Robust Regularized Learning Using Distributed Approximating Functional Networks

Zhuoer Shi, D. S. Zhang, and Donald J. Kouri
Department of Physics and Chemistry
University of Houston, Houston, TX 77204
zshi@bayou.uh.edu, dszhang@bayou.uh.edu, kouri@uh.edu

David K. Hoffman
Department of Chemistry and Ames Laboratory
Iowa State University, Ames, IA 50011
hoffman@qdynamic.ameslab.gov

Abstract

We present a novel polynomial functional neural networks using Distributed Approximating Functional (DAF) wavelets (infinitely smooth filters in both time and frequency regimes), for signal estimation and surface fitting. The remarkable advantage of these polynomial nets is that the functional space smoothness is identical to the state space smoothness (consisting of the weighting vectors). The constrained cost energy function using optimal regularization programming endows the networks with a natural time-varying filtering feature. Theoretical analysis and an application show that the approach is extremely stable and efficient for signal processing and curve/surface fitting.

I. Introduction

For a robust estimation, the output not only approaches the signal value, but also suppresses the noise distortion. Simultaneously, the state space should be smooth to ensure stability. Based on these facts, one finds that the least mean square (LMS) error

$$E_A = \int [Y(X) - \hat{Y}(W,X)]^2 dX \ , \tag{1}$$

the regularization constraints of order r

$$E_R = \int [\frac{\partial^r \hat{Y}(W,X)}{\partial X^r}]^2 dX \ , \tag{2}$$

and the condition of the system

$$E_W = \frac{\|W\|}{\sum_i |g(x_i)|^2} \tag{3}$$

are the three dominant factors that must be accounted for in designing a robust estimation system.

Distributed Approximating Functionals (DAFs), which can be constructed as a window modulated interpolating shell, were introduced as a powerful grid method for numerically solving PDEs with extremely high accuracy and computational efficiency [3,9,10]. DAF neural networks (DAFNNs) are presented for biomedical signal processing. DAFNNs possess several advantages: (1) the DAF wavelet is infinitely smooth in both time and frequency domains. (2) For essentially arbitrary order of the Hermite polynomial, the DAFs possess an approximately constant shape while other approximating functions become more oscillatory as the regularization order increases. (3) The translation invariance of the DAF approximation ensures feature preservation in state space. (4) Complicated mathematical operations, such as differentiation or integration, can be carried out conveniently using the DAF approximation. (5) The identical smoothness of the DAF signal space and state space implies the inherent robustness of estimation.

II. DAF Wavelet Networks

Signal filtering can be regarded as a kind of approximation with noise suppression. According to DAF theory, a signal approximation in DAF space is expressed as

$$\hat{g}(x) = \sum_i g(x_i) d_{DAF}(x - x_i) \tag{4}$$

where the $d_{DAF}(x)$ is a generalized symmetric Delta functional. We choose it as a Gauss modulated interpolating shell, or the so-call distributed approximating functional (DAF) wavelet. The Hermite-type DAF wavelet is given by [10].

$$d_M(x \mid s) = \frac{1}{s} \exp(\frac{-x^2}{2s^2}) \sum_{n=0}^{M/2} (\frac{-1}{4})^n \frac{1}{\sqrt{2p}\, n!} H_{2n}(\frac{x}{\sqrt{2}s}) \tag{5}$$

The function H_{2n} is the Hermite polynomial of even order $2n$. The Hermite polynomial H is generated by the usual recursion

$$H_n(x) = \begin{cases} 1, & n=0 \\ 2x, & n=1 \\ 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x), & n>1 \end{cases} \quad (6)$$

The predominant advantage of the Hermite polynomial approximation is high-order derivative preservation. The qualitative behavior of one Hermite DAF is shown in Fig.1.

