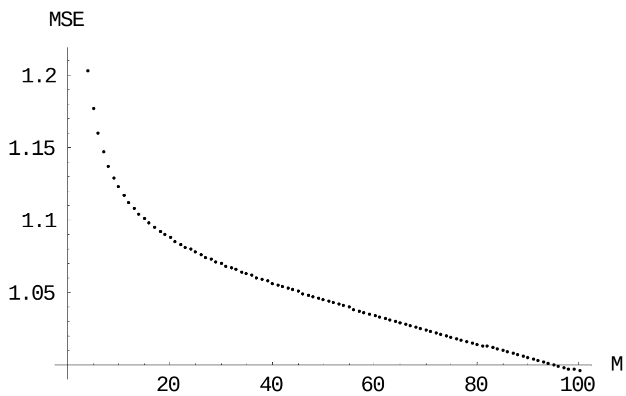

Fig. 4. $f_3(t) = 1 + 2\sin(20\pi t)$, $q_3(t) = 1 - e^{-0.003(t-0.3)}$

As Fig. 5 shows, the MSE range has been increased very much again. In the next step we try to reduce error range with setting parameters of ***q(t)***. The new result has been plotted in Fig. 6.

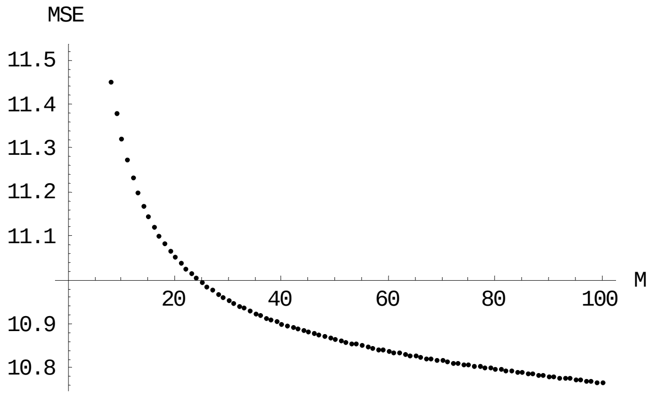

Fig. 5. $f_4(t) = 1 + 2\sin(200\pi t)$, $q_4(t) = 1 - e^{-0.003(t-0.3)}$

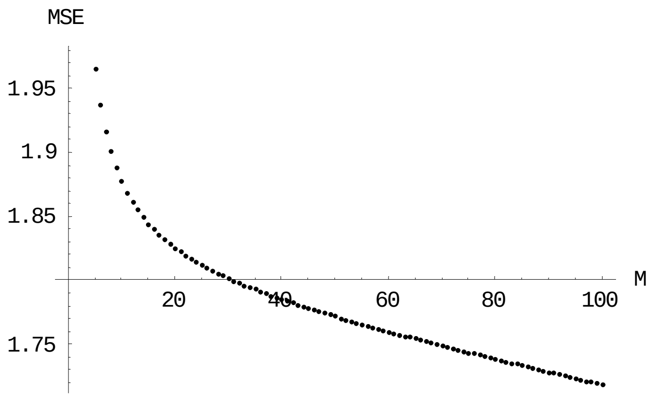

Fig. 6. $f_5(t) = 1 + 2\sin(200\pi t)$, $q_5(t) = 1 - e^{-0.003(t-0.05)}$

We can conclude that the error range will be varied with respect to ***f(t)*** and ***q(t)*** then we suppose this model for parameters of these functions and find the minimum of error for this model.

$$f_6(t) = 1 + 2\sin(2\pi t) \ , \ q_4(t) = 1 - e^{-b(t-c)} \qquad (17)$$

Fig.7 shows the effect of ***b*** and ***c*** parameters on error when M=100. We find the minimum of error with these values:

b=1.6E-14, c=0.32

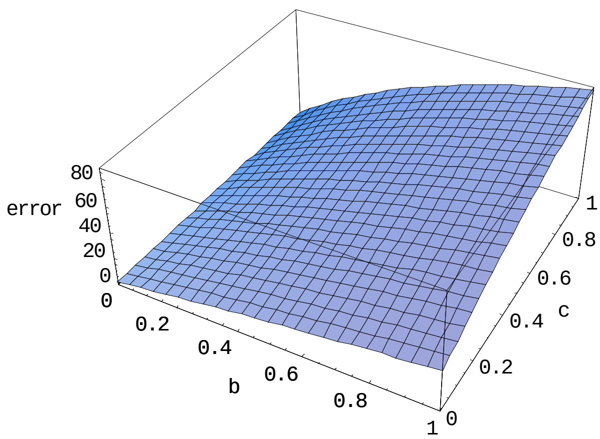

Fig. 7. $f_6(t) = 1 + 2\sin(2\pi t)$, $q_4(t) = 1 - e^{-b(t-c)}$

Now draw MSE with respect to M for above values of b and c. the result has been shown in Fig. 8. You can see that the MSE value increases with respect to M and this is not our desirable. Then we change value *b* to find the descent graph.

