Wavelets and variance reduction in non-parametric transfer function estimation.

S.G. Douma T.J. de Hoog P.M.J. Van den Hof
Signals, Systems and Control Group, Department of Applied Physics,
Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
s.g.douma@tn.tudelft.nl

Abstract

A variance reduction scheme is presented for non-parametric transfer function estimators based on the use of wavelets as an alternative to the traditional spectral windowing. The latter can be generalized into a variance reduction method based on *thresholding* (omitting or altering) the coefficients of an orthogonal series expansion of the estimator to be smoothed. Crucial is the choice of threshold level, distinguishing between coefficients related predominantly to estimation errors and those associated with the underlying true function. The standard wavelet threshold operation with a constant or level-dependent threshold can not be applied to wavelet coefficients of spectral denstity functions. The non-stationarity in the statistical properties of these estimators reveals itself in the wavelet domain as significant peaks. An efficient threshold level should follow the standard deviation of each wavelet coefficient. New exact expressions of the standard deviation are presented, using the fact that we are dealing with functions associated with linear time invariant systems. An estimator based on these expressions proves to provide an appropriate threshold level.

1 Introduction

The application of wavelets in practice generally follows the wavelet denoising scheme of Donoho (1995), based on the compression ability of wavelets and the associated possibility of efficient noise reduction. In the field of system identification the influence of wavelets has been relatively small. The application has been studied for variance reduction in auto spectral density function estimators, [9][11], and research is undertaken to develop model structures in terms of wavelets [1]. Bodin employs wavelets for smoothing the empirical transfer function estimate by combining wavelet denoising with the traditional spectral windowing [2].

We will introduce a technique for variance reduction in non-parametric transfer function estimators based on the application of wavelets, as an alternative to the use of spectral windows. The theory behind the method is characterized by the fact that explicit use is made of the properties of wavelets and of the fact that the functions under consideration are linear time invariant. In particular, an estimator of the standard deviation of the wavelet coefficients is presented, serving as a threshold level corresponding with the behaviour of the estimation errors in the wavelet coefficients.

2 Background

The systems under consideration are taken to be scalar, linear, time-invariant and discrete. Discrete data $\{u(t), y(t)\}$ of length N is available. The observed output $y(t)$ is considered as consisting of the output $\tilde{y}(t)$ due to the system dynamics and an additive term $v(t)$, representing all random influences. Both the input $u(t)$ and the noise term $v(t)$ have normal distributions $\mathcal{N}(0, \sigma_u)$ and $\mathcal{N}(0, \sigma_v)$ and are mutually uncorrelated. The transfer function $G_o(f)$ is estimated by use of non-parametric estimators of the spectral density functions $\Phi_{\mathrm{u}}(f)$ and $\Phi_{yu}(f)$,

$$\hat{G}(f) = \frac{\hat{\Phi}_{yu}(f)}{\hat{\Phi}_{\mathrm{u}}(f)}. \tag{1}$$

Well known estimators, found in e.g. ([10][6]), are given by $\hat{\Phi}_{yu}(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{C}_{yu}(\tau)e^{-i2\pi f\tau}$,where $\hat{C}_{yu}(\tau)$ represents the estimator of the covariance function $C_{yu}(\tau)$ given by

$$\hat{C}_{yu}(\tau) = \begin{cases} \frac{1}{N}\sum\limits_{t=1}^{N-\tau} y(t)u(t+\tau) & , N > \tau > 0 \\ \frac{1}{N}\sum\limits_{t=1-\tau}^{N} y(t)u(t+\tau) & , -N < \tau \le 0 \end{cases} \tag{2}$$

These estimators are not consistent:

$$var[\hat{\Phi}_{yu}(f)] \approx \begin{cases} \Phi_{\mathrm{u}}(f)\Phi_y(f), \ f \neq 0 \text{ or } \pm f_{(N)} \\ \Phi_{\mathrm{u}}(f)\Phi_y(f) + \left|\Phi_{yu}(f)\right|^2, \\ \qquad f = 0, \pm f_{(N)} \end{cases} \tag{3}$$

with $f_{(N)}$ the Nyquist frequency. Though asymptotically unbiased and uncorrelated in neighbouring frequencies, their variance is independent of the sample size N.[1] The resulting characteristic erratic and wildly fluctuating behaviour necessitates further processing in order to reduce their variance.

