Case Studies on Discrete Wavelet Denoising via Kernel Based Nonlinear Component Analysis

Zhengmao Ye, Hang Yin
College of Engineering, Southern University
Baton Rouge, LA70813, USA

Yongmao Ye
Liaoning Radio and Television Station
ShenYang, Liaoning, 110003, China

Abstract — Complex real world problems are essentially nonlinear. Linear models are relatively simple but inaccurate to describe the nonlinear aspects of dynamic system behaviors. Denoising techniques have been broadly applied to numerous applications in the spatial domain, frequency domain, and time domain. To increase the adaptability of denoising techniques to signal processing of arbitrary nonlinear systems, kernel based nonlinear component analysis is proposed to enhance wavelet denoising. In the multilevel wavelet decomposition, the low frequency approximations and high frequency details are produced at each level. Discrete wavelet transform (DWT) will help to decompose low frequency approximations exclusively at all the succeeding levels, while wavelet packet transform decomposes both approximations and high frequency details at each level. DWT is selected for wavelet denoising in this study, where details at each level and the approximation at specified level are all subject to simplification using nonlinear component analysis. Case studies of typical nonlinear denoising problems in various domains are conducted. The results manifest strong feasibility and adaptability across diverse denoising problems of nonlinear systems.

1. Introduction

Signal and image processing is an important branch of science and engineering. It mainly focuses on measuring, recognizing, analyzing, filtering, enhancing, compressing and synthesizing the signals and images. It deals with all types of linear and nonlinear systems. To extract the intrinsic signals and images from measurements and observations, denoising has always been adopted as the initial stage before pattern classification, modeling, control and optimization could be conducted. Both wavelet transform and component analysis techniques are broadly used for signal and image denoising. In addition artificial intelligence schemes such as fuzzy logic and neural networks are also frequently involved in noise suppression. To enhance the quality of denoising for highly nonlinear systems, integration of 2 popular schemes is proposed to achieve efficient denoising and to be adaptable to more general data types without excessive computation complexity needed [1-4].

Discrete wavelet transform (DWT) conducts the wavelet transform via a discrete set of wavelet scales and translations on a basis of defined rules. It actually decomposes the signal into an orthogonal set of wavelets. DWT decomposes a signal to equal length of wavelet coefficient spectra to the number of the signal data or possible shorter length, so no information redundancy occurs and thus it does have many applications. A watermarking algorithm uses wavelet multi-resolution structure to obtain frequency components, both wavelet and spatial transformation have been conducted [5]. (2D) DWT has been presented for digital compression and reconstruction of fingerprint images to manifest the unique biometric patterns based on reoccurring data. Its feasibility has been testified via information metric analysis such as discrete entropy and discrete energy [6]. DWT is also applied to the fractal-based denoising in the wavelet domain. Prediction of the fractal code for a noiseless image can be made in terms of noisy observations. The fractal-based denoising scheme results in better estimations of source images [7]. For the blurry aerial images, denoising is indeed necessary for better target detection and pattern recognition. Both (2D) DWT and wavelet packets could be used with satisfactory results together with soft thresholding. Metrics such as mutual information have also been applied to make comparisons between two techniques [8]. Being a noval scheme, dual watermarking based on both DWT and chaotic schemes is proposed to protect information security against various malicious attacks. It is shown that information integrity has been well preserved in both the frequency domain and spatial domain via quantitative analysis [9].

On the other hand, component analysis is another good means to process signals and images. It has the potential to be applied to robotic controlled remote surgery. The real-time biomedical sample analysis with a Raman spectrometer mounted on the end-effector of the medical robot has been designed, which is used to differentiate among diverse animal samples of liver, lung, kidney and glands. Principal component analysis (PCA) instead can be used for in-vivo sample differentiation. Based on Raman frequency shift information between incident light and scattered light, the resonant frequency of the biomedical sample could be uniquely located. When PCA is combined with other approach such as fuzzy filtering for denoising and clustering analysis, a systematic approach is provided for biomedical sample characterization [10-11]. The extension of PC allows for extracting a few relevant features from high-dimension fuzzy data. The linear neural networks are exploited for information compression without the procedure of explicit matrix diagonalization. It is demonstrated that this scheme provides concise representations of lengthy fuzzy data from experiment data [12]. Component analysis can be applied to biometric identification as well. Sensing information can be digitized into data matrices. It can be indicated by plaintexts consisting of inherent signatures. Both PCA and Independent Component Analysis (ICA) could be introduced for decision

