Wavelet Thresholding: Beyond the Gaussian I.I.D. Situation

Michael H. Neumann[1] *and Rainer von Sachs*[2]

[1] Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany
[2] Department of Mathematics, University of Kaiserslautern, Germany

Abstract

With this article we first like to a give a brief review on wavelet thresholding methods in non-Gaussian and non-i.i.d. situations, respectively. Many of these applications are based on *Gaussian approximations* of the empirical coefficients. For regression and density estimation with independent observations, we establish joint asymptotic normality of the empirical coefficients by means of strong approximations. Then we describe how one can prove asymptotic normality under mixing conditions on the observations by cumulant techniques.
In the second part, we apply these non-linear adaptive shrinking schemes to spectral estimation problems for both a stationary and a non-stationary time series setup. For the latter one, in a model of Dahlhaus ([Da93]) on the *evolutionary spectrum* of a *locally stationary* time series, we present two different approaches. Moreover, we show that in classes of anisotropic function spaces an appropriately chosen wavelet basis automatically adapts to possibly different degrees of regularity for the different directions. The resulting fully-adaptive spectral estimator attains the rate that is optimal in the idealized Gaussian white noise model up to a logarithmic factor.

1 Introduction

Nonlinear wavelet methods have become very popular since the seminal papers by Donoho and Johnstone (see [DJ92], [DJ94], e.g.). They rediscovered a phenomenon, originally detected by Nemirovskii, Polyak and Tsybakov in [NTP85] and [N85] for Sobolev smoothness classes, that linear estimators are unable to attain the optimal uniform convergence rate in balls of Besov spaces $B^m_{p,q}$ with $p < 2$. Moreover, they showed that appropriately thresholded wavelet estimators attain the minimax bound up to a small constant. Whereas the nonlinear estimators proposed in [NTP85] and [N85] did not lead to new developments beyond the Gaussian white noise model, the work by Donoho and Johnstone has led to a real breakthrough in statistical estimation theory.

This new method can be described as follows. Assume we have function-valued observations in the Gaussian white noise model

$$dY(t) \;=\; f(t)\,dt \;+\; \epsilon\,dW(t), \quad t \in [0,1], \tag{1}$$

where $W(t)$ is a standard Wiener process, and where the noise level ϵ is assumed to be small. There are two basic requirements, which are essential for the advantages of these new methods over traditional linear estimators. First, we have an orthonormal system $\{\phi_{lk}\}_{k\in I_l^0} \cup \{\psi_{jk}\}_{j\ge l;k\in I_j}$ of basis functions, which are essentially generated by dilations and translations of a scaling function ϕ and a so-called wavelet ψ. Here the index sets I_l^0 and I_j are of cardinality $O(2^l)$ and 2^j, respectively. These basis functions, which are both localized in spatial position and frequency, are capable of optimally compressing functions both under more traditional "homogeneous" smoothness assumptions like Hölder or L_2-Sobolev as well as in the case of "inhomogeneous" smoothness like L_p-Sobolev or Besov $B^m_{p,q}$ with $p < 2$. Having in mind the Fourier series expansion $f = \sum \alpha_k\phi_{lk} + \sum \alpha_{jk}\psi_{jk}$, one calculates first empirical versions $\widetilde{\alpha}_k = \int \phi_{lk}(t)\,dY(t)$ and $\widetilde{\alpha}_{jk} = \int \psi_{jk}(t)\,dY(t)$ of the wavelet coefficients. These empirical coefficients are again independently normally distributed with homogeneous variances ϵ^2.

In the case of "homogeneous" smoothness classes $\mathcal{F}$, the coefficients within an arbitrary scale are roughly of the same order of magnitude. Strictly speaking, the functional $\sup_{f\in\mathcal{F}}\left\{\sum_k \min\{\epsilon^2, \alpha_{jk}^2\}\right\}$ is not essentially smaller than $\sup_{f\in\mathcal{F}}\left\{\min\{2^j\epsilon^2, \sum_k \alpha_{jk}^2\}\right\}$. Hence, we do not lose very much by a levelwise inclusion/exclusion strategy, i.e. by either taking all empirical coefficients from a certain resolution level j or by neglecting the whole level. This leads to linear estimation rules, which are known to be able to provide the optimal rate of convergence.

In contrast, in "inhomogeneous" smoothness classes, the coefficients within a given scale are allowed to be much less homogeneous. For certain j depending on ϵ, the quantity $\sup_{f\in\mathcal{F}}\left\{\sum_k \min\{\epsilon^2, \alpha_{jk}^2\}\right\}$ will be essentially smaller than $\sup_{f\in\mathcal{F}}\left\{\min\{2^j\epsilon^2, \sum_k \alpha_{jk}^2\}\right\}$. This calls for a coordinatewise decision rule whether or not an empirical coefficient should be included in the estimator of f. Donoho and Johnstone proposed in [DJ92] two nonlinear rules to treat the empirical coefficients:

1) hard thresholding
$$\delta^{(h)}(\widetilde{\alpha}_{jk}, \lambda) \;=\; \widetilde{\alpha}_{jk} I\left(|\widetilde{\alpha}_{jk}| \ge \lambda\right)$$

and

2) soft thresholding
$$\delta^{(s)}(\widetilde{\alpha}_{jk}, \lambda) \;=\; \left(|\widetilde{\alpha}_{jk}| - \lambda\right)_+ \operatorname{sgn}(\widetilde{\alpha}_{jk}).$$

In the following $\delta^{(\cdot)}$ is used to (somewhat sloppily) denote either $\delta^{(h)}$ or $\delta^{(s)}$.