The DAF wavelet neural nets possess the alternative feature of the commonly used DAF approximation as

$$\hat{g}(x) = \sum_i w(i) d_{DAF}(x - x_i) \quad (7)$$

The weights $w(i)$ of the nets determine the superposition $\hat{g}(x)$ to the original signal $g(x) \in L^2(R)$. It is easy to show that the weights are closely related to the DAF sampling $g(x_i)$. The irregular finite discrete time samplers of the original signal are selected for network learning. If the observed signal is limited to an interval I containing a total of N discrete samples, $I = \{0, 1, \ldots, N\text{-}1\}$, the square error of the signal is digitized as

$$E_A = \sum_{n=0}^{N-1} [g(n) - \hat{g}(n)]^2 \quad (8)$$

This cost function is commonly used for training and is referred to as the minimum mean square error (MMSE) rule.

However, if the observed signal is noise corrupted, the network under MMSE training causes unstable reconstruction because the MMSE recovers the noise as well as the signal. The signal-noise-ratio (SNR) cannot be improved much. Even for a noise-free signal, MMSE may lead to Gibbs-like undulations in the signal, which is harmful for calculating accurate derivatives. Therefore the network structure should be modified to deal with the particular situation. In this paper, we present a regularization design of the cost function for network training. It generates edge-preserved filters and reduces distortion. An additional smooth derivative term, E_r, is introduced to modify the original cost function as

$$\begin{aligned} E &= E_A + lE_r \\ &= \sum_k [g(k) - \hat{g}(k)]^2 + l \int_R [\frac{\partial^r \hat{g}(x)}{\partial x^r}]^2 dx \end{aligned} \quad (9)$$

The Lagrange factor l introduces a compromise between the orders of approximation and smoothness. Generally, the derivative order $r \geq 2$ is used to evaluate the smoothness of the signals. Using the properties of the Hermite DAFs, we find that the derivative term of the regularized cost function E_r can be expressed in a comparatively simple convolution form,

$$\frac{\partial^r \hat{g}(x)}{\partial x^r} = \sum_i w(i) d_M^{(r)}(x - x_i \mid s) \quad (10)$$

where $d_M^{(r)}(x - x_i \mid s)$ is termed a "differentiating DAF" and is given by

$$\begin{aligned} d_M^{(r)}(x - x_i \mid s) &= \frac{(-1)^r}{2^{r/2} s^{r+1}} \exp(\frac{-(x-x_i)^2}{2s^2}) \\ &\times \sum_{n=0}^{M/2} (\frac{-1}{4})^n \frac{1}{\sqrt{2p}\, n!} H_{2n+r}(\frac{x-x_i}{\sqrt{2}s}) \end{aligned} \quad (11)$$

It is the rth derivative of $d_M(x - x_i \mid s)$. Because the DAF is smooth, any order derivative can be obtained. Another constraint in state space is taken to increase the stability of system as

$$E_W = \frac{\sum_i |w(i)|^2}{\sum_i |g(x_i)|^2} \quad (13)$$

Thus the complete cost function utilized for DAF wavelet net training is given by

$$\begin{aligned} E &= E_A + lE_r + hE_W \\ &= \sum_k [g(k) - \hat{g}(k)]^2 + l \int_R [\frac{\partial^r \hat{g}(x)}{\partial x^r}]^2 dx + h \frac{\sum_i |w(i)|^2}{\sum_i |g(x_i)|} \end{aligned} \quad (14)$$

The combined constraints above enable the DAF networks a natural time-varying and nonlinear process, that the DAF wavelet nets are adaptive to the local behaviors of the signal, which are characterized by its smoothness and stability.

III. Simulations

Automatic diagnosis of electrocardiogram (ECG or EKG) signals is based on the detection of abnormalities. ECG signal processing is a crucial step for obtaining a noise-free signal and for improving diagnostic accuracy. A typical raw ECG signal is given in Fig. 2. The letters P, Q, R, S, T and U label the medically interesting features. For example, normal P waves rate 60-100 bpm with <10% variations. Their heights are <2.5mm and widths <0.11s in lead II. A normal PR interval ranges from 0.12 to 0.20s (3-5 small squares). A normal QRS complex has duration of <0.12s (3 small squares) [12].

An important task of ECG signal filtering is to preserve the true magnitudes of the waves, protect the true intervals (starting and ending points) and segments, and suppress distortions induced by noise. The most common noise in an ECG signal is AC interference (about 50Hz-60Hz in the frequency regime). Traditional filtering methods (low-pass, and band-elimination filters, etc.) encounter difficulties in dealing with the AC noise because the signal and the noise overlap the same band. As a consequence, experienced doctors are required to carry out time-consuming manual diagnoses.

Another application is for electromyography (EMG) filtering. Surface EMG has been used to evaluate muscle activation patterns in-patients with gait disorders since the mid 1900s. In experimental as well as routine recording of muscle action potentials, signal cross-talk from various sources cannot always be avoided [2]. In particular EMG-investigations within the areas of physical science, orthopedics or ergonomics, where the collection of data has to be carried out under field conditions, the measured signals are often incorrect due to movement of the subject. In particular DC off-set-voltages, movement of electrodes and cables, 50 Hz interference and electrostatic interference should all be considered. But even with the utmost case, movement artifacts, particularly in studies of movement, cannot be completely avoided. Thus for a number of quantitative signal processing procedures, an elimination of interference has to be carried out.

Experimental electrocardiogram (ECG) and electromyography (EMG) signals are employed for testing the new technique. Signal processing is crucial for obtaining a noise-free signal and for improving diagnostic accuracy. The raw ECG is given in Fig. 3(a). Note that it has typical thorn-like electromagnetic interference. The original measured EMG has similar features (Fig.4 (a)). Fig. 3(b) and Fig. 4(b) are the results of low-pass filter smoothing of the two signals. The magnitudes of the waves are significantly reduced and the small waves almost disappear completely. Such a low-pass filtering result can cause significant diagnostic errors. Fig. 3(c) and Fig.4 (c) show the results obtained using our DAF wavelet neural nets. Obviously, our method provides better feature-preserving filtering.

III. Conclusion

Regularized DAF wavelet neural networks are proposed for non-stationary biomedical signal processing. The DAF approximation shells possess infinite smoothness in both physical and frequency domains, which enable the high-resolution time-varying analysis of the signal. The optimal signal filtering solution is obtained using a combination of several different contributions to a “cost function”. Measured ECG and EMG signals are employed for testing the new technique. The simulations show that our method is both efficient and robust for time-varying filtering

References

[1] D. L. Donoho, “De-noising by soft-threshold,” IEEE Trans. Information Theory, vol. 41, no.3, pp. 613~627, 1995.

[2] T. H. J. M. Gootzen, “Muscle Fibre and Motor Unit Action Potentials,” Ph.D. Thesis, Univ. of Nijmegen, 1990.

[3] D. K. Hoffman, G. W. Wei, D. S. Zhang, D. J. Kouri, “Shannon-Gabor wavelet distributed approximating functional,” Chemical Pyiscs Letters, Vol. 287, pp.119-124, 1998.

[4] Z. Shi, “Nonlinear processing of ECG signal,” B.Sc. Thesis, Xidian Univ., 1991.

[5] Z. Shi, Z. Bao, L. C. Jiao, “Nonlinear ECG filtering by group normalized wavelet packets”, IEEE International Symposium on Information Theory, Ulm, Germany, 1997

[6] Zhuoer Shi, L. C. Jiao, Z. Bao, “Regularized spline networks,” IJCNN’95, Beijing, China. (also J. Xidian Univ., Vol.22, No.5, pp.78~86, 1995.)

[7] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: part I—theory," IEEE Trans. SP, Vol. 41, No.2, pp.821~833, February 1993

[8] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: part II—Efficient design and applications," IEEE Trans. SP, Vol. 41, No.2, pp.834~848, February 1993

[9] G. W. Wei, D. S. Zhang, D. J. Kouri, and D. K. Hoffman, “Lagrange distributed approximating Functionals,” Phys. Rev. Lett., Vol. 79, No.5, pp. 775~779, 1997

[10] D. S. Zhang, G. W. Wei, D. J. Kouri, and D.K. Hoffman, “Numerical method for the nonlinear Fokker-Planck equation,” Phys. Rev. E, Vol.56, No.1, pp.1197-1206, 1997.

[11] V. R. Zschorlich, “Digital filtering of EMG-signals,” Electromyogr. Clin. Neurophysiol., Vol. 29, pp. 81-86, 1989.

[12] http://homepages.enterprise.net/djenkins/.

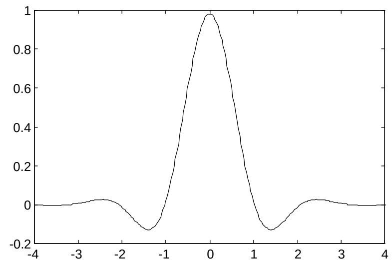


Fig.1. Hermite DAF (M=8, s=1).

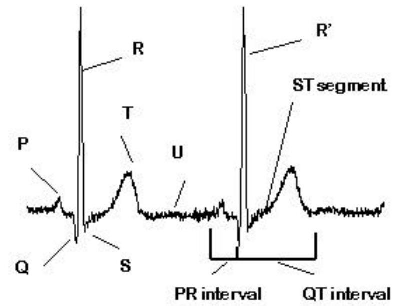
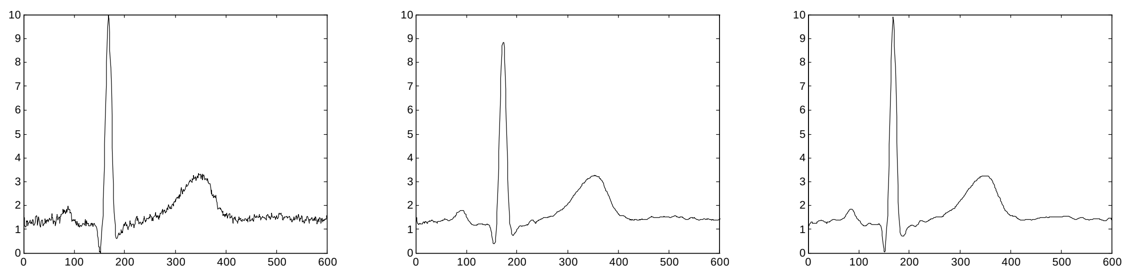
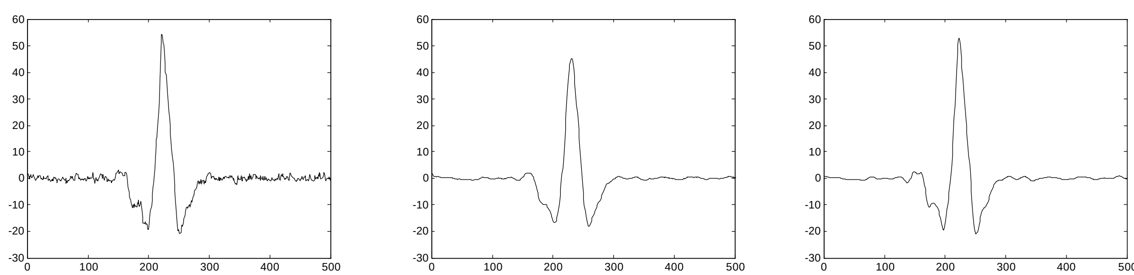


Fig.2 ECG Criterion Characteristics for Diagnosis



(a) (b) (c)

Fig.3 ECG filtering (a) Original ECG (b) Low pass filtering (c) DAF wavelet net filtering



(a) (b) (c)

Fig.4 EMG filtering (a) Original EMG (b) Low pass filtering (c) DAF wavelet net filtering