CDC00-REG1120

# 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
Email: s.g.douma@tn.tudelft.nl

*Abstract*— **This article presents a variance reduction scheme 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. The efficiency depends on the degree in which the information on the true function and on the estimation errors is separated in the transform domain and on the choice of threshold operation. 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 nonstationarity 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 be an appropriate threshold level.**

*Keywords*—**System identification, transfer function estimation, wavelets, variance reduction, spectral estimation, frequency dependent window**

## I. Introduction

Wavelets, developed since the 1980's, have found their use in various fields of science as mathematics, physics, geology and image processing. Extensive literature is available on the theory of wavelets and their properties ([11][9][5]). Their application 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 their influence has not been of comparable degree. The application has been studied for variance reduction in auto spectral density function estimators ([12][14]) and research is undertaken to develop model structures in terms of wavelets ([6][1][2]). Bodin (1995) employs wavelets for smoothing the empirical transfer function estimate by combining wavelet denoising with the traditional spectral windowing.

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 in the wavelet coefficients is presented, serving as a threshold level corresponding with the behaviour of the estimation errors in the wavelet coefficients.

Background information and properties of the spectral estimators are given in sections II. Section III deals with wavelet thresholding and emphasizes its potential by showing that resulting variance reduction follows from windowing with a frequency dependent window. Section IV shows that the direct application of the standard wavelet thresholding to the empirical transfer function estimate is infeasible. Section V then presents the new variance reduction method. Section VI motivates the use of a threshold level following the standard deviation in the wavelet coefficients. An estimator of the standard deviation is derived in section VII, which is shown to serve as an appropriate threshold level.

## II. Background

### A. Situation under consideration

We adopt the following system representation in the frequency domain:

$$Y(f) = G_o(f)U(f) + V(f), \tag{1}$$

where the transfer function $G_o(f)$, given by

$$G_o(f) = \sum_{k=-\infty}^{\infty} g_o(k)e^{-i2\pi fk}, \tag{2}$$

relates the Discrete Time Fourier Transforms of the input $u(t)$ and the observed output $y(t) = \tilde{y}(t) + v(t)$.[1] The underlying system is taken to be scalar, linear, time-invariant and discrete, letting the variable $t$ be an integer. The data $\{u(t), y(t)\}$ is known within an interval $[1, N]$, while the additive term $v(t)$ to the output $\tilde{y}(t)$ of the system represents all random influences. Both the input $u(t)$ and the noise term $v(t)$ are considered to be zero mean Gaussian distributed stationary stochastic sequences.

The finiteness and random behaviour of the observed input and output data implies that the derivation of the transfer function $G_o(f)$ amounts to estimating this function. Here non-parametric estimators will be considered.

### B. Transfer function estimator

The deterministic behaviour of the underlying system can be deduced from the observed stochastic input and output data by use of their ensemble properties. The transfer function $G_o(f)$ is expressed in the cross spectral density function of the input and the output and the spectral density of the input by

$$G_o(f) = \frac{\Phi_{yu}(f)}{\Phi_{\mathrm{u}}(f)}. \tag{3}$$

[1] Due to the fact that the energy of a random process is unlimited, the standard Discrete Time Fourier Transform has to be modified for stochastic signals. Define, for data length $N$, $U_N(f) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} u(t)e^{-i2\pi ft}$ for any (stochastic) signal $u(t)$. Omittance of the subscript $N$, as in equation 1, will denote the limit, if existing, of the transform for $N \to \infty$.

This expression is exact under the condition that the input $u(t)$ is uncorrelated with the additive noise term $v(t)$.

Expression (3) allows the problem of estimating the transfer function $G_o(f)$ to be transferred to the problem of estimating the (cross) spectral density functions $\Phi_{yu}(f)$ and $\Phi_{\mathrm{u}}(f)$. Well-known estimators, found in e.g. [13] and [8], are given, for input and output observations of data length $N$, by

$$\hat{\Phi}_{yu}(f) = Y_N(f)U_N^*(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{C}_{yu}(\tau)e^{-i2\pi f\tau}, \quad (4)$$

where $\hat{C}_{yu}(\tau)$ represents the estimator of the covariance function $C_{yu}(\tau)$ given by

$$\hat{C}_{yu}(\tau) = \begin{cases} \frac{1}{N}\sum_{t=1}^{N-\tau} y(t)u(t+\tau) & , N > \tau > 0 \\ \frac{1}{N}\sum_{t=1-\tau}^{N} y(t)u(t+\tau) & , -N < \tau \leq 0 \end{cases} \quad (5)$$

The corresponding transfer function estimator, given by

$$\hat{G}_{ETFE}(f) = \frac{\hat{\Phi}_{yu}(f)}{\hat{\Phi}_{\mathrm{u}}(f)} = \frac{Y_N(f)U_N^*(f)}{U_N(f)U_N^*(f)} = \frac{Y_N(f)}{U_N(f)} \quad (6)$$

is commonly known as the *empirical transfer function estimate*.