[DJ92] and [DJKP95] contain two main results, which characterize the performance of thresholded wavelet estimators in model (1). If the thresholds $\lambda = \lambda_j(\mathcal{F}, \epsilon)$ are chosen in a level-dependent optimal way, then the estimator

$$\widehat{f} \;=\; \sum_k \widetilde{\alpha}_k\,\phi_{lk} \;+\; \sum_{j\ge l}\sum_k \delta^{(\cdot)}(\widetilde{\alpha}_{jk}, \lambda_j)\,\psi_{jk}$$

is minimax up to a multiplicative constant, if $\mathcal{F}$ is a ball in a certain Besov space. However, the optimal choice of the thresholds λ_j depends on the set $\mathcal{F}$ and the noise level ϵ. In simple estimation problems there exist some proposals to choose these λ_j automatically.

Assuming Gaussian errors, in [DJ93] Donoho and Johnstone proposed to use Stein's Unbiased Risk Estimate (SURE) in conjunction with sample splitting at the higher resolution levels, which was further investigated by Nason in [Na94]. In view of approximation results connecting non-Gaussian regression with model (1), as described e.g. in the present paper, we conjecture that SURE can also be applied in non-Gaussian regression. The application of the classical leave-one-out cross-validation was investigated by Neumann and Spokoiny in [NS95].

A much simpler approach is to set a uniform threshold. Assume that n coefficients are to be thresholded. It is shown in [DJKP95] that by the choice

$$\lambda = \epsilon\sqrt{2\log(n)}$$

we obtain an estimator, which is *simultaneously minimax* in a wide range of smoothness classes up to a logarithmic factor. Note that by using such a uniform threshold there is no intention for an optimal trade-off of bias and variance of the resulting estimator. This choice is rather motivated by significantly testing what is the information content of the noisy wavelet coefficients.

These results are undoubtly of considerable theoretical interest, however it is of great importance to know how these methods can be transferred to practically more relevant estimation problems. Perhaps the first practical application of the shrinkage scheme developed in [DJ92] was to density estimation by Johnstone, Kerkyacharian and Picard in [JKP92]. There, the authors proposed nonrandom thresholds $\lambda_j = K\sqrt{j/n}$, which were motivated by considerations on the tail behaviour of the empirical coefficients $\tilde{\alpha}_{jk} = n^{-1}\sum_i \psi_{jk}(X_i)$, where $X_i \sim f$ are independent. They also indicated that soft thresholds originally developed for the case of Gaussian noise can be applied as well.

Patil in [P94] for hazard rate estimation and Gao in [G93] for spectral density estimation of a stationary Gaussian time series also developed specific threshold rules on the basis of upper estimates of the tail probabilities of appropriate empirical coefficients.

In [DeJ93] Delyon and Juditskii extended these results to general estimation problems, which allow to consider empirical coefficients with a certain tail behavior. They proposed specific threshold choices in dependence on the decay of probabilities of large deviations of the empirical coefficients. These results can be applied to a large number of curve estimation problems, and imply that in many cases the same rates for the risk as in model (1) are attainable.

The abovementioned papers provide certainly reasonable rules for the choice of the thresholds in particular estimation problems. However, a careful analysis of all these problems (density estimation, non-Gaussian regression, spectral density estimation, density estimation and regression with dependent observations) shows that, under weak regularity conditions, all these problems are essentially of the same nature. It turns out that the empirical coefficients at the scales j with $2^j \ll n$ are asymptotically Gaussian. Obviously, a simple central limit theorem would not be sufficient for proving risk equivalence of thresholded estimators in non-Gaussian models to the case of Gaussian noise. However, in many situations one can state asymptotic normality by showing that

$$P\left(\pm(\tilde{\alpha}_{jk} - \alpha_{jk})/\sigma_{jk} \geq x\right) = (1 - \Phi(x))\,(1 + o(1)) \tag{2}$$

holds uniformly in $(j,k) \in \mathcal{J}_n$, $-\infty \leq x \leq \Delta_n$, where $\sigma_{jk}^2 := \operatorname{var}(\tilde{\alpha}_{jk}) = O(n^{-1})$, $\Delta_n \asymp n^{\mu}$ for some $\mu > 0$ and $\mathcal{J}_n$ such that

$$\sum_{(j,k)\notin\mathcal{J}_n} \alpha_{jk}^2 = O(n^{-1}). \tag{3}$$

(Usually the restriction $\mathcal{J}_n = \{(j,k) \mid 2^j \leq Cn^{1-\delta}\}$ for some $\delta > 0$ will be appropriate.)

Now we can define an accompanying Gaussian model as

$$\xi_{jk} = \alpha_{jk} + \varepsilon_{jk}, \quad (j,k) \in \mathcal{J}_n, \tag{4}$$

where $\varepsilon_{jk} \sim N(0, \sigma_{jk}^2)$. Essentially by integration by parts, we then can show that

$$\sum_{(j,k)\in\mathcal{J}_n} E\left(\delta^{(\cdot)}(\tilde{\alpha}_{jk}, \lambda_{jk}) - \alpha_{jk}\right)^2 = (1 + o(1)) \sum_{(j,k)\in\mathcal{J}_n} E\left(\delta^{(\cdot)}(\xi_{jk}, \lambda_{jk}) - \alpha_{jk}\right)^2 + O(n^{-1}). \tag{5}$$

For the sake of greater generality, here we allow the thresholds to depend on n, j and k, although in practice one often uses thresholds only depending on the sample size n and/or the resolution scale j. Location dependent thresholds being proportional to σ_{jk} are reasonable in the case of heteroscedastic error models.

Once we have shown (3) and (5), the application of thresholding methods developed for the case of Gaussian noise is justified. In particular, if we assume that the unknown curve is a member of some smoothness class $\mathcal{F}$ with degree of smoothness m, we are able to attain the "classical" rate $n^{-2m/(2m+1)}$ for the L_2-risk by exactly the same treatment of the empirical coefficients as in the Gaussian case. Assume we have any one-dimensional curve estimation problem, which allows to define empirical coefficients satisfying (2). We consider the estimator

$$\widehat{f} = \sum_k \tilde{\alpha}_k \phi_{lk} + \sum_{(j,k)\in\mathcal{J}_n} \delta^{(\cdot)}(\tilde{\alpha}_{jk}, \lambda_j)\psi_{jk}.$$

Then we immediately obtain the following theorem from known results in Gaussian regression.

