1999년도 대한전기학회 하계학술대회 논문집 1999. 7.19-21

Bayesian 방법에 의한 잡음감소 방법에 관한 연구

이 문 직, 정 진 현
광운대학교 제어계측 공학과

Wavelet Denoising based on a Bayesian Approach

Moon Jik Lee , Chin Hyun Chung
Dept. of Control & Instrumentation Engineering. Kwangwoon Univ.

Abstract - The classical solution to the noise removal problem is the Wiener filter, which utilizes the second-order statistics of the Fourier decomposition. We discuss a Bayesian formalism which gives rise to a type of wavelet threshold estimation in non-parametric regression. A prior distribution is imposed on the wavelet coefficients of the unknown response function, designed to capture the sparseness of wavelet expansion common to most application. For the prior specified, the posterior median yields a thresholding procedure

1. Introduction

Consider the standard non-parametric regression problem:

$$y_i = g(t_i) + \varepsilon_i, \qquad i = 1, \ldots n. \qquad (1)$$

where $t_i = i/n$, ε_i are independent identically distributed normal variables with zero mean and variance σ^2, and we wish to recover the unknown function g from the noisy data without assuming any particular parametric form.

The function g is expanded in wavelet series in a way similar to the generalized Fourier series approach. The usual approach is to expand the noisy data in wavelet series, extract the 'significant' wavelet coefficients by thresholding, and then to invert the wavelet transform of the de-noised coefficients. Donoho and Johnstone[3] showed that such wavelet estimators with a properly chosen thresholding rule have various important optimality properties. The choice of thresholding rule, therefore, becomes a crucial step in the estimation procedure. In this paper we consider a thresholding within a Bayesian framework. In this Bayesian approach a prior distribution is imposed on the wavelet coefficients of the unknown response function. The prior model is designed to capture the sparseness of wavelet expansion common to most applications. Then, the function is estimated by applying some Bayes rule considered in the literature corresponds to an L^2-loss based on the wavelet coefficients. In this paper, instead of the L^2-loss, we proposed to use of a weighted combination of L^1-losses based on the wavelet coefficients. These losses correspond to L^1-losses based on the function and on its derivatives; *such losses are naturally measures* for spatially inhomogeneous functions. The corresponding Bayes rule is the posterior median and, for a certain prior, yields a thresholding procedure.

2. Wavelet estimators

2.1 Wavelet transform

Wavelet series are generated by dilation and translation of a function ψ, called the mother wavelet:

$\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$, $j, k \in Z$. *For suitable choices* of ϕ, the corresponding set of ψ_{jk} forms an orthonormal basis in $L^2(R)$. The wavelet series representation of a function $g \in L^2(R)$ is then :

$$g(t) = \sum_{j\in Z}\sum_{k\in Z} w_{jk}\psi_{jk}(t)$$

where the wavelet coefficients w_{jk} are given by :

$$w_{jk} = \int_R g(t)\psi_{jk}(t)\,dt.$$

In contrast to standard Fourier series, wavelets are local in both frequency/scale (via dilation) *and in time (via translation). This localization* allows parsimonious representation for a wide *set of different functions in wavelet series.* In technical term of corresponding regularity properties, one can generate an unconditional wavelet basis in a wide set of function spaces.

2.2 Wavelet Shrinkage

Given observed discrete data $Y = (y_1, \ldots, y_n)^T$ *from model* (1), we *may find* the vector $\hat{d}$ of its sample discrete wavelet coefficients by performing the discrete wavelet transform of Y:

$$\hat{Y} = WY.$$

where W is the DWT-matrix with (jk, i) entry. given by $\sqrt{(n)}\, W_{jk,i} \simeq \psi_{jk}(i/n) = 2^{j/2}\psi(2^j i/n - k)$.

The population discrete wavelet coefficients d_{jk} are defined as the DWT of the vector of function values $g(t_i)$, $i=1,\ldots,n$. These are related to the wavelet coefficients $w_{jk} = \int_R g(t)\psi_{jk}(t)\,dt$ by $d_{jk} \simeq \sqrt{n}\, w_{jk}$. The $\sqrt{n}$ factor essentially arises from the difference between continuous and discrete orthogonality conditions. Because of the orthogonality of W, the DWT of a white noise is also an array ε_{jk} of independent $N(0, \sigma^2)$ random variables and, hence, equally contaminates the population discrete wavelet coefficients d_{jk} :

$$\widehat{d_{jk}} = d_{jk} + \varepsilon_{jk}\ , \qquad j=0,\ldots,2^j-1. \qquad (2)$$

The next step is to extract those coefficients that really contain information about unknown function g and discard the others. This can be done by thresholding the sample discrete wavelet coefficients $\widehat{d_{jk}}$. The intuitive idea is that the true function g has a parsimonious wavelet expansion, i.e. only a few 'large' $\widehat{d_{jk}}$ essentially contain real information about g. If we decide which ones these are, we can estimate them and set all the others equal to zero. Donoho and Johnstone proposed the hard and soft thresholding rules:

$$T_{hard}(\widehat{d_{jk}}, \lambda) = \widehat{d_{jk}} I(|\widehat{d_{jk}}| > \lambda),$$
$$T_{soft}(\widehat{d_{jk}}, \lambda) = sign(\widehat{d_{jk}}) \max(0, |\widehat{d_{jk}}| - \lambda). \qquad (3)$$

where $\lambda \geq 0$ is a threshold parameter and I is the usual indicator function. In application, hard thresholding generally reproduces peak heights and discontinuous better, but at some cost in visual smoothness. By defining $d_{jk}^{new} = T_{hard}(\widehat{d_{jk}}, \lambda)$ or $d_{jk}^{new} = T_{hard}(\widehat{d_{jk}}, \lambda)$, one can then reconstruct $\hat{g}$ by the inverse DWT:

$$\hat{g} = W^T d^{new} \qquad (4)$$

The choice of λ is therefore crucial : if the threshold is too large then the wavelet shrinkage estimator will tend to overfit or underfit the data. Donoho and Johonstone proposed the universal threshold $\lambda_{DJ} = \sigma\sqrt{2\log(n)}$ called by them as visuShrink. Despite the simplicity of such a threshold, they showed that the resulting nonlinear wavelet estimator is spatially adaptive.

2.3 Bayesian Thresholding Rule

A large variety of different functions allow parsimonious representation in wavelet series where there are only a few non-negligible coefficients present in the expansion. We incorporate this characteristic feature of wavelet bases by replacing the following prior on the population discrete wavelet coefficients d_{jk}:

$$d_{jk} \sim \pi_j N(0, \tau^2_j) + (1-\pi_j)\delta(0),$$
$$j=0,\ldots,J-1; \qquad k=0,\ldots,2^j-1$$

(5)

where $0 \leq \pi_j \leq 1$, $\delta(0)$ is a point mass at zero, and d_{jk} are independent. The hyperparameters π_j and τ^2_j either zero with probability $1-\pi_j$ or with probability π_j is normally distributed with zero mean and variance τ^2_j. The probability π_j gives the proportion of non-zero wavelet coefficients at resolution level j while the variance τ^2_j is a measure of their magnitudes.

Subject to the prior, the posterior distribution $d_{jk} | \widehat{d_{jk}}$ is also a mixture of corresponding posterior normal distribution and $\delta(0)$. Hence, the posterior cumulative distribution function $F(d_{jk} | \widehat{d_{jk}})$, letting Φ be the standard normal cumulative distribution functions, is :

$$F(d_{jk} | \widehat{d_{jk}}) = \frac{1}{1+w_{jk}} \Phi\left(\frac{d_{jk} - \widehat{d_{jk}}\tau_j^2/(\sigma^2+\tau_j^2)}{\sigma\tau_j/\sqrt{\sigma^2+\tau_j^2}} \right) + \frac{w_{jk}}{1+w_{jk}} I(d_{jk} \geq 0),$$

(6)

where the posterior odds ratio for the component at zero is :

$$w_{jk} = \frac{1-\pi_j}{\pi_j} \frac{\sqrt{\tau_j^2}}{\sigma} \exp\left(-\frac{\tau_j^2 \widehat{d_{jk}}^2}{2\sigma^2(\tau_j^2+\sigma^2)} \right). \qquad (7)$$

The traditional Bayes rule corresponding to the L^2-loss considered in the literature is not a thresholding rule but a shrinkage. Instead, we proposed to use of any weighted combination of L^1-losses on the individual wavelet coefficients. Whichever weighted combination used, the corresponding Bayes rule will be obtained by taking the posterior median of each coefficients.

$$Med(d_{jk} | \widehat{d_{jk}}) = sign(\widehat{d_{jk}}) \max(0, \zeta_{jk}),$$

where:

$$\zeta_{jk} = \frac{\tau_j^2}{\sigma^2+\tau_j^2} |\widehat{d_{jk}}| - \frac{\tau_j\sigma}{\sqrt{\sigma^2+\tau_j^2}} \Phi^{-1}\left(\frac{1+\min(w_{jk}, 1)}{2} \right). (8)$$

The quantity ζ_{jk} is negative for all $\widehat{d_{jk}}$ in some implicity defined interval $[-\lambda_j, \lambda_j]$, and hence d_{jk} is zero whenever $|\widehat{d_{jk}}|$ falls below the threshold λ_j. The posterior median is therefore a level-dependant thresholding rule with thresholds λ_j. For large $\widehat{d_{jk}}$ the thresholding rule asymptotes to linear shrinkage by a factor of $\tau_j^2/(\sigma^2+\tau_j^2)$, since the second term in (8) becomes negligible as $|\widehat{d_{jk}}| \to \infty$.

2.3 Simulation

Figure 1 is the plot when Hyper parameters were choosen as τ^2 = 25, π=0.05, while σ was fixed at 1. We applied this algorithm to the "Einstein" image for three different levels of Gaussian white noise contamination. Figure 2 shows four images orignal, noisy, restored with weiner and bayes image. The Bayesian image appears to be both sharper and less noisy.

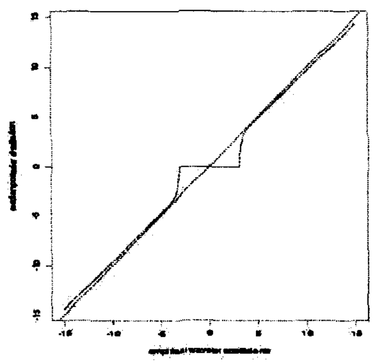

Figure 1. The median of posterior distribution (solid line) as function of the empirical wavelet coefficients.

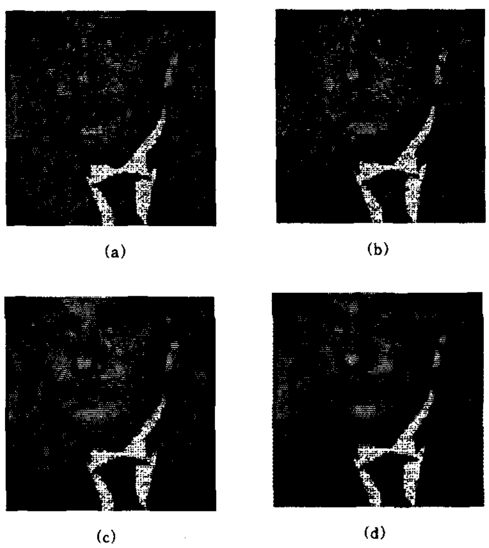

Figure 2. Noise reduction example. (a) Original image (cropped). (b) Image contaminated with additive Gaussian shite noise (SNR = 9.00dB). (c) Image restored using Winer filter (SNR = 11.88dB). (d) Image restored using Bayesian estimator(SNR = 14.83dB)

3. Conclusion

Removal of noise from images relies on difference in the statistical properties of noise and signal. The Bayesian estimator described above provides a natural extension for incorporating the higher-order statistical regularity present in the point statistics of subband representations. The estimator is based on two factors - a subband representation and a statistical model - both of which can be generalized. Theorically, one would like a direct link from the properties of the subband pdf to the quality of noise removal, which could then be used to optimize the choice of subband transform. In addition, the statistical model should account for joint statistics of wavelet coefficients, both within and between bands. Finally this type of statistical image model can be useful in other applications, such as image compression or texture synthesis.

[참 고 문 헌]

[1] aghuveer M. Rao and Ajit S. Bopardikar, Wavelet Transforms, Addison Wesley, 1988
[2] S. James press, Bayesian Statistics: Priciple, Models, and applications, Wiley, 1989
[3] Donoho,D.L & Johnstone, I.M., Ideal spatial adaption by wavelet shrinkage. Biometrika 81, 425-455, 1994
[4] Chipman, H.A., Kolaczyk, E.D. & McCulloch, R.E. Adaptive Bayesian wavelet shrinkage. Journal of the american Statistical Association 92,1997
[5] Johnstone, I.M. Minimax Bayes, asymptotic minimax and sparse wavelet priors. Statistical Decision Theory and Related Topics, V,Gupta, S.S and Berger, J.O. (Eds.), pp.303-326, New-York: Springer-Verlag, 1994.
[6] E.M Stein Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Prinstone, New Jersey, 1970.
[7] Meyer, Y. Wavelets and Operators, Cambridge: Cambridge University Press, 1992
[8] Eero P. Simoncelli, Edward H. Adelson, Noise removal via Bayesian wavelet coring, Proc 3rd IEEE Int'l Conf Image Processing, Vol. I, pp. 379-382., 1996.