### C. Properties of the spectral estimators

Both the (cross) spectral density function estimators $\hat{\Phi}_{yu}(f)$ and $\hat{\Phi}_{\mathrm{u}}(f)$ and the empirical transfer function estimate $\hat{G}_{ETFE}(f)$ are not consistent. Though asymptotically unbiased and uncorrelated in neighbouring frequencies, they do not converge in the mean square sense to the functions to be estimated.[2] In particular, with $f_{(N)}$ the Nyquist frequency (e.g. [8], [13])

$$var[\hat{\Phi}_{yu}(f)] \approx \begin{cases} \Phi_{\mathrm{u}}(f)\Phi_y(f) + |\Phi_{yu}(f)|^2, \ f = 0, \pm f_{(N)} \\ \Phi_{\mathrm{u}}(f)\Phi_y(f), \ f \neq 0 \text{ or } \pm f_{(N)} \end{cases} \quad (7)$$

The variance is independent of the sample size $N$, which, in combination with the neighbouring values being uncorrelated, results in an erratic and wildly fluctuating form making them of practically no use as an estimate of the spectral density functions. The resulting empirical transfer function estimate exhibits a similar behaviour. To turn $\hat{\Phi}_{yu}(f)$, $\hat{\Phi}_u(f)$ and $\hat{G}_{ETFE}(f)$ into proper estimators, further processing is required by which the variance is reduced.

## III. 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 expansion of the spectral estimator $\hat{\Phi}(f)$ in orthogonal basis functions $\phi_l(f)$.

### A. Wavelets

We consider the series expansion of a spectral estimator in periodic orthogonal discrete wavelets allowing for a multiresolution analysis. For details the reader is referred to [9][5][11]. The Periodic Discrete Wavelet Transform ($c_j(k)$, $d_j(k)$) is induced by projecting the sampled spectral estimator $\hat{\Phi}(f_l)$ onto the two-dimensional family of orthonormal functions $\varphi_{j,k}(f_l)$ and $\psi_{j,k}(f_l)$, generated by dilation and translation of a basic scaling function $\varphi(f_l)$ and associated wavelet $\psi(f_l)$ (for $k, j \in \mathbb{Z}$)

$$\begin{aligned} \varphi_{j,k}(f_l) &= 2^{-\frac{j}{2}}\varphi(2^{-j}f_l - k) = 2^{-\frac{j}{2}}\varphi_j(f_l - 2^j k) \\ \psi_{j,k}(f_l) &= 2^{-\frac{j}{2}}\psi(2^{-j}f_l - k) = 2^{-\frac{j}{2}}\psi_j(f_l - 2^j k). \end{aligned} \quad (8)$$

Note that the translations are taken in steps of $2^j$. But for this downsampling operation the wavelet transform could be interpreted as a straightforward convolution of the estimator $\hat{\Phi}(f_l)$ and the scaling function $\varphi_j(-f_l)$ or wavelet $\psi_j(-f_l)$ at scale $j$. We use a transform known as the Shift Invariant Wavelet Transform, induced by the coefficients $\bar{c}_j(f_l)$ and $\bar{d}_j(f_l)$ obtained by projecting $2N$ samples of one period of the spectral estimator $\hat{\Phi}(f_l)$ onto all the $2N$ integer translates of basis functions $\bar{\varphi}_j(f_l)$ and $\bar{\psi}_j(f_l)$,

$$\begin{aligned} \bar{c}_j(f_l) &= \left\langle \hat{\Phi}(f_l), \bar{\varphi}_{j,k}(f_l) \right\rangle = \frac{1}{2N}\sum_{m=1}^{2N} \hat{\Phi}(\zeta_m)\bar{\varphi}_j^*(\zeta_m - f_l) \\ \bar{d}_j(f_l) &= \left\langle \hat{\Phi}(f_l), \bar{\psi}_{j,k}(f_l) \right\rangle = \frac{1}{2N}\sum_{m=1}^{2N} \hat{\Phi}(\zeta_m)\bar{\psi}_j^*(\zeta_m - f_l). \end{aligned} \quad (9)$$

The original 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 wavelet transform transforms $2N = 2^{\bar{M}}$ samples of one period of the sampled spectral estimator $\hat{\Phi}(f_l)$ into $2N$ coefficients $c_{j_M}(k)$ and $d_j(k)$ representing the correspondence between $\hat{\Phi}(f_l)$ and the periodized discrete scaling functions $\varphi_{j,k}(f_l)$ and wavelets $\psi_{j,k}(f_l)$ of a particular scale $j \in [0, \bar{M}]$ and frequency $2^j k$ ($k \in [0, 2^{\bar{M}-j}]$).