Theorem 1.1 *Let $\mathcal{F}$ be a ball in a Besov space $B_{p,q}^m$, $m,p,q \geq 1$. Assume that (2) and (3) are satisfied uniformly in $(j,k) \in \mathcal{J}_n$ and $f \in \mathcal{F}$. Then,*

(i) for an optimal choice of the thresholds $\lambda_j = \lambda(n,j,\mathcal{F})$,

$$\sup_{f\in\mathcal{F}} \left\{\|\widehat{f} - f\|_{L_2}^2\right\} = O\left(n^{-2m/(2m+1)}\right),$$

(ii) for thresholds λ_{jk} satisfying $\sigma_{jk}\sqrt{2\log(\#\mathcal{J}_n)} \leq \lambda_{jk} \leq C\ n^{-1/2}\sqrt{\log(n)}$ for any positive constant C,

$$\sup_{f\in\mathcal{F}} \left\{\|\widehat{f} - f\|_{L_2}^2\right\} = O\left((\log(n)/n)^{-2m/(2m+1)}\right).$$

This approach has been used by Neumann and Spokoiny, [NS95], in non-Gaussian regression, by Neumann, [Ne94], in spectral density estimation for stationary time series, by Neumann and von Sachs, [NvS94], in spectral density estimation in the locally stationary case and by Dahlhaus, Neumann and von Sachs, [DNvS95], for the estimation of the (time varying) autoregression coefficients in a nonstationary model.

In the following section we indicate how to derive the basis approximation (2) in regression and density estimation with dependent observations. We also describe an alternative approach: By means of strong approximations, a large set of empirical coefficients is *simultaneously* approximated by those in a certain Gaussian observation model. The third section deals with the application of the prescribed methods and techniques to the spectral density estimation problem. First we emphasize that this problem is (asymptotically) analogous to the regression problem but without knowledge of the exact underlying error distribution and dependence structure of the errors. Then, we present the main techniques how to derive the necessary approximations for the stationary time series set up: we distinguish between the case of a Gaussian series, which has been treated in [G93], and a general (i.e. non-Gaussian) series with techniques as derived by [Ne94].

These techniques give us the basis to also investigate the instationary time series setup: In a model of a *locally stationary* process with time-varying (*evolutionary*) spectrum, as derived by [Da93], we address a particular 2-d problem with the same features as for the 1-d stationary situation but with an additional (time) dimension - with additional (unknown) dependence structure.

Moreover, far more generally, we like to treat our problem by embedding it into n-d (anisotropic) function estimation problems: in [NvS94] a minimax result has been derived for a typical (2-d) function class, an anisotropic Sobolev class, with different degree of smoothness in different directions. We show that this minimax rate is attained if we base our threshold estimator on an appropriate wavelet basis, the so-called tensor product basis. With this, we end up to show that a certain wavelet estimator of the evolutionary spectrum, under minimal regularity assumptions, nearly attains this rate in a ball within this Sobolev class. This estimator, derived in [NvS94] arises as a modification of a first approach of [vSS94], which turns out to be able to adapt to the different degrees of smoothness of the time-dependent spectrum in the time-frequency plane.

Finally some simulation examples on a time-dependent spectrum, which shows some typical local features, confirm what has been derived by our asymptotic results.

2 Asymptotic normality in regression and density estimation

In this section we consider nonparametric regression and density estimation and establish connections to the idealized Gaussian model (1). In the first part we assume independent observations and also indicate an alternative way to show the equivalence to the Gaussian case via strong approximations. In the second part we drop the sometimes unrealistic assumption of independence and focus on the derivation of (2) in cases of weak dependence.

2.1 Regression and density estimation with independent observations

We start with the regression case, where we assume to observe Y_i from the model

$$Y_i \;=\; m(x_i) \,+\, \varepsilon_i, \quad i = 1, \ldots, n \tag{6}$$

with $E\varepsilon_i = 0$. For simplicity we assume that the x_i's are taken from a triangular array of regularly spaced points $x_i = x_i(n)$, i.e.

(A1) $\int_0^{x_i} g(t)\,dt \;=\; i/n$ for some positive, continuous density g.

The errors ε_i are independently, not necessarily identically distributed with

(A2) $C_1 \;\leq\; \sigma_i^2\,; = \,\mathrm{var}(\varepsilon_i) \;\leq\; C_2$.

Again for simplicity, we assume that all moments are uniformly bounded, i.e.

(A3) $E|\varepsilon_i|^M \;\leq\; C_M$ for all $M \in \mathbb{N}$ and for some fixed constants $C_M < \infty$.

Further, we assume

(A4) ϕ and ψ are of bounded total variation on $[0,1]$.

As a potential asymptotic approximation to (6) we consider the model

$$Z_i \;=\; m(x_i) \,+\, \xi_i, \quad i = 1, \ldots, n, \tag{7}$$

where the ξ_i's are independent with $\xi \sim N(0, \sigma_i^2)$. For simplicity of notation we define $\psi_{l-1,k} = \phi_{lk}$ and $\alpha_{l-1,k} = \alpha_k$.