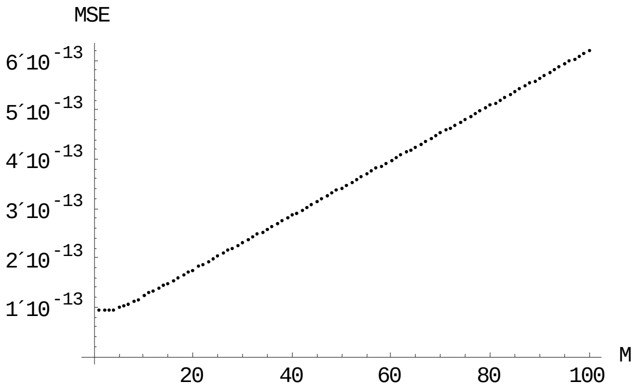

Fig. 8. $f_7(t) = 1 + 2\sin(2\pi t)$, $q_7(t) = 1 - e^{-1.6\times10^{-14}(t-0.32)}$

We set typical value 0.03 for *b* and MSE is drawn in Fig. 9 for this assumption.

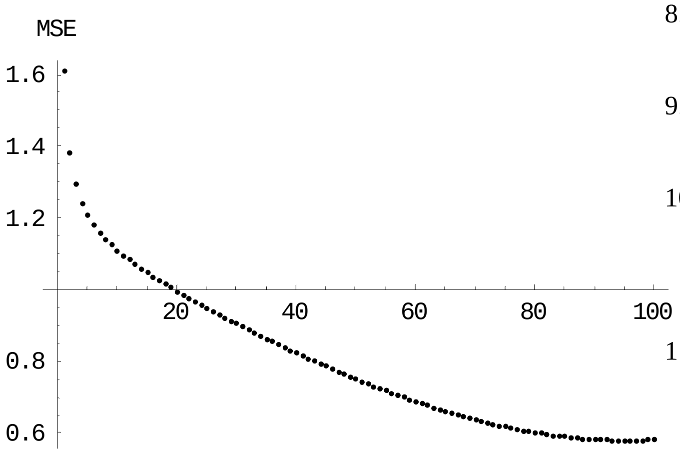

Fig. 9. $f_8(t) = 1 + 2\sin(2\pi t)$, $q_8(t) = 1 - e^{-0.03(t-0.32)}$

As it is clear in Fig. 9, the error range is between 1.6 and 0.6 for first 100 samples of data.

V. CONCLUSION

In this model we conclude that we could control variation of MSE error affected by increase in frequency, with parameters of deterministic function. These parameters set in such a way that error decreased with respect to number of samples value. it means that we could make better approximation with new sample arrival and decrease Mean Square Error.

REFERENCES

1. A. Papoulis, Signal Analysis. New York: McGraw-Hill, 1977.
2. L. R. Rabiner and R. W. Schafer, "Recursive and nonrecursive realizations of digital filters designed by frequency sampling techniques," IEEE 7'rans. Audio Electroacoust. vol. AU-19, pp. 200-207, Sept. 1971.
3. A. Papoulis, Probability, Random Variables and Stochastic Process. New York: McGraw-Hill, 2002, pp. 485-487.
4. A. Papoulis, "The Two-to-One Rule in Data Smoothing", IEEE transactions on information theory, vol. n-23, NO. 5, September 1977
5. A. M. Prokhorenkov, "Methods for Identification of Random Process Characteristics in Information Processing Systems," IEEE transactions on instrumentation and measurement, vol. 51, NO. 3, June 2002
6. J. S. Bendat and A. G. Piersol, Random Data. New York: Wiley, 1986, p. 545.
7. BYY data smoothing based learning on a small size of samples, Lei Xu, Neural Networks, 1999. IJCNN '99. International Joint Conference on Volume 1, 10-16 July 1999 Page(s):546 - 551 vol.1
8. Data smoothing algorithm, Van Kley, E.R., Position Location and Navigation Symposium, 1994., IEEE, 11-15 April 1994 Page(s):323 – 328
9. A matched filter data smoothing algorithm, Jacobson, L.A., Nuclear Science, IEEE Transactions on Volume 36, Issue 1, Part 1, Feb. 1989 Page(s):1227 – 1231
10. Data smoothing by cubic spline filters, Gung Feng, Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] Volume 46, Issue 10, Oct. 1998 Page(s):2790 – 2796
11. A note on "Data smoothing by cubic spline filters", Gang Feng, Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on], Volume 47, Issue 9, Sept. 1999 Page(s):2588 – 2589