[1] An important exception is formed in case of a periodic input.

3 Wavelets and wavelet denoising

Variance reduction is traditionally achieved by means of spectral windowing. The mechanism involved in the use of lag windows can be generalized by considering series expansions of the spectral estimator $\hat{\Phi}(f)$ in orthogonal basis functions $\phi_l(f)$.

3.1 Wavelets

We consider periodic orthogonal discrete wavelets allowing for a multiresolution analysis (see e.g. [7][3][8]). Use is made of a transform known as the Shift Invariant Wavelet Transform (cf. [4]), obtained by projecting $2N$ samples of one period of the spectral estimator $\hat{\Phi}(f_l)$ onto all the $2N$ integer translates of wavelet and scaling functions $\bar{\varphi}_j(f_l)$ and $\bar{\psi}_j(f_l)$ of scale j,

$$\begin{aligned} \bar{c}_j(f_l) &= \frac{1}{2N}\sum_{m=1}^{2N}\hat{\Phi}(\zeta_m)\bar{\varphi}_j^*(\zeta_m - f_l) \\ \bar{d}_j(f_l) &= \frac{1}{2N}\sum_{m=1}^{2N}\hat{\Phi}(\zeta_m)\bar{\psi}_j^*(\zeta_m - f_l). \end{aligned} \quad (4)$$

The standard wavelet coefficients $d_j(k)$ follow from a downsampling with a factor 2^j ($d_j(k) = \downarrow 2^j\left[\bar{d}_j(f_l)\right]$). The Inverse Discrete Fourier Transforms of the basis functions $\varphi_j(-f_l)$ and $\bar{\psi}_j(-f_l)$, denoted by $G_j(\tau) = G(2^{j-1}\tau)\prod_{l=1}^{j-1}H(2^{l-1}\tau)$ and $H_j(\tau) = \prod_{l=1}^{j}H(2^{l-1}\tau)$ are complementary low and a high pass filters.

3.2 Thresholding expansion coefficients

The spectral estimator $\hat{\Phi}(f)$ is considered to consist of the true function $\Phi(f)$ and an additive error term $V(f)$, representing the estimation errors, i.e. $\hat{\Phi}(f) = \Phi(f) + V(f)$. The linear operation of projecting $\hat{\Phi}(f)$ onto orthonormal basis functions $\phi_l(f)$[2] correspondingly results in expansion coefficients $\hat{\alpha}_l = \alpha_l + \alpha_{l,V}$. A smoothed estimator $\hat{\Phi}_\delta(f)$ follows from

$$\hat{\Phi}_\delta(f) = \sum_{l=1}^{\infty}\delta_l\hat{\alpha}_l\phi_l(f), \quad (5)$$

where the scaling factor δ_l can reduce the influence of the error term $V(f)$ by altering or omitting in the series expansion those coefficients $\hat{\alpha}_l$ contributing, in some norm, more to the error term $V(f)$ than to the true function $\Phi(f)$. In the field of wavelets such an operation is known as *thresholding* [5]. The resulting variance reduction can be seen as a smoothing operation with a frequency dependent window $P(f,g) = \sum_{l=1}^{\infty}\delta_l\phi_l^*(g)\phi_l(f)$; as follows from,

$$\begin{aligned} \hat{\Phi}_\delta(f) &= \sum_{l=1}^{\infty}\delta_l\left\{\int_{-\infty}^{\infty}\hat{\Phi}(g)\phi_l^*(g)dg\right\}\phi_l(f) \\ &= \int_{-\infty}^{\infty}\hat{\Phi}(g)\left\{\sum_{l=1}^{\infty}\delta_l\phi_l^*(g)\phi_l(f)\right\}dg. \end{aligned} \quad (6)$$

[2]Here ϕ_l represents either a scaling function or wavelet with associated coefficients α_l.