We consider the empirical coefficients

$$\widetilde{\alpha}_{jk} \;=\; \sum_i (x_i - x_{i-1}) w_{jk}(i) Y_i$$

and

$$\widetilde{\theta}_{jk} \;=\; \sum_i (x_i - x_{i-1}) w_{jk}(i) Z_i,$$

where

$$w_{jk}(i) \;=\; \int_{x_{i-1}}^{x_i} \psi_{jk}(t)\,dt/(x_i - x_{i-1}).$$

To obtain a connection between $\widetilde{\alpha}_{jk}$ and $\widetilde{\theta}_{jk}$, we consider the partial sum processes

$$S_j \;=\; \sum_{i \leq j} \varepsilon_i (x_i - x_{i-1})$$

and

$$T_j \;=\; \sum_{i \leq j} \xi_i (x_i - x_{i-1}).$$

Under (A1) through (A3), we infer by Corollary 4 in [S91] that there exists a joint probability space on which

$$P\left(\max_{1 \leq i \leq n} \{|S_i \,-\, T_i|\} > n^{\gamma - 1} \right) \;=\; O(n^{-\lambda}) \tag{8}$$

holds for arbitrarily small $\gamma > 0$ and arbitrarily large $\lambda < \infty$. By (A4) we get that

$$\sum_{i=1}^{n} |w_{jk}(i) - w_{jk}(i+1)| + w_{jk}(n) = O(2^{j/2}),$$

which yields by summation by parts that

$$\begin{aligned} |\widetilde{\alpha}_{jk} - \widetilde{\theta}_{jk}| &= |w_{jk}(i)(\varepsilon_i - \xi_i)(x_i - x_{i-1})| \\ &= \left| \sum_{i=1}^{n-1} (w_{jk}(i) - w_{jk}(i+1))(S_i - T_i) + w_{jk}(n)(S_n - T_n) \right| \\ &= O\left(2^{j/2} \max_{1 \le i \le n} \{|S_i - T_i|\} \right). \end{aligned}$$

This implies the following lemma.

Lemma 2.1 *Assume (A1) through (A4). Then, on an appropriate probability space,*

$$P\left(|\widetilde{\alpha}_{jk} - \widetilde{\theta}_{jk}| > n^{\gamma - \delta/2} n^{-1/2} \quad \text{for any} \quad (j,k) \in \mathcal{J}_n \right) = O(n^{-\lambda}) \tag{9}$$

for any $\gamma > 0$, $\lambda < \infty$, *where* $\mathcal{J}_n = \{(j,k) \mid 2^j \le n^{1-\delta}\}$ *for some* $\delta > 0$.

If γ is chosen such that $\gamma < \delta/2$, then the maximal difference between $\widetilde{\alpha}_{jk}$ and $\widetilde{\theta}_{jk}$ is of smaller order than the noise level $n^{-1/2}$. This result can be used in particular to prove (2). But there are also other interesting applications of (9). Due to the possible reduction to the Gaussian case, we can immediately derive the asymptotics of thresholded estimators in the L_∞-norm from known results in the Gaussian case. Further, we can also use (9) to derive tests against the hypothesis $H_0: \ m \equiv const$, which in particular would mean that $\alpha_{jk} = 0$ for all (j,k), $j \ge l$. This has been done in a recent work by Fan in [F94]. Moreover, we can apply (9) to derive asymptotic confidence sets.

If we replace assumptions (A1) and (A2) by

(A1') $x_i = i/n$,

(A2') $\mathrm{var}(\varepsilon_i) \equiv \sigma^2 \in (0, \infty)$,

and additionally assume

(A5) $TV_{[0,1]}(m) \le C$,

then we can also establish the asymptotic equivalence to the (homogeneous) Gaussian white noise model (1). This is in accordance to equivalence results in a decision theoretic framework by Brown and Low in [BL92]. Let

$$\widetilde{\widetilde{\theta}}_{jk} = \int \psi_{jk}(t)\, dY(t)$$

be defined by the observations from model (1).

Lemma 2.2 *Assume (A1'), (A2') and (A3) through (A5). Then there exists a joint probability space such that*

$$P\left(|\widetilde{\alpha}_{jk} - \widetilde{\widetilde{\theta}}_{jk}| > n^{\gamma-\delta/2}n^{-1/2} \quad \textit{for any} \quad (j,k) \in \mathcal{J}_n\right) = O(n^{-\lambda}) \tag{10}$$

for any $\gamma > 0$, $\lambda < \infty$.

Proof. Instead of $\{S_j\}$ we consider the linearly interpolated partial sum process on [0,1]:

$$S(t) = \sum_{j \le nt} \varepsilon_j + (nt - [nt])\varepsilon_{[nt]+1}.$$

Again by Corollary 4 in [S91], we obtain that, on an appropriate probability space,

$$P\left(\sup_{0\le t\le 1}\left\{|S(t) - \sigma n^{1/2}W(t)|\right\} > n^{\gamma}\right) = O(n^{-\lambda}) \tag{11}$$

holds for arbitrary $\gamma > 0$, $\lambda < \infty$. Further, we conclude by (A5) that

$$|\Delta_{jk}| = \left|E\widetilde{\alpha}_{jk} - E\widetilde{\widetilde{\theta}}_{jk}\right| \le \sum_i \int_{x_{i-1}}^{x_i} |\psi_{jk}(t)||m(x_i) - m(t)|\,dt = O(2^{j/2}n^{-1}) \tag{12}$$

holds uniformly in $(j,k) \in \mathcal{J}_n$.

By integration by parts we obtain

$$\begin{aligned}
\widetilde{\alpha}_{jk} - \widetilde{\widetilde{\theta}}_{jk} &= \Delta_{jk} + n^{-1}\int \psi_{jk}(t)\,dS(t) - \sigma n^{-1/2}\int \psi_{jk}(t)\,dW(t) \\
&= \Delta_{jk} + n^{-1}\int (S(t) - \sigma n^{1/2}W(t))\,d\psi_{jk}(t) + n^{-1}(S(1) - \sigma n^{1/2}W(1))\psi_{jk}(1) \\
&= O\left(2^{j/2}n^{-1} + 2^{j/2}n^{-1}\sup_{0\le t\le 1}\{|S(t) - \sigma n^{1/2}W(t)|\}\right),
\end{aligned}$$

which proves the lemma in conjunction with (11) and (12).

The approximation by the continuous model (1) has certain advantages over an approximation by the discrete model (7). The coefficients $\widetilde{\widetilde{\theta}}_{jk}$ are exactly unbiased and independently distributed with homogeneous variances, which makes their analysis easier.

Now we turn to the problem of density estimation. We assume that we have independent observations $X_1, \ldots, X_n$, identically distributed with density f supported on $[0,1]$ and cumulative distribution function F.