The Inverse Discrete Fourier Transforms of basis functions $\varphi_j(-f_l)$ and $\psi_j(-f_l)$ (and $\bar{\varphi}_j(-f_l)$ and $\bar{\psi}_j(-f_l)$, as they only differ in the manner of translation) are given by

$$\begin{aligned} G_j(\tau) &= \mathcal{F}^{-1}\left\{\psi_j(-f_l)\right\} = G(2^{j-1}\tau)\prod_{l=1}^{j-1} H(2^{l-1}\tau) \\ H_j(\tau) &= \mathcal{F}^{-1}\left\{\varphi_j(-f_l)\right\} = \prod_{l=1}^{j} H(2^{l-1}\tau) \end{aligned} \quad (10)$$

where the filters $H(\tau)$ and $G(\tau)$ are conjugate quadrature mirror filters. The filters $H_j$ and $G_j$ are complementary low-pass and a high-pass band filter, respectively. Figure 1 shows how each wavelet coefficient contains information on details in the spectral estimator not only associated with a particular place in the estimator but as well with a particular scale, or "frequency band".

### B. Thresholding expansion coefficients

The spectral estimator is considered to consist of the true spectral function $\Phi(f)$ and an additive error term $V(f)$, representing the estimation errors.[3] The linear operation of project-

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

[3] Note that we consider the estimator $\hat{\Phi}(f)$ to be unbiased. If this assumption is not valid, in the following discussion the true spectral density function $\Phi(f)$ should be replaced by the sum of $\Phi(f)$ and a term representing the bias. Variance reduction is then optimized with respect to the biased function.

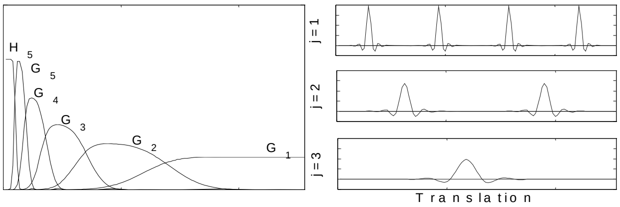


Fig. 1. Left: *Partitioning of the 'wavelet frequency domain' into bands related to wavelets and scaling functions of scale $j$.* Right: *Translation and dilation of a scaling function (coifman-5). The scaling functions are localized in time and in scale.*

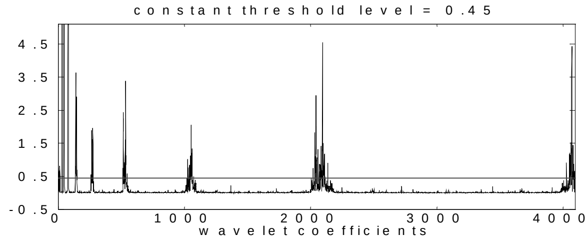


Fig. 2. *Wavelet coefficients of $\hat{G}_{ETFE}(f_l)$ associated with the rotating drive system of figure 5. Coefficients above 500 contain no significant information on the true function. The large peaks are due to the influence of the stochastic input outside the time interval under consideration. Here even for $N = 4096$ a constant threshold level can not be used.*

ing $\hat{\Phi}(f)$ onto orthonormal basis functions $\phi_l(f)$[4] then results in expansion coefficients $\hat{\alpha}_l$ to consist of the sum of an element $\alpha_l$ associated with the true spectral function $\Phi(f)$ and an element $\alpha_{l,V}$ related to the estimation error $V(f)$,

$$\begin{aligned} \hat{\Phi}(f) &= \Phi(f) + V(f) \\ \hat{\alpha}_l &= \alpha_l + \alpha_{l,V}. \end{aligned} \tag{11}$$

A new estimator $\hat{\Phi}_\delta(f)$, based on altering the expansion coefficients $\hat{\alpha}_l$ with a factor $\delta_l$, is given by

$$\hat{\Phi}_\delta(f) = \sum_{l=1}^{\infty} \delta_l \hat{\alpha}_l \phi_l(f). \tag{12}$$

The scaling factor $\delta_l$ can reduce the influence of the error term $V(f)$ in the spectral estimator by altering or omitting in the series expansion those basis functions $\phi_l(f)$, whose coefficients $\hat{\alpha}_l$ contribute, in some norm, more to the error term $V(f)$ than to the true spectral function $\Phi(f)$. The independency of each of the coefficient guarantees that altering or removing one coefficient will not affect the others. In the field of wavelet theory an operation as 12 is known as *thresholding* the expansion coefficients $\hat{\alpha}_l$ [7].