4 Proposal of a variance reduction scheme

We propose a variance reduction scheme based on the wavelet denoising of expression (6) of the spectral density function estimators $\hat{\Phi}_u(g)$ and $\hat{\Phi}_{yu}(f)$:

$$\hat{G}_\delta(f) = \frac{\int_{-f(N)}^{f(N)}P(f,g)\hat{\Phi}_{yu}(g)dg}{\int_{-f(N)}^{f(N)}P(f,g)\hat{\Phi}_u(g)dg}. \quad (7)$$

The same threshold operation $\delta_{j,k}$ following from denoising $\hat{\Phi}_{yu}(f)$, is applied to the wavelet coefficients of $\hat{\Phi}_u(f)$.

5 Thresholding the cross spectral density function estimator

Figure 1 depicts the wavelet coefficients of an estimator of $\Phi_{yu}(f)$. The wavelet transform has been able to compress the information on the true function in a small number of relatively large coefficients, allowing for an effective variance reduction. However, the choice of threshold level must be based, without knowledge on the true function, on the availability of a single realization only.

The variance of the wavelet coefficients $d_j(f_l)$ can, with the shift invariant coefficients $\bar{d}_j(f_l)$ of expression (4), be expressed in terms of the covariance matrix of the estimation error $V(f)$,

$$var[\bar{d}_j(f_l)] = E\left[\left|\bar{d}_j(f_l) - E[\bar{d}_j(f_l)]\right|\right]^2 = \quad (8)$$
$$\frac{1}{(2N)^2}\sum_{m=1}^{2N}\sum_{n=1}^{2N}cov[V(\zeta_m),V(\zeta_n)]\bar{\psi}_j(\zeta_m - f_l)\bar{\psi}_j^*(\zeta_n - f_l).$$

Since the statistical properties of the spectral estimator $\hat{\Phi}_{yu}(f)$ are proportional to the underlying true function (see (3)), (8) shows the standard deviation in the wavelet coefficients to vary greatly over the coefficients (see figure 1). The generally used constant threshold level will not perform satisfactorily. A proper threshold level should follow the significantly varying standard deviation.

6 Estimating the variance in the wavelet coefficients

In practice, the specific features of the variance in the wavelet coefficients of $\hat{\Phi}_{yu}(f)$ have to be estimated from one realization of the data. Available approximate expressions, as (3), of the covariance of $\hat{\Phi}_{yu}(f)$ are too inaccurate when based on a small number of data points. Exact expressions that are valid for finite data sets can be derived though, in terms of the covariance function estimator $\hat{C}_{yu}(\tau)$, by applying the Fourier Transform to (8).

6.1 Theoretical expression for the variance in wavelet coefficients of a cross spectral density function estimator

From (8), with $V(f_l) * \bar{\psi}_j(-f_l)$ denoting the convolution between $V(f_l)$ and the reversed shift invariant wavelet function $\bar{\psi}_j(-f_l)$, it follows that

$$var[\bar{d}_j(f_l)] = E\left[\{V(f_l) * \bar{\psi}_j(-f_l)\}\{V(f_l) * \bar{\psi}_j(-f_l)\}^*\right].$$

The straightforward application of the Inverse Discrete Time Fourier Transform (IDFT) and exploiting the dual relationship between convolution and multiplication yields,

$$\begin{aligned}
&\mathcal{F}^{-1}\left\{var[\bar{d}_j(f_l)]\right\} \\
&= \frac{1}{2N}\sum_{k=1}^{2N} E\left[v(k)v^*(k-\tau)\right] G_j(k)G_j^*(k-\tau) \\
&= \frac{1}{2N}\sum_{k=1}^{2N} E\left[\left(\hat{C}_{yu}(k) - E[\hat{C}_{yu}(\tau)]\right)\cdot\right. \\
&\left.\left(\hat{C}^*_{yu}(k-\tau) - E[\hat{C}_{yu}(k-\tau)]\right)\right] G_j(k)G_j^*(k-\tau). \quad (9)
\end{aligned}$$

Here $v(\tau)$ denotes the IDFT of $V(f)$, easily seen to represent the estimation errors in the covariance function estimator $\hat{C}_{yu}(\tau)$. It is precisely for the covariance function of the covariance function estimator $\hat{C}_{yu}(\tau)$ that an exact expression can be derived, (compare with [10])

$$\begin{aligned}
&cov\left[\hat{C}_{yu}(k), \hat{C}_{yu}(k-\tau)\right] \\
= &\frac{1}{N^2}\sum_t\sum_s \{C_{\tilde{y}\tilde{y}}(s-t)C_{\mathrm{u}}(s-t-\tau) \\
&+C_{\tilde{y}u}(s-t+k-\tau)C_{u\tilde{y}}(s-t-k)\} \\
&+\frac{1}{N^2}\sum_t\sum_s C_{vv}(s-t)C_{\mathrm{u}}(s-t-\tau), \quad (10)
\end{aligned}$$

where the ranges of summation should be taken as

$$\begin{array}{lll}
\left.\begin{array}{l} t=[1,N-k] \\ s=[1:N-k+\tau] \end{array}\right\} & , & k \geq 0 \text{ and } k-\tau \geq 0 \\
\left.\begin{array}{l} t=[1-k,N] \\ s=[1:N-k+\tau] \end{array}\right\} & , & k < 0 \text{ and } k-\tau \geq 0 \\
\left.\begin{array}{l} t=[1,N-\tau] \\ s=[1-k+\tau:N] \end{array}\right\} & , & k \geq 0 \text{ and } k-\tau < 0 \\
\left.\begin{array}{l} t=[1-k,N] \\ s=[1-k+\tau:N] \end{array}\right\} & , & k < 0 \text{ and } k-\tau < 0\ .
\end{array}$$

Expressions (9) and (10) are exact even for finite data sets. By a simple Discrete Fourier transformation and appropriate downsampling the exact variance in the wavelet coefficients is obtained.

6.2 Estimating the variance in the wavelet coefficients

An estimator of the variance in the wavelet coefficients of $\hat{\Phi}_{yu}(f)$ from a single realization of the estimator is obtained by using estimators $\hat{C}_{yy}(\tau)$, $\hat{C}_{\mathrm{u}}(\tau)$, $\hat{C}_{yu}(\tau)$ and $\hat{C}_{uy}(k)$ as in (2) for the occurring covariance functions in expression (10). Note that, $C_{yy}(\tau) = C_{\tilde{y}\tilde{y}}(\tau) + C_{vv}(\tau)$ and $C_{yu}(\tau) = C_{\tilde{y}u}(\tau)$, as $u(t)$ and $v(t)$ are assumed to be uncorrelated.

Figure 1 depicts an example, based on the test system found in [2], of an estimation of the standard deviation in the wavelet domain based on one realization of the data and the

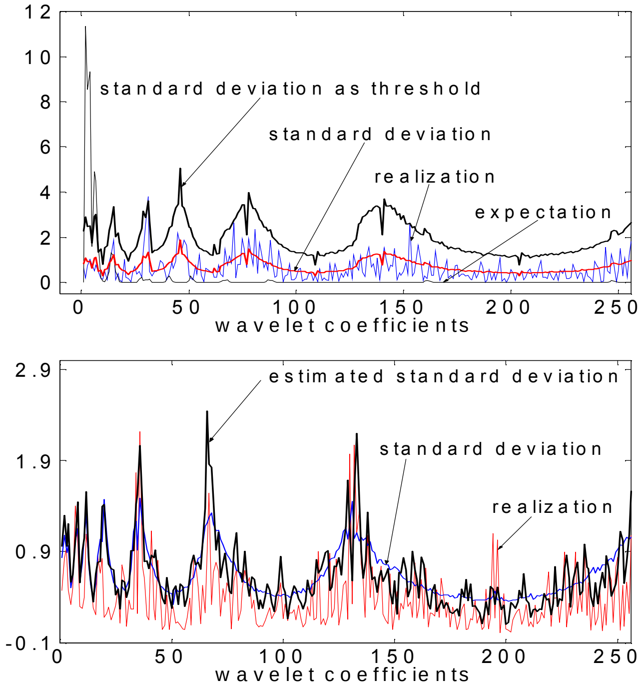

Figure 1: Top: *Wavelet coefficients (coifman-5) of a realization of an estimator* $\hat{\Phi}_{yu}(f)$*, with expectation equal to the true function* $\Phi_{yu}(f)$*. The variance in the estimation errors is highly nonstationary over the coefficients. An efficient threshold would be based on scaling the standard deviation. The standard deviation is calculated from a Monte Carlo simulation* (1000 *realizations) based on* $N = 512$, $v(t)$ *and* $u(t)$ *being white noise with* $\sigma_u^2 = 1$ *and* $\sigma_v^2 = 0.0025$. Bottom: *Estimation of the standard deviation based on a single realization using expressions* (9) *and* (10) *. The estimate provides for an effective threshold.*

use of expressions (9) and (10) . As an estimate of the standard deviation itself the estimator is obviously quite poor being based on one realization. However, the features of the particular realization of the estimation error in the wavelet coefficients are followed remarkably well. The estimator practically encloses the estimation error in the wavelet coefficients of the particular realization on which the estimator is based. As such it serves as a very efficient threshold by being able to clearly assess the presence of estimation errors in each wavelet coefficient.

Monte Carlo simulations indeed reveal that thresholding with this estimate results in a low mean square error, even lower than when applying the real standard deviation (cf. figure 2). The present implementation of (10) , however, is such that only data sets upto $N = 512$ are feasible. It is for these small data sets that often the functions to be smoothed do not contain sufficient information. Figure 3 shows an example of this situation, where no information on the high frequency peak is available for identification. The fact that for this data length the proposed method performs similar to the optimal spectral windowing is quite promising. The objective of improving upon spectral windowing could be obtained when the implementational issue will be solved.

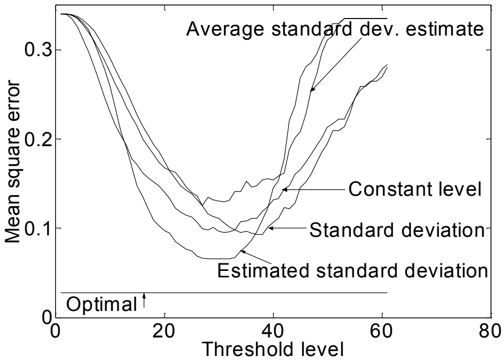

Figure 2: *Mean square error for thresholding wavelet coefficients of an estimator $\hat{\Phi}_{yu}(f_l)$ with constant threshold, standard deviation, an estimate of the average standard deviation and with the estimated standard deviation. For comparison the theoretical minimum, achieved with $\delta_{j,k} = \frac{|d_j(k)|^2}{|d_j(k)|^2+\sigma^2}$, is given.*

7 Conclusions

Wavelet denoising provides for an effective variance reduction in a transfer function estimator, when based on the consecutive thresholding of the coefficients of the (cross) spectral density function estimator in its numerator and denominator. Theoretical expressions of the standard deviation in the wavelet coefficients allow for the derivation of threshold level which, based on one realization, efficiently removes the estimation errors. Further research is required on the choice of wavelet type, since wavelet denoising is based on the assumption of a compact representation of the function to be smoothed. Empirical results indicate that the proposed method is a promising alternative to the traditional spectral windowing.