With $\widetilde{X}_i = F(X_i) \sim \text{uniform}[0,1]$ we define the empirical process

$$U_n(t) = n^{-1/2}\sum_{i=1}^{n}\left(I(\widetilde{X}_i \le t) - t\right). \tag{13}$$

By Theorem 2 in Section 12.1 of [SW86] we infer that, on an appropriate probability space,

$$P\left(\|U_n - B_n\|_\infty > n^{-1/2}(c_1\log(n) + x)\right) \le c_2\exp(-c_3 x) \tag{14}$$

holds for appropriate c_1, c_2 and c_3, where $\{B_n\}$ is a sequence of Brownian bridges. Since

$$U_n(F(s)) \;=\; n^{1/2}(F_n(s) \;-\; F(s)),$$

where F_n is the empirical c.d.f. of the sample $X_1, \ldots, X_n$, we obtain

$$P\left(\sup_{0\le s\le 1}\left\{|(F_n(s) \;-\; F(s) \;-\; n^{-1/2}B_n(F(s))|\right\} > n^{-1}(c_1\log(n)+x)\right) \;\le\; c_2\exp(-c_3 x). \tag{15}$$

Now we are in the position to establish the connection to the Gaussian case for the empirical coefficients. Let

$$\tilde{\alpha}_{jk} \;=\; n^{-1}\sum_i \psi_{jk}(X_i).$$

Then

$$\begin{aligned}\tilde{\alpha}_{jk} \;-\; \alpha_{jk} \;&=\; \int \psi_{jk}(t)\, d(F_n(t) - F(t)) \\ &=\; n^{-1/2}\int \psi_{jk}(t)\, dB_n(F(t)) \\ &\quad - \int \left[(F_n(t) - F(t)) \;-\; n^{-1/2}B_n(F(t))\right]\, d\psi_{jk}(t).\end{aligned} \tag{16}$$

By (15) we see that the second term on the right-hand side of (16) is $O(n^{\gamma-1}2^{j/2})$ for any fixed $\gamma > 0$ with overwhelming probability. With $B_n(t) = W(t) - tW(1)$ (neglecting the index n), the first term on the right-hand side writes as

$$\begin{aligned} n^{-1/2}\int \psi_{jk}(t)\, dB_n(F(t)) \;&=\; n^{-1/2}\int \psi_{jk}(t)\, dW(F(t)) \;-\; n^{-1/2}W(1)\int \psi_{jk}(t)\, dF(t) \\ &\sim\; n^{-1/2}\int \psi_{jk}(t)f^{1/2}(t)\, dW(t) \;-\; n^{-1/2}W(1)\alpha_{jk}. \end{aligned} \tag{17}$$

Because of $\sum_{j,k}\alpha_{jk}^2 \;=\; \int |f(t)|^2\, dt \;<\; \infty$, the second term on the right-hand side of (17) will be negligible for most of the (j,k)'s. Moreover, under appropriate smoothness assumptions,

$$\sum_k \alpha_{jk}^2 \;\le\; C(j) \longrightarrow 0 \quad \text{as } j \to \infty,$$

which implies that $\tilde{\alpha}_{jk} - \alpha_{jk}$ can be well approximated by $n^{-1/2}\int \psi_{jk}(t)f^{1/2}(t)\, dW(t)$ for $(j,k) \in \left\{(j,k) \mid\; n^{\delta} \le 2^j \le n^{1-\delta}\right\}$ and any $\delta > 0$. This reflects the well-known fact that, in a shrinking neighborhood of some density f_0, density estimation is asymptotically equivalent to the statistical experiment given by observations

$$dY(t) \;=\; f(t)dt \;+\; n^{-1/2}f_0^{1/2}(t)\, dW(t),$$

cf. [Nu94].

2.2 Regression and density estimation for dependent observations

Again we begin with regression, i.e. we assume (6) and only drop the assumption of independence of the ε_i's. In [Bri94] Brillinger assumes a stationary error process $\{\varepsilon_i\}$ and shows asymptotic normality of the empirical coefficients under the assumption that

$$\sum_m C_m z^m/m! \;<\; \infty \tag{18}$$

holds for $|z|$ small enough, where

$$C_m \;=\; \sum_{u_1,\ldots,u_{m-1}} |c_m(u_1,\ldots,u_{m-1})|$$

and

$$c_m(u_1,\ldots,u_{m-1}) \;=\; \operatorname{cum}(\varepsilon_{i+u_1},\ldots,\varepsilon_{i+u_{m-1}},\varepsilon_i).$$

He remarks that (18) is satisfied in particular if the error process $\{\varepsilon_i\}$ is Gaussian with $C_2 < \infty$.

However, to cover both usual mixing conditions and non-Gaussian processes, we need a weaker substitute of (18). Assume that

(A6) $E|\varepsilon_i|^p \;\leq\; C^p(p!)^\kappa \quad \forall p \geq 2$

holds for appropriate $C < \infty$ and $\kappa \geq 0$. Note that many of the distributions that can be found in textbooks satisfy (A6) for an appropriate choice of κ. In [JK70] we can find closed forms of higher order cumulants of the exponential, gamma, inverse Gaussian and F-distribution, which show that this condition is satisfied for $\kappa = 1$.

Further, we assume that

(A7) $\{\varepsilon_i\}$ is α-mixing with $\alpha(s) \leq C\exp(-b|s|)$.