### C. Threshold level

A proper bias and variance trade-off, inherent in the thresholding operation 12, can not be made on the basis of the availability, in practice, of a single realization of the spectral estimator $\hat{\Phi}(f_l)$. In order to remove those wavelet coefficients predominantly representing information on the estimation error $V(f)$, we have to resort to ensemble properties of the wavelet coefficients. Donoho and Johnstone (1995) showed that the maximum values of a particular realization of a Gaussian white noise are related by a factor, asymptotically $\sqrt{2\log(N)}$, to the standard deviation. Omittance of all coefficients under the so-called *universal* threshold $\sigma\sqrt{2\log(N)}$ ensures that, if the expectation of a coefficient $\hat{\alpha}_l$ is zero, the probability of keeping the noisy coefficient vanishes asymptotically.

In practice, due to the finiteness of the data, one resorts to scaling the standard deviation with a factor used as a design variable comparable to the bandwidth parameter in spectral windowing. [14][12] [5]

[4]Here $\phi_l$ represents either a scaling function or wavelet with associated coefficients $\alpha_l$.

### D. Why wavelets are suitable

The variance reduction resulting from the threshold operation can be made more explicit. Equation (12) can be written as,[5]

$$\begin{aligned} \hat{\Phi}_\delta(f) &= \sum_{l=1}^{\infty} \delta_l \left\{ \int_{-\infty}^{\infty} \hat{\Phi}(g)\phi_l^*(g)dg \right\} \phi_l(f) \\ &= \int_{-\infty}^{\infty} \hat{\Phi}(g) \left\{ \sum_{l=1}^{\infty} \delta_l \phi_l^*(g)\phi_l(f) \right\} dg \\ &= \int_{-\infty}^{\infty} \hat{\Phi}(g) P(f,g) dg. \end{aligned} \tag{13}$$

The smoothing window $P(f,g)$ results from altering the influence of some of the basis functions in the series expansion of the estimator $\hat{\Phi}(f)$. Alternatively, the omittance of coefficients in any series expansion results in the projection onto a lower dimensional subspace.

In traditional spectral windowing the application of the lag window to the covariance function estimator $\hat{C}(\tau)$ results, due to the specific character of the complex exponentials associated with the Discrete Time Fourier Transform, in a convolution with a spectral window of fixed form. For general basis functions, the resulting window will be frequency dependent.

Note that the omittance of all wavelet coefficients upto a certain scale $j$, will result exactly in the application of a lag window $H_j(\tau)$ and resulting smoothing operation with a fixed spectral window $\varphi_j(-f)$ (fig. 1). The potency in the use of wavelets lies with the fact that each wavelet coefficient can not only be associated with a "frequency band", but with a particular interval around $f = 2^j k$ as well. As the coefficients are independent due to the orthonormality of the wavelets, the smoothing operation, the trade-off between bias and variance, can be made locally.

## IV. Wavelet denoising directly to the empirical transfer function estimate

The direct application of the thresholding scheme to the empirical transfer function estimate seems at first sight to be attractive. Well-known expressions by Ljung (1985) for the statistical properties of $\hat{G}_{ETFE}(f)$ indicate the estimator to be asymptotically unbiased and uncorrelated, exhibiting a variance given

[5]For convenience, continuous functions are considered. Strictly speaking, $\hat{\Phi}(f)$ is not necessarily contained in $L^2(\mathbb{R})$. In practice, we consider $\hat{\Phi}(f_l)$ which will be a function in $\ell^2(\mathbb{R})$, allowing for an exact representation in basis functions $\phi(f_l) \in \ell^2(\mathbb{R})$.

by the signal to noise ratio [10]. For white input and noise sequences, or at least sequences having a smooth spectral density function, the standard deviation in the associated wavelet coefficients is then constant. A simple constant threshold would provide for a proper thresholding operation.

However, the behaviour of $\hat{G}_{ETFE}(f)$ will be erratic corresponding to the particular realization of the input, with severe outliers where the denominator in (6) is close to zero. But more important, the expressions of Ljung are derived under the assumption that the input $u(t)$ is deterministic or periodic. In a general setup, the past values of the input $u(t)$ are not known and the input $u(t)$ will be stochastic as its realizations differ for different experiments. This implies that the influence of the values of the input outside the time interval under consideration should be ascribed to the variance. (see also [4]) The additional influence to the variance is proportional to the magnitude of the underlying true transfer function. For transfer functions exhibiting both large and small values in their magnitude this implies that the variance in both $\hat{G}_{ETFE}(f)$ and the associated wavelet coefficients is severely nonstationary (fig. 2).

The application of a narrow Hamming window prior to the application of the wavelet denoising scheme, as proposed by Bodin (1997), will reduce both the magnitude of the variance and its variation over frequencies by correlating the values of $\hat{G}_{ETFE}(f)$. However, Bodin's method is still based on the variance expression of Ljung (1987) and does not consider the influence of the stochastic input outside the interval under consideration. Particularly in the situation of a transfer function to be estimated exhibiting significant variations in its magnitude, the nonstationarity in the variance in the wavelet coefficients will be significant even after the application of a spectral window. Moreover, the choice of the right combination of threshold level and bandwidth of the spectral window is not trivial.

Instead, we suggest the use of a threshold level explicitly following the nonstationary standard deviation in each coefficient. This is possible by exploiting the fact that the functions are associated with linear time-invariant systems.

## V. Proposal of a variance reduction scheme

We propose a variance reduction scheme for smoothing an empirical transfer function estimate $\hat{G}_{ETFE}(f)$, based on the wavelet denoising of expression (13) of the cross spectral density function estimator $\hat{\Phi}_{yu}(f)$ and the spectral density function estimator $\hat{\Phi}_{\mathrm{u}}(f)$ in expression (6):

$$\hat{G}_{\delta}(f) = \frac{\int_{-\frac{1}{2}}^{\frac{1}{2}} P^{num}(f,g)\hat{\Phi}_{yu}(g)dg}{\int_{-\frac{1}{2}}^{\frac{1}{2}} P^{den}(f,g)\hat{\Phi}_{\mathrm{u}}(g)dg}. \quad (14)$$

Though the literature on spectral windowing (e.g. [13]) suggests the use of a window with different bandwidth for $\hat{\Phi}_{yu}(f)$ and for $\hat{\Phi}_{\mathrm{u}}(f)$, it should be realized that the cross spectral density function estimator $Y_N(f)U_N(f)$ is directly related, through the true transfer function, to the spectral density function estimator $|U_N(f)|^2$. The variations following the features of the particular realization of $U_N(f)$, are more significantly present in the term $Y_N(f)U_N(f)$ than the influence of the true transfer function on its expectation. Empirical results verify that the strong correlation between numerator and denominator requires an identical smoothing operation.

As it is the cross spectral density function in the numerator which contains the system information we are interested in, we base the wavelet denoising on thresholding the wavelet coefficients of $\hat{\Phi}_{yu}(f)$. Subsequently, the same threshold operation $\delta_{j,k}$ is applied to the wavelet coefficients of the denominator $\hat{\Phi}_{\mathrm{u}}(f)$.

## VI. Thresholding the cross spectral density function estimator

Figure 3 depicts the wavelet coefficients of an estimator of the spectral density function $\Phi_{yu}(f)$. The wavelet transform has been able to compress the information on the true function in a small number of relatively large coefficients. However, the choice of threshold level must be based, without knowledge on the true function, on the availability of a single realization only. A proper threshold level should follow the significantly varying standard deviation.

The variance of the wavelet coefficients $d_j(k)$ can be expressed in terms of the covariance matrix of the estimation error $V(f)$. With expressions (11), (8) and (9), we have for the shift invariant wavelet coefficients $\bar{d}_j(f_l)$

$$\begin{aligned}
&var[\bar{d}_j(f_l)] = E[|\bar{d}_j(f_l) - E[\bar{d}_j(f_l)]|]^2 \\
=&E[\{\frac{1}{(2N)^2}\sum_{m=1}^{2N} V(\zeta_m)\bar{\psi}_j(\zeta_m - f_l)\} \cdot \\
&\{\sum_{n=1}^{2N} V^*(\zeta_n)\bar{\psi}_j^*(\zeta_n - f_l)\}] \\
=&\frac{1}{(2N)^2}\sum_{m=1}^{2N}\sum_{n=1}^{2N} cov[V(\zeta_m), V(\zeta_n)]\bar{\psi}_j(\zeta_m - f_l)\bar{\psi}_j^*(\zeta_n - f_l).
\end{aligned} \quad (15)$$

The standard deviation for the wavelet coefficients $d_j(k)$ follows from downsampling $(var[d_j(k)] =\downarrow 2[var[\bar{d}_j(f_l)]])$. Since the statistical properties of the spectral estimator $\hat{\Phi}_{yu}(f)$ are proportional to the underlying true function (see expression (7)), expression (15) shows the standard deviation in the wavelet coefficients to vary greatly over the coefficients (see figure 3).