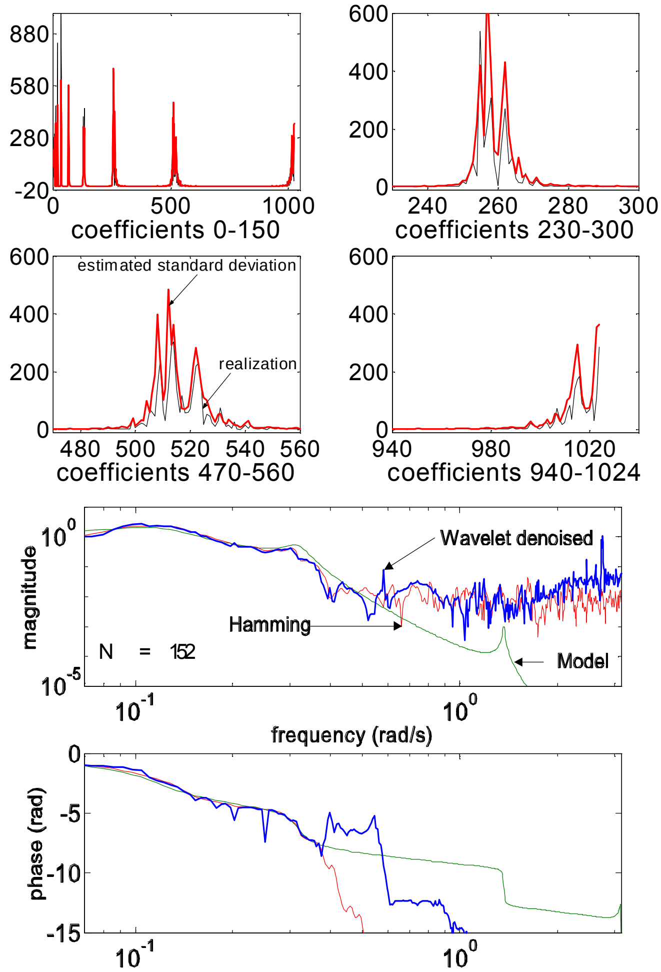

Figure 3: *Smoothing of a transfer function estimate of a rotating drive system ($N = 512$) with an optimized Hamming window versus smoothing with the proposed denoising method. As reference a sine-swept model is given. On top are depicted the wavelet coefficients of the associated $\hat{\Phi}_{yu}(f)$ with the estimated standard deviation used as threshold.*

References

[1] M. Basseville, A. Benveniste, K.C. Chou, S.A. Golden, R. Nikoukhah, and A.S. Willsky. Modeling and estimation of multiscale stochastic processes. *IEEE Transactions on Information Theory*, 38(2):766–784, March 1992.

[2] P. Bodin. *On the Selection of Best Orthonormal Basis in System Identification and Signal Processing*. PhD thesis, Royal Institute of Technology, Department of Signals, Sensors and Systems, Stockholm, Sweden, 1997.

[3] C.S. Burrus, R.A. Gopinath, and H. Guo. *Introduction to Wavelets and Wavelet Transform. A Primer*. Prentice-Hall, New Yersey, USA, 1998.

[4] R.R. Coifman and D.L. Donoho. Translation invariant denoising. In A. Antoniades, editor, *Wavelets and Statistics, Springer Lecture Notes*. Springer-Verlag, 1995.

[5] D.L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[6] G. M. Jenkins and D. W Watts. *Spectral Analysis and its Applications*. Holden-Day, San Francisco, California, 1968.

[7] G. Kaiser. *A Friendly Guide to Wavelets*. Birkhauser, Boston, USA, 1994.

[8] Y. Meyer. *Wavelets. Algorithms and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 1993. Translated and Revised by Ryan, R.D.

[9] P. Moulin. Wavelet thresholding techniques for power spectrum estimation. *IEEE Transactions on Signal Processing*, 42(11):3126–3136, November 1994.

[10] M.B. Priestley. *Spectral Analysis and Time Series. Univariate Series*, volume 1. Academic Press, London, UK, 1981.

[11] A.T Walden, D.B. Percival, and E.J. McCoy. Spectrum estimation by wavelet thresholding of multitaper estimators. Technical report, Imperial College of Science, Technology and Medicine, London, UK, 1997.