Then we can show by 2) of Theorem 3 in [SJ88] that

$$\sup_{1\leq i_1\leq\infty} \left\{ \sum_{i_2,\ldots,i_p=1}^{\infty} |\operatorname{cum}(\varepsilon_{i_1},\ldots,\varepsilon_{i_p})| \right\} \;\leq\; \widetilde{C}^p(p!)^{3+\kappa} \tag{19}$$

holds for all $p = 2,3,\ldots$ and appropriate $\widetilde{C}$, cf. Remark 3.1 in [Ne94]. Let

$$\tilde{\alpha}_{jk} \;=\; \sum_i \tilde{w}_{jk}(i)Y_i$$

with $\tilde{w}_{jk}(i) = \int_{x_{i-1}}^{x_i} \psi_{jk}(t)\,dt$. Then we obtain that

$$\begin{aligned}
\operatorname{cum}_p(\tilde{\alpha}_{jk}) \;&=\; \operatorname{cum}_p(\sum_i \tilde{w}_{jk}(i)\varepsilon_i) \\
&=\; \sum_{i_1}\sum_{i_2,\ldots,i_p} \tilde{w}_{jk}(i_1)\cdots\tilde{w}_{jk}(i_p)\operatorname{cum}(\varepsilon_{i_1},\ldots,\varepsilon_{i_p}) \\
&=\; O\left(n^{-1}\,(n^{-1}2^{j/2})^{p-2}\,\widetilde{C}^p(p!)^{3+\kappa}\right)
\end{aligned}$$

holds uniformly in $p \geq 2$.

If $\sigma_{jk} = \operatorname{var}(\tilde{\alpha}_{jk}) \geq Cn^{-1}$ for some $C > 0$, we have

$$|\operatorname{cum}_p(\tilde{\alpha}_{jk})/\sigma_{jk})| \;=\; O\left(C^p(p!)^{3+\kappa}\,(2^j n^{-1})^{(p-2)/2}\right) \tag{20}$$

uniformly in $p \geq 2$. This implies by Lemma 1 in [RSS78] that

$$P\left(\pm(\tilde{\alpha}_{jk} \,-\, E\tilde{\alpha}_{jk})/\sigma_{jk} \geq x\right) \;=\; (1 \,-\, \Phi(x))(1 \,+\, o(1)) \tag{21}$$

holds uniformly in those $(j,k) \in \mathcal{J}_n$ with $\sigma_{jk} \geq Cn^{-1}$, and uniformly on some interval $-\infty < x \leq \Delta_n$, $\Delta_n \asymp n^{\mu}$ for some $\mu > 0$.

In cases where σ_{jk} is less than $n^{-1/2}$, we first consider the theoretical quantities

$$\tilde{\tilde{\alpha}}_{jk} \;=\; \tilde{\alpha}_{jk} \,+\, \vartheta_{jk},$$

where $\vartheta_{jk} \sim N(0, n^{-1} - \sigma_{jk}^2)$ is independent of $\tilde{\alpha}_{jk}$. The new random variable $\tilde{\tilde{\alpha}}_{jk}$ has a variance equal to n^{-1}, and its higher order cumulants coincide with those of $\tilde{\alpha}_{jk}$. Hence, we can establish (21) for $\tilde{\tilde{\alpha}}_{jk}$ instead of $\tilde{\alpha}_{jk}$. On the other hand, we can easily show for thresholded estimators that

$$E\left(\delta^{(\cdot)}(\tilde{\alpha}_{jk},\lambda) \,-\, \alpha_{jk}\right)^2 \;\leq\; 2E\left(\delta^{(\cdot)}(\tilde{\tilde{\alpha}}_{jk},\lambda) \,-\, \alpha_{jk}\right)^2$$

holds, which gives us again the connection to a certain Gaussian model.

The case of density estimation is quite similar. We assume that

(A8) $\{X_i\}$ is ϕ-mixing with $\phi(s) \leq C\exp(-b|s|)$.

By 1b) of Theorem 5 in [SJ88], with $\beta = \delta = 1$, we obtain that

$$|\operatorname{cum}_p(\tilde{\alpha}_{jk})| \;=\; p!\, 8^{p-1}\,(Cn^{-1}2^{j/2})^{p-2}\, \Lambda_n^{p-2}(\phi, 4(p-2))\;\; n\,\Lambda_n(\phi,4)\, E|\psi_{jk}(X_i)|^2,$$

where

$$\Lambda_n(\phi,u) \;=\; \max\left\{1, \max_{1\leq s\leq n}\{\sum_{t=s}^{n}\phi^{1/u}(t-s)\}\right\}.$$

As

$$\Lambda_n(\phi, 4(p-2)) \;=\; O\left(\sum_{t=0}^{\infty}\exp(-\frac{b|t|}{4(p-2)})\right) \;=\; O(p-2),$$

we get, using $m^m \leq m!\exp(m)$, that

$$|\operatorname{cum}_p(\tilde{\alpha}_{jk})| \;=\; O\left((p!)^2\, n^{-1}\,(Cn^{-1}2^{j/2})^{p-2}\right) \tag{22}$$

holds uniformly in (j,k) and $p \geq 2$.

Hence, we can derive asymptotic normality analogously to the regression case.

3 Spectral density estimation

3.1 Spectrum estimation for stationary processes

In this context the aim is to estimate the spectral density

$$f(\omega) \;=\; \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} cov\{X_t; X_{t+s}\}\exp(-i\omega s), \quad \omega \in [-\pi,\pi],$$

of a stationary process $\{X_t\}$. A basis for a wavelet estimator will be the periodogram of the observed (tapered) data $X_t, 1 \leq t \leq T$:

$$I_T(\omega) = (2\pi\;H_{2,T})^{-1}\;|\sum_{s=1}^{T} h(\frac{s}{T})\, X_s\,\exp(-i\omega s)|^2, \tag{23}$$

where $h : [0,1] \to [0,1]$ is a sufficiently smooth taper-function ("window") and $H_{2,T} = \sum_{s=1}^{T} h^2(s/T)$ the appropriate norming factor with $H_{2,T} \sim T$ (see, e.g., [Da93]).

Using data-tapers, e.g. a Hanning-window which is of cosine form

$$h(u) = \begin{cases} \frac{1}{2}(1-(\cos 2\pi u)), & \text{if } u \in [0, \frac{1}{2}] \\ h(1-u), & \text{if } u \in [\frac{1}{2}, 1] \end{cases}, \tag{24}$$

is a well-known remedy in spectral estimation to reduce leakage effects, which occur in particular for spectra with a high dynamic range.