## VII. 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 an input $u(t)$ and output $y(t)$. The main problem lies with the fact that available expressions of the covariance of the cross spectral density function estimator like (7) are too approximate for estimators based on a small number of input and output data. 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 expression (15).

### A. *Theoretical expression for the variance in wavelet coefficients of a cross spectral density function estimator*

From expression (15), with $V(f_l)*\bar{\psi}_j(-f_l)$ denoting the convolution between $V(f_l)$ and the reversed shift invariant wavelet transform $\bar{\psi}_j(-f_l)$, it follows that

$$var[\bar{d}_j(f_l)] = E\left[\{V(f_l) * \bar{\psi}_j(-f_l)\}\{V(f_l) * \bar{\psi}_j(-f_l)\}^*\right]. \tag{16}$$

The straightforward application of the Inverse Discrete Time Fourier Transform to expression (16) yields for the (periodic) shift invariant wavelet coefficients $\bar{d}_j(f_l)$ of scale $j$,

$$\begin{aligned} & \frac{1}{2N}\sum_{l=1}^{2N} var[\bar{d}_j(f_l)]e^{i2\pi f_l \tau} \\ = \; & E\left[\{v(\tau)G_j(\tau)\} * \{v(\tau)G_j(\tau)\}^*\right] \\ = \; & E\left[\frac{1}{2N}\sum_{k=1}^{2N} v(k)G_j(k)v^*(k-\tau)G_j^*(k-\tau)\right], \end{aligned} \tag{17}$$

where $G_j(\tau)$ denotes the wavelet filter corresponding to the wavelets $\psi_j(f)$ of scale $j$ (cf. 10) and $v(\tau)$ the Inverse Discrete Time Fourier Transform of $V(f)$. Relating the expectation operator $E[\cdot]$ to the stochastic variables only and noting from (4) that the term $v(k)$ represents the estimation errors in the covariance function estimator $\hat{C}_{yu}(\tau)$, results in

$$\begin{aligned} & \mathcal{F}^{-1}\left\{var[\bar{d}_j(f_l)]\right\} \\ & = \frac{1}{2N}\sum_{k=1}^{2N} E\left[v(k)v^*(k-\tau)\right] G_j(k)G_j^*(k-\tau) \\ & = \frac{1}{2N}\sum_{k=1}^{2N} E\left[\left(\hat{C}_{yu}(k) - E[\hat{C}_{yu}(\tau)]\right) \cdot \right. \\ & \left.\left(\hat{C}^*_{yu}(k-\tau) - E[\hat{C}_{yu}(k-\tau)]\right)\right] G_j(k)G_j^*(k-\tau). \end{aligned} \tag{18}$$

The Inverse Discrete Fourier Transform of the variance in the shift invariant wavelet coefficients $\bar{d}_j(k)$ is seen to be expressed in terms of the covariance function of the covariance function estimator $\hat{C}_{yu}(\tau)$ and the wavelet filters $G_j(\tau)$ of the associated wavelets $\psi_j(f)$. It is precisely for the covariance function estimator $\hat{C}_{yu}(\tau)$ that an exact expression can be derived,

$$\begin{aligned} & cov\left[\hat{C}_{yu}(k), \hat{C}_{yu}(k-\tau)\right] \\ = \; & \frac{1}{N^2}\sum_t\sum_s \{C_{\tilde{y}\tilde{y}}(s-t)C_{\mathrm{u}}(s-t-\tau) \\ & + C_{\tilde{y}u}(s-t+k-\tau)C_{u\tilde{y}}(s-t-k)\} \\ & + \frac{1}{N^2}\sum_t\sum_s C_{vv}(s-t)C_{\mathrm{u}}(s-t-\tau), \end{aligned} \tag{19}$$

where the ranges of summation should be taken as

$$\left.\begin{array}{ll} \left.\begin{array}{l} t = [1, N-k] \\ s = [1 : N-k+\tau] \end{array}\right\} , & k \geq 0 \text{ and } k-\tau \geq 0 \\ \left.\begin{array}{l} t = [1-k, N] \\ s = [1 : N-k+\tau] \end{array}\right\} , & k < 0 \text{ and } k-\tau \geq 0 \\ \left.\begin{array}{l} t = [1, N-\tau] \\ s = [1-k+\tau : N] \end{array}\right\} , & k \geq 0 \text{ and } k-\tau < 0 \\ \left.\begin{array}{l} t = [1-k, N] \\ s = [1-k+\tau : N] \end{array}\right\} , & k < 0 \text{ and } k-\tau < 0 \end{array}\right. . \tag{20}$$

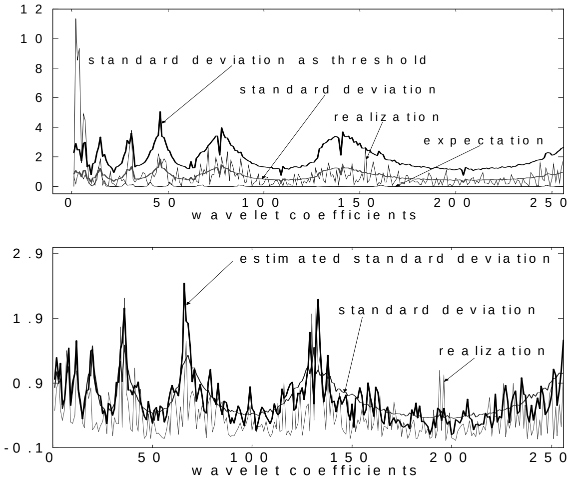


Fig. 3. Top: *Wavelet coefficients (coifman-5) of a realization of an estimator $\hat{\Phi}_{yu}(f)$, with expectation equal to the coefficients of the true function $\Phi_{yu}(f)$. An efficient threshold is based on scaling the standard deviation. The standard deviation is calculated from a Monte Carlo simulation of* 1000 *realizations of $\hat{\Phi}_{yu}(f_l)$ based on a white noise and input signal of length $N = 512$ with variance $\sigma_u^2 = 1$ and a noise term $v(t)$ additive to the output with variance of $\sigma_v^2 = 0.0025$.* Bottom: *Estimation of the standard deviation based on a single realization using expressions* (18) *and* (19) .

Expressions (18) and (19) are exact even for finite data sets. By a simple Discrete Fourier transformation and appropriate downsampling the variance in the wavelet coefficients is obtained.

### *B. 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 $\hat{C}_{uy}(k)$ with expression (5) for the occurring covariance functions in expression (19). 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. Therefore, the estimate follows with

$$\begin{aligned} & \widehat{cov}\left[\hat{C}_{yu}(k), \hat{C}_{yu}(k-\tau)\right] \\ = \; & \frac{1}{N^2}\sum_t\sum_s \left\{\hat{C}_{yy}(s-t)\hat{C}_{\mathrm{u}}(s-t-\tau)\right. \\ & \left. + \hat{C}_{yu}(s-t+k-\tau)\hat{C}_{uy}(s-t-k)\right\}. \end{aligned} \tag{21}$$

Figure 3 depicts an example, based on the test system found in [3], of an estimation of the standard deviation in the wavelet domain based on one realization of the input and output data and the use of expressions (18) and (21). As an estimate of the standard deviation itself the estimator is naturally quite poor since it is based on one realization only. Although following the peaks, the estimator is too erratic to closely approximate the true (Monte Carlo) standard deviation. However, the features of the particular realization of the estimation error in the wavelet coefficients on which the estimation is based are followed remarkably well by the estimation of the standard deviation. 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

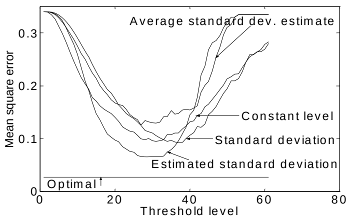


Fig. 4. *Mean square error for thresholding wavelet coefficients of an estimator $\hat{\Phi}_{yu}(f_l)$ with constant threshold, standard deviation, an estimate of the average standard deviation and with the estimated standard deviation. For comparison the theoretical minimum, achieved with $\delta_{j,k} = \frac{|d_j(k)|^2}{|d_j(k)|^2+\sigma^2}$, is given.*

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 4) .

The method is well capable of properly identifying those coefficients predominantly containing estimation errors. In fact, with expressions (18) and (21) it is the data of one realization itself which provides this information. However, the present implementation of expression (21) is such that only data sets upto $N = 512$ are feasible. It is for these small data sets that often the $\hat{G}_{ETFE}(f)$ to be smoothed does not contain sufficient information. Figure 5 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.

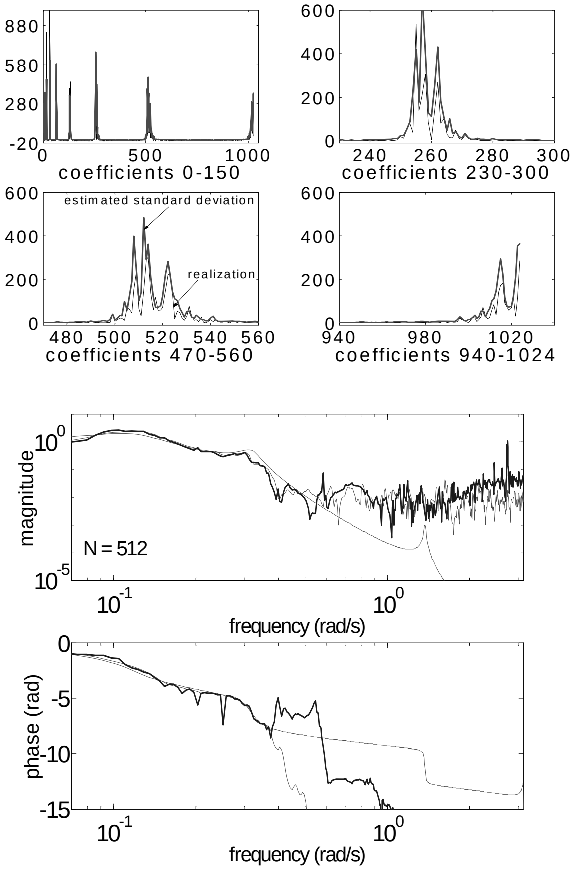


Fig. 5. *Smoothing of an empirical transfer function estimate of a rotating drive system ($N = 512$) with the optimal Hamming window versus smoothing with the proposed denoising method. As reference a sine-swept model is given. On top are depicted the wavelet coefficients of the associated $\hat{\Phi}_{yu}(f)$ with the estimated standard deviation used as threshold.*

## VIII. Conclusions

Wavelet denoising provides for an effective variance reduction in the empirical transfer function estimate, when based on the consecutive thresholding of the wavelet 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. The absence of a theoretical expression for the variance in $\hat{G}_{ETFE}(f)$ for finite stochastic data sets prevents the direct application of the wavelet denoising scheme to $\hat{G}_{ETFE}(f)$. 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.

## 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** (1992), no. 2, 766–784.

[2] A. Benveniste, A. Juditsky, B. Deylon, Q. Zhang, and P-Y. Glorennec, *Wavelets in identification*, Proceedings of 10th IFAC Symposium on System Identification, SYSID '94 **2** (1994), 27–48.

[3] P. Bodin, *On the selection of best orthonormal basis in system identification and signal processing*, Ph.D. thesis, Royal Institute of Technology, Department of Signals, Sensors and Systems, Stockholm, Sweden, 1997.

[4] P.M.T. Broersen, *A comparison of transfer function estimators*, IEEE Transactions on Instrumentation and Measurement **44** (1995), no. 3, 657–661.

[5] C.S. Burrus, R.A. Gopinath, and H. Guo, *Introduction to wavelets and wavelet transform. a primer*, Prentice-Hall, New Yersey, USA, 1998.

[6] K.C Chou, A.S. Willsky, and Benveniste A., *Multiscale recursive estimation, data fusion, and regularization*, IEEE Transactions on Automatic Control **39** (1994), no. 3, 464–477.

[7] D.L. Donoho, *De-noising by soft-thresholding*, IEEE Transactions on Information Theory **41** (1995), no. 3, 613–627.

[8] G. M. Jenkins and D. W Watts, *Spectral analysis and its applications*, Holden-Day, San Francisco, California, 1968.

[9] G. Kaiser, *A friendly guide to wavelets*, Birkhauser, Boston, USA, 1994.

[10] L. Ljung, *System identification: Theory for the user*, P T R Prentice Hall, Englewood Cliffs, New Yersey, USA, 1987.

[11] Y. Meyer, *Wavelets. algorithms and applications*, Society for Industrial and Applied Mathematics, Philadelphia, USA, 1993, Translated and Revised by Ryan, R.D.

[12] P. Moulin, *Wavelet thresholding techniques for power spectrum estimation*, IEEE Transactions on Signal Processing **42** (1994), no. 11, 3126–3136.

[13] M.B. Priestley, *Spectral analysis and time series. univariate series*, vol. 1, Academic Press, London, UK, 1981.

[14] A.T Walden, D.B. Percival, and E.J. McCoy, *Spectrum estimation by wavelet thresholding of multitaper estimators*, Tech. report, Imperial College of Science, Technology and Medicine, London, UK, 1997, http://www.ma.ic.ac.uk.