making [13]. For moving object detection, it also contains unique signatures. The underlying factors can be revealed via ICA based synthesis. This method has been evaluated using metrics well defined in the information theory [14]. The component analysis can also be expanded to other relevant fields. In the complex flocculation mechanism study, kernel based nonlinear component analysis (NCA) has been employed, which offers more flexibility to solve large dimensional nonlinear multivariate problems. It can also be easily extended to other high-dimensional nonlinear cases. The impact of salinity on the synthetic floc strength has been examined via NCA. It is observed that salinity turns out to be a dominating factor for mechanical behaviors of flocs [15].

Seeing the crucial roles of DWT and NCA on a wide variety of denoising problems, no matter on linear or nonlinear systems. Integration of DWT and NCA has been proposed for more general nonlinear denoising issues with higher adaptability across several domains in this study.

2. 1D Discrete Wavelet Transform (DWT)

Discrete Wavelet Transform (DWT) is introduced for denoising and in general it is applied at multiple levels. The decomposition at each level produces the scaling function and wavelet function in 1D DWT, together with both the corresponding approximation and detail coefficients. Each function is reduced to the half size after decomposition. For the multiple-level DWT, the approximation component is subject to further decomposition at each level while all the detail components are kept for further analysis. Information loss between two adjacent levels of approximations gives rise to the detail component. The simple 1D Haar wavelet transform is formulated as below. The expansion function sets consist of integer translations and binary scaling of the real squared scaling function $\varphi(x)$ that is integrable. Each component of the set $\{\varphi_{n,k}(x)\}$ is expressed as (1), for all pairs of integers n and k in the Hilbert space $L^2(R)$. The scaling function $\varphi(x)$ in fact changes with n.

$$\varphi_{n,k}(x) = 2^{n/2}\,\varphi(2^n x - k) \tag{1}$$

The wavelet functions $\psi(x)$ is also defined with integer translations and binary scaling. It spans difference between any two adjacent scaling subspaces. Individual component of wavelet function set $\{\psi_{i,k}(x)\}$ is expressed as (2).

$$\psi_{n,k}(x) = 2^{n/2}\,\psi(2^n x - k) \tag{2}$$

for all pairs of integers n and k in the Hilbert space $L^2(R)$. The wavelet function $\psi(x)$ also changes with n.

The scaling function $\varphi(x)$ and wavelet function $\psi(x)$ can be self-defined. For instance, the scaling function $\varphi(x)$ has been originally defined as a unit height and unit width function. Meanwhile when the scale n is fixed, it follows that (3) and (4) hold.

$$\varphi(x) = \varphi(2x) + \varphi(2x-1) \tag{3}$$
$$\psi(x) = \varphi(2x) - \varphi(2x-1) \tag{4}$$

The wavelet series expansion of f(x) is formulated based on both the scaling function $\varphi(x)$ and wavelet function $\psi(x)$.

$$f(x) = \sum_k \alpha_{n0}(k)\,\varphi_{n,k}(x) + \sum_{n=n_0}^{\infty}\sum_k \beta_n(k)\,\psi_{n,k}(x) \tag{5}$$

where n_0 represents the specified starting scale. $\alpha_{n0}(k)$ is the approximation coefficient and $\beta_{n0}(k)$ is the detail coefficient, which are defined as the inner products of the function f(x) itself with the 1D scaling function and 1D wavelet function, respectively. It will be applied in this study.

3. Nonlinear Component Analysis

For a number of short discrete-time sequences containing noises, it is straightforward to apply component analysis. For a finite sufficient long discrete-time sequences (size: LN; N>>L) corrupted by noises, interferences and disturbances instead, it is also reasonable to be viewed as multiple (L) observations of the short sequences of equal size (N), so a set of column vectors can be produced. In this way, to make it simple, component analysis could be actually applied to the typical sufficient long discrete-time sequence without a set of long sequences being needed.

For DWT denoising at the level M in context, wavelet decomposition is performed from the 1st to Mth level on the scaling functions. Then the approximation coefficient at the level M and all the wavelet coefficients from the 1st to Mth level can be collected. Meanwhile it has been applied to all L observations simultaneously. Now a 3D array of N×L×(M+1) is constructed with (M+1) numbers of N×L matrices. Each one is used to estimate the covariance matrix based on the proposed nonlinear component analysis.

Compared with the principal component analysis, kernel based nonlinear component analysis is more adaptable to the complex problems with highly nonlinear behaviors. The nonlinear kernel can be applied to approximate all the covariance matrices instead when the data is not linearly separable at all. Typical nonlinear kernels include Gaussian kernel, Cauchy kernel, Laplace exponential kernel, and so on. Both exponential and non-exponential kernel functions are frequently used (Fig. 1).
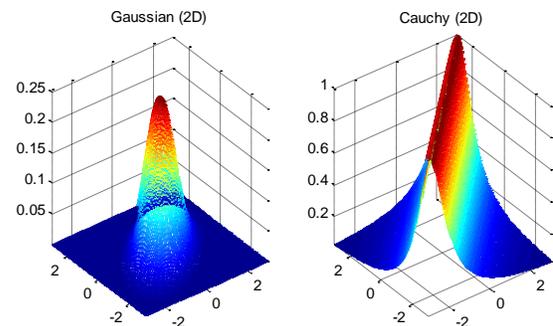

Fig. 1. Typical Exponential and Non-Exponential Kernel Functions

The popular functions of Gaussian kernel, Cauchy kernel, and Laplace exponential kernel can be formulated as (6).

$$
\begin{aligned}
k_G(x_1, x_2) &= e^{-\frac{\|x_1 - x_2\|^2}{2\sigma^2}} \\[2mm]
k_C(x_1, x_2) &= \frac{1}{1 + \frac{1}{\sigma^2}\|x_1 - x_2\|^2} \\[2mm]
k_L(x_1, x_2) &= e^{-\frac{\|x_1 - x_2\|}{\sigma}}
\end{aligned}
\tag{6}
$$

where sigma σ in (6) is the square root of the variance. In Gaussian kernel k_G, σ can be adjusted so that a tradeoff between nonlinearity and sensitivity can be made. It is an exponential kernel that measures the closeness. The Cauchy kernel instead produces broad influence and sensitivity over the high dimensional space. The Laplace kernel is often used to reduce sensitivity against variations of the parameter σ. Gaussian kernel is commonly adopted, whose results via numerical simulations will be shown in following sessions.

For the nonlinear kernel approach with the centered data ($\sum_{i=1}^{N} \Phi(x_i) = 0$), i=1, 2, ..., N, the diagonalization is implemented on the resulting covariance matrix COV_{NCA} in (7), (M+1) times in total. Φ defines a nonlinear function, such as the three kernel functions defined above.

$$COV_{NCA} = \frac{1}{N} \sum_{i=1}^{N} \Phi(x_i) \Phi(x_i)^T \qquad (7)$$

If the centered data is unavailable initially, Standard Normal Variate (SNV) can be easily employed to normalize the data to zero mean with the standard deviation scaled to one. Now (7) results in the problem of eigenvalues and eigenvectors.

$$\lambda V = COV_{NCA} V = \frac{1}{N} \sum_{i=1}^{N} [\Phi(x_i) V] \Phi(x_i) \qquad (8)$$

where λ acts as the eigenvalue and V is the corresponding eigenvector of COV_{NCA}. The eigenvector corresponding to the largest eigenvalue is known as the most significant principal component or first principal component (PC1), showing the direction of maximal variability. $\Phi(x_i)$ is the centered data and COV_{NCA} is a positive definite matrix. An alternative approach to compute the nonlinear principal component could be Singular Value Decomposition (SVD). The results from two approaches will be identical to each other. The nonlinear kernel function is easily defined as an inner product in the feature space being denoted as (9). The eigen-space is generated in the span of $\Phi(x_i)$, $\Phi(x_2)$, $\Phi(x_3)$, ..., and $\Phi(x_N)$.

$$k(x_i, x_j) = \left\langle \Phi(x_i), \Phi(x_j) \right\rangle \qquad (9)$$

4. Numerical Simulations

With respect to the approximation coefficient at the level M and all the detail coefficients from the 1st to M^{th} level, NCA will be applied to each of L corrupted finite sequences of size N for denoising, It leads to another covariance matrix of equal size N×L. In particular, all principal components for the approximation coefficient and M detail coefficients can be computed using NCA, while those eigenvalues less than one would be directly ignored. Then principal components from low frequency smooth approximation will be kept for reconstruction via inverse transform. Principal components from the coarse detail instead essentially represent insignificant high frequency components and noises, which are subject to suppressing. Via integration of DWT and NCA, all qualified principal components of

details at levels 1 through M are computed and kept. Hard thresholding will be applied to these principal components of the detail coefficients. Specifically two simple schemes are implemented. Firstly, hard thresholding is performed on most significant principal components (PC1) at levels 1 to M on details exclusively. Secondly, hard thresholding is applied to all those principal components greater than one at levels 1 to M on details.

Now all detail coefficients at levels 1 though M have been thresholded for noise reduction purposes. For principal components of approximation coefficient at the level M instead, leave it alone for reconstruction via inverse transform as long as the corresponding eigenvalues are greater than one. The next step turns out to be the wavelet reconstruction starting from the level M back to level 1. The centered data after denoising is reached. After conducting reverse process of standard normal variate and then combining the L short sequences, a similar finite sufficient long sequence to the original signal after denoising has been formulated. Without loss of generality, the proposed scheme will be tested multiple times in terms of noise reduction for diverse nonlinear issues in the spatial domain, frequency domain and time domain, respectively, so as to show its feasibility and broad adaptability.

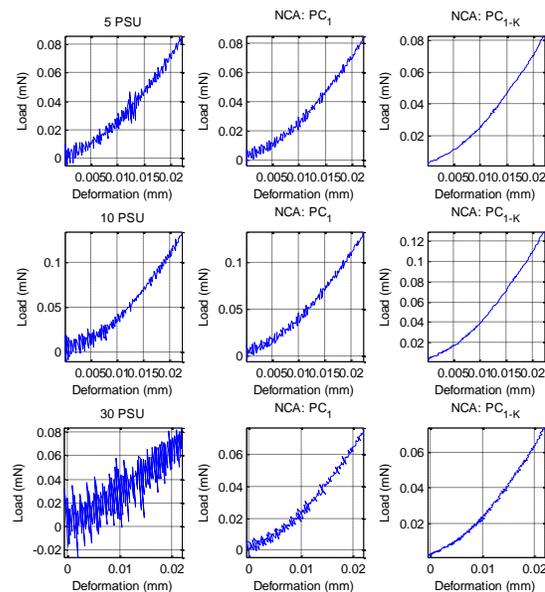# 5. Case Study 1: Spatial Domain Problem

Fig. 2. Denoising in Spatial Domain Problem

To analyze complex mechanisms involved in flocculation and coagulation processes with highly nonlinear behaviors, the impact of salinity on the synthetic floc strength has been examined. The scale of salinity in the synthetic flocs can be adjusted to different scales such as 5 PSU, 10 PSU and 30 PSU, respectively. Because the flocculation processes of suspended clay particles are performed in the stirring bath, it is not guaranteed at all that the data collected will be accurate. Via data acquisition, however, nonlinear dynamic relationships between load and deformation could be

captured for each case. It provides 3 sets (5 PSU, 10 PSU and 30 PSU) of source data in the spatial domain. Based on the proposed scheme, DWT and NCA are applied for denosing with satisfied results. For each set of data in Fig. 2, comparisons are made among source data and two individual types of the denoised signals. The wavelet decomposition is performed up to the Mth level. Those PCs whose eigenvalues are less than one should be omitted for both approximation and detail coefficients. In addition detail coefficients are also subject to hard thresholdng on either PC1 or all PCs across all M levels prior to reconstruction. From the simulation results, denoising is well implemented on both cases. But hard thresholdng on all PCs will produce better denoising results than that on PC1 exclusively.

6. Case Study 2: Frequency Domain Problem

The wavelets also have the local property in the time domain and frequency domain from data points of view. Raman spectroscopy can be applied to biomedical analysis in a non-destructive manner. The frequency domain Raman spectra, however, are relatively weak signals whose features are easily corrupted by noises. Thus nonlinear denoising should be introduced beforehand. In this study, the source Raman spectra from samples of Mice glands, liver and lung are collected for signal denoising. Once again, for each set of Raman spectra in Fig. 3, comparisons are made among source data and two individual types of denoised data. From the simulation results, denoising is well implemented on both cases of nonlinear denoising via integration of DWT and NCA. However, hard thresholdng on all relevant PCs of detail coefficients will produce better denoising results than that on PC1 of the detail coefficients exclusively. Similar results are reached as those cases in the spatial domain.
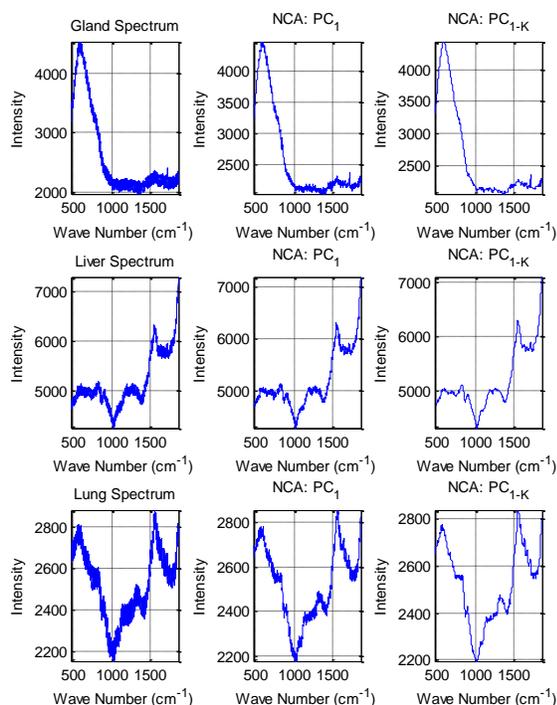

Fig. 3. Denoising in Frequency Domain Problem

7. Case Study 3: Time Domain Problem

The same wavelet analysis approach can also be expanded to denoising in the time domain problems. Modeling and prediction in the stock market are both tough problems in finance and no existing theory could solve all the tough problems. The fluctuation of data and nonlinear behaviors occur all the time. It is also a discrete time domain problem varying daily or monthly. Denoising could be employed as well in order to simplify the model for better prediction. In Fig. 3, the data draft of Dow Jones Industrial Average Stock Market Index over 100 years is collected, stating from 1043.67 in January 1921 to 30606.48 in December 2020. Similar to both the spatial domain and frequency domain examples, integration of DWT and NCA schemes proposed has been demonstrated to be a powerful denoising tool. Meanwhile, when the resulting PCs for approximation stay the same, better denoising results can be obtained if hard thresholdng is performed on all relevant PCs of details than on the most significant PC1 of details themselves. It has shown clearly that the proposed integration scheme of DWT and NCA could be applied to the general types of nonlinear denoising problems.
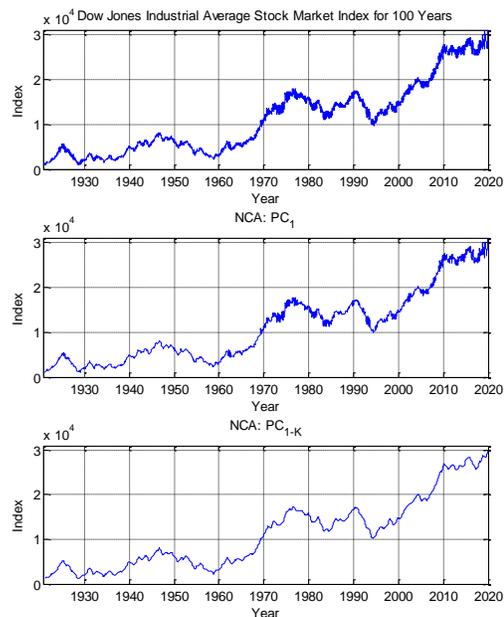

Fig. 4. Denoising in Time Domain Problem

Conclusions

Integration of DWT and NCA has been presented for discrete time signal denoising in this research. The nonlinear nature of the noise reduction approach allows for potential applications in numerous different problems across diverse disciplines. Without loss of generality, case studies have been conducted accordingly, covering mechanical nonlinear behavior analysis of the salinity impact on the flocculation process in the spatial domain, Raman spectral analysis for sample differentiation in the frequency domain, and Dow Jones Industrial Average Stock Market Index analysis and potential prediction in the time domain. Outcomes of these case studies have indicated the effectiveness of the proposed

nonlinear denoising approach using both DWT and NCA. It has the potential to be applied to diverse cases of noise reduction problems, even under highly nonlinear behaviors.

References

[1] R. Schilling, S. Harris, "Fundamental of Digital Signal Processing Using Matlab", Cengage Learning, 2005

[2] R. Gonzalez and R. Woods, "Digital Image Processing," 2nd Edition, Prentice Hall, 2002

[3] S. Hykin, "Neural Networks –A Comprehensive Foundation", 2nd Edition, 1999, Prentice Hall

[4] R. Duda, P. Hart and D. Stork, "Pattern Classification," 2nd Edition, John Wiley & Sons, 2000

[5] M. Tsai, K. Yu, Y. Chen, "Joint Wavelet and Spatial Transformation for Digital Watermarking", IEEE Transactions on Consumer Electronics, v 46, n 1, p 237, FEB 2000

[6] Z. Ye, H. Mohamadian and Y. Ye, "Information Measures for Biometric Identification via 2D Discrete Wavelet Transform", Proceedings of the 2007 IEEE International Conference on Automation Science and Engineering (CASE 2007), pp. 835-840, Sept. 22-25, 2007, Scottsdale, Arizona, USA

[7] M. Ghazel, G. Freeman, and E. Vrscay, "Fractal-Wavelet Image Denoising Revisited", IEEE Transactions on Image Processing, Vol. 15, No. 9, September, 2006

[8] Z. Ye, H. Mohamadian and Y. Ye, "Quantitative Effects of Discrete Wavelet Transforms and Wavelet Packets on Aerial Digital Image Denoising", Proceedings of the 2009 IEEE International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE 2009), pp. 416-420, November 10-13, 2009, Toluca, Mexico

[9] Z. Ye, H. Yin, and Y. Ye, "Quantitative Comparison of Discrete Wavelet Chaotic Watermarking and Dual Chaotic Watermarking", Proceedings of IEEE International Conference on Computing, Communication and Automation, December 14-15, 2018, Greater Noida, India

[10] Z. Ye and G. Auner, "Principal Component Analysis for Biomedical Sample Identification", Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (SMC 2004), pp. 1348-1353, Oct. 10-13, 2004, Hague, Netherlands

[11] Z. Ye, "Artificial Intelligence Approach for Biomedical Sample Characterization Using Raman Spectroscopy", IEEE Transactions on Automation Science and Engineering, Vol. 2, No. 1, pp. 67-73, January, 2005

[12] T. Denoeux, M. Masson, "Principal Component Analysis of Fuzzy Data Using Autoassociative Neural Networks", IEEE Transactions on Fuzzy Systems, June 2004

[13] Z.. Ye, Y. Ye and H. Mohamadian, "Biometric Identification via PCA and ICA Based Pattern Recognition", Proceedings of the Sixth IEEE International Conference on Control and Automation, Guangzhou, China, pp. 1600-1604, May 30 - June 1, 2007

[14] Z. Ye, H. Mohamadian and Y. Ye, "Independent Component Analysis for Spatial Object Recognition with Applications of Information Theory Synthesis", Proceedings of the 2008 IEEE World Congress on Computational Intelligence (WCCI 2008), pp. 3640-3645, Hong Kong, June 1-6, 2008

[15] H. Yin, P. Carriere, H. Lawson, H. Mohamadian and Zhengmao Ye, "Characterization of Salinity Impact on Synthetic Floc Strength Via Nonlinear Component Analysis", Information Management and Big Data: Communications in Computer and Information Science Book Series, Vol. 1070, Ch. 2, pp. 8-20, Springer International Publishing, Springer, Switzerland, April 2020