In general, the spectral estimation problem can be considered as sort of a multiplicative regression problem with fixed design (cf. model (6)):

$$I_T(\omega_k) = f(\omega_k) \cdot e_k, \quad \omega_k = 2\pi k/T,\ 1 \le k \le T. \tag{25}$$

(By taking logarithm, (25) writes in the familiar form of an additive regression model.) In contrast to pure regression, the additional difficulty in the "approximate model" (25) is that we neither know the exact underlying error distribution nor its dependence structure. Well-known properties of the periodogram are only of asymptotic nature, e.g. its asymptotic exponential distribution and the asymptotic uncorrelation at two distinct Fourier frequencies ω_k and $\omega_j, j \neq k$. Also, we have to cope with heteroscedasticity: in the multiplicative model the (asymptotic) variance depends on the unknown spectral densities $f(\omega_k)$. Nevertheless, it is possible to transfer all the interesting properties of coordinatewise wavelet thresholding which hold for the simple Gaussian model (1): Here the (empirical) wavelet coefficients are built by projection of the spectrum (periodogram) onto the elements of a 2π-periodic wavelet basis of $L_2([-\pi, \pi])$ (see [Dau92], Ch. 9.3, e.g.). Note that, by the abovementioned problem of heteroscedasticity, the asymptotic variance of the empirical coefficients depends on the unknown spectral density, cf. Lemma 3.1 (b) below.

In deriving asymptotic theory for the resulting wavelet estimator again techniques to prove asymptotic normality as in (2) are applicable, i.e. for a general time series we rely on the results of [Ne94] on the risk equivalence to the case of Gaussian noise. If, however, Gaussianity of the underlying time series is assumed, more specific threshold rules can be developed (as in [G93]) by a different technique: Due to the Gaussianity one can represent the empirical coefficients as quadratic forms of some independent and normally distributed random variables and use the chi-square (exponential) distribution to estimate their tail probabilities. This leads to thresholds of order $T^{-1/2}\log(T)$ which cause a loss in efficiency of order $(\log(T))^2$ instead of the usual $\log(T)$ as in the Gaussian error case. However, this approach allows to include all resolution levels.

In both situations we end up with similar results for bias and variance. In contrast to the situation of a Gaussian time series, in general we explicitly need an upper bound for the cumulants of higher order, as described in the introduction. This bound can be derived by a very general lemma on quadratic forms (see [Ne94] and [NvS94], Lemma 3.1, respectively). - The following rates for bias, variance and higher cumulants are given to hold uniformly in $(j,k) \in \mathcal{J}_n$, which has been defined along with eq. (2).

Lemma 3.1 *Under appropriate assumptions as in [Ne94], uniformly in* $(j,k) \in \mathcal{J}_n$,

(a) $E\ \tilde{\alpha}_{jk} \ = \ \alpha_{jk} \ + \ o(T^{-1/2})$

(b) $\sigma_{jk}^2 \ := \ \mathrm{var}(\tilde{\alpha}_{jk}) \ = \ 2\pi \ T^{-1} \int_{\pi}^{\pi} \{f(\omega)\}^2 \, \psi_{jk}(\omega) \ [\psi_{jk}(\omega) + \psi_{jk}(-\omega)] \, d\omega$
$+ \, o(T^{-1}) \ + \ O(2^{-j} T^{-1})$

(c) $|\mathrm{cum}_p(\tilde{\alpha}_{jk}/\sigma_{jk})| \ \le \ (p!)^{2+2\gamma} \, (CT^{\mu})^{-(p-2)}$ *for* $p \ge 3$,
γ *as in Assumption (B5) below and appropriate* $C, \mu > 0$.

As in Section 2.2 this helps us to derive a strong form of asymptotic normality which puts special emphasis on moderate and large deviations, i.e. in the form of equation (2).

Finally, in relation to Theorem 1.1, the resulting theorems for the nearly-optimal rates of the coordinatewise thresholded wavelet spectral estimator are the following:

Theorem 3.2 *Let* $\mathcal{F} = B_{p,q}^m(C)$ *be some ball in the Besov space* $B_{p,q}^m$ *with either* $p \ge 1$ *and* $m > 1/p$, *or* $m, p \ge 1$ *for* $f \in \mathcal{F}$ *being of bounded variation.*

A) For Gaussian time series (see [G93], Theorem 1): Let $\hat{f}$ *be the wavelet estimator based on thresholds* $\lambda_T \ = \ 2 \ \log(T) \ T^{-1/2}$ *(for all levels* j*). Then*

$$\sup_{f \in \mathcal{F}} \{E \ \|\hat{f} \ - \ f\|^2_{L_2([-\pi,\pi])}\} \ = \ O\left(T^{-2m/(2m+1)} \ (\log(T))^2\right). \tag{26}$$

B) For general (i.e. non-Gaussian) time series (see [Ne94], Theorem 5.2): Let $\hat{f}$ *be the wavelet estimator based on universal thresholds* $\lambda_T \ = \ \max_{(j,k) \in \mathcal{J}_n} \{\sigma_{jk}\} \sqrt{2 \log(\# \mathcal{J}_n)}$ *(for levels* j *with* $2^j \le T^{1-\delta}$ *). Then*

$$\sup_{f \in \mathcal{F}} \{E \ \|\hat{f} \ - \ f\|^2_{L_2([-\pi,\pi])}\} \ = \ O\left((\log(T)/T)^{-2m/(2m+1)}\right). \tag{27}$$

Note that, as always, in the estimator $\hat{f}$ thresholding is not applied to the coarsest levels (with $j < j_0$, say, with some appropriately chosen j_0.)

3.2 Spectrum estimation for locally stationary processes

In this final section, we turn to our main example for a highly "non-standard" situation in which wavelet thresholding works: With thresholding of empirical coefficients built by projection of suitably time-localized periodograms of a locally stationary time series (see Dahlhaus, [Da93]) we address the problem of non-Gaussian and highly dependent 2-dimensional (spectral) curve estimation: we shall see below that despite the fairly complicated dependence structure of our observations (i.e. the local periodograms) we are able to derive rigorous asymptotic results which are based on either one of the presented techniques for the stationary SDE-problem of section 3.1. That is, surprisingly enough, wavelet thresholding even works for situations where we have little knowledge on the

actual underlying distribution and correlation structure of the empirical coefficients - as long as arguments of asymptotic equivalence or tail estimation apply.
Before we introduce the model of a locally stationary process we like to mention that in this section not only we distinguish according to the distribution of the underlying time series - as we did while discussing the stationary situation. Moreover we consider wavelet thresholding w.r.t. two different wavelet bases used for our two-dimensional estimation problem: In the following subsection we show that in order to generally treat problems in anisotropic smoothness classes with different degrees of regularity in different directions, one should use a particular tensor product basis, i.e. one with separate scale parameter for every dimension. This basis is superior to its one-scale multiresolution analog, if we compare the resulting minimax rates for estimation in the specific anisotropic smoothness class. Moreover we shall show that this rate is attained by our coordinatewise thresholded estimator. And finally, in the context of our particular application, we give a fully adaptive estimator of the evolutionary spectrum which is able to automatically adapt to different structure of the unknown object in time and in frequency direction.

3.2.1 Anisotropic smoothness classes In [NvS94] a general multidimensional estimation problem was studied introducing so-called anisotropic function spaces (see Nikol'skii, [Ni75]) with possibly different degrees of regularity in different directions. It was shown, for a typical example of a 2-d anisotropic Sobolev class

$$\mathcal{F}_{p_1,p_2}^{m_1,m_2} \;=\; \mathcal{F}(m_1,m_2,p_1,p_2,C) \;=\; \left\{ f \;\middle|\; \sum_{i=1}^{2} \left(\|f\|_{p_i} \,+\, \|\frac{\partial^{m_i}}{\partial x_i^{m_i}} f\|_{p_i} \right) \;\le\; C \right\} \tag{28}$$

for any positive constant C, with $m_i \ge 1, p_i \ge 1$ and $m_i > 1/p_i$, $i = 1, 2$, that thresholded wavelet estimators based on the following "tensor product" basis in $L_2([0,1]\times[0,1])$ attain the optimal rate of convergence within this class (see [NvS94], Theorem 2.2):

Theorem 3.3 *Let $\widehat{f}_\epsilon$ be the thresholded estimator based on a wavelet expansion w.r.t. basis $\mathcal{B}$ defined straight below and based on optimal thresholds which depend on the m_1, m_2 and on the noise level ϵ as defined in a 2-d Gaussian white noise model, analogous to model (1). If Assumption (B1) below is fulfilled, then*

$$\sup_{f\in\mathcal{F}_{p_1,p_2}^{m_1,m_2}} \left\{ E\|\widehat{f}_\epsilon - f\|^2 \right\} \asymp \inf_{\widetilde{f}} \sup_{f\in\mathcal{F}_{p_1,p_2}^{m_1,m_2}} \left\{ E\|\widetilde{f} - f\|^2 \right\} \;=\; O\left(\epsilon^{2\vartheta(m_1,m_2)} \right), \tag{29}$$

where

$$\vartheta(m_1,m_2) \;=\; \frac{2m_1m_2}{2m_1m_2 + m_1 + m_2}. \tag{30}$$

The basis $\mathcal{B}$ of $L_2([0,1]\times[0,1])$ is obtained as

$$\begin{aligned} \mathcal{B} \;&=\; \{\phi_{lk_1}(x_1)\phi_{lk_2}(x_2)\}_{k_1,k_2} \cup \left(\bigcup_{j_1\ge l} \{\psi_{j_1k_1}(x_1)\phi_{lk_2}(x_2)\}_{k_1,k_2} \right) \\ &\cup \left(\bigcup_{j_2\ge l} \{\phi_{lk_1}(x_1)\psi_{j_2k_2}(x_2)\}_{k_1,k_2} \right) \cup \left(\bigcup_{j_1,j_2\ge l} \{\psi_{j_1k_1}(x_1)\psi_{j_2k_2}(x_2)\}_{k_1,k_2} \right). \end{aligned} \tag{31}$$

Here $\{\phi_{lk}\}_k \cup \{\psi_{jk}\}_{j\geq l;k}$ are chosen to build a orthonormal basis of $L_2[0,1]$, where the functions ϕ and ψ satisfy, for $m \geq \max\{m_1, m_2\}$,

(B1) (i) ϕ and ψ are in C^m,
(ii) $\int \phi(t)\,dt \;=\; 1$,
(iii) $\int \psi(t)t^k\,dt \;=\; 0 \quad$ for $\quad 0 \leq k \leq m-1$.

Such bases are given by Meyer (1991) and Cohen, Daubechies and Vial (1993).

Another construction, which corresponds to what is also known as the "non-standard decomposition" (see, e.g. [BCR93]), and which is a frequently used alternative to the so-called "standard decomposition" of $L_2([0,1]\times[0,1])$, is given by

$$\begin{aligned}\widetilde{\mathcal{B}} \;&=\; \{\phi_{lk_1}(x_1)\phi_{lk_2}(x_2)\}_{k_1,k_2} \\ &\cup \bigcup_{j\geq l}\{\phi_{jk_1}(x_1)\psi_{jk_2}(x_2), \psi_{jk_1}(x_1)\phi_{jk_2}(x_2), \psi_{jk_1}(x_1)\psi_{jk_2}(x_2)\}_{k_1,k_2}.\end{aligned} \tag{32}$$

For details on $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ we refer to [NvS94]. Note that we can also use different one-dimensional bases to build a two-dimensional basis, which is done in Subsection 3.2.3 in view of the special problem considered there.

By Theorem 2.4 of [NvS94], we can compare these two bases w.r.t. their minimax risk as long as coordinatewise shrinkage is considered: It turns out that basis $\mathcal{B}$ is superior to basis $\widetilde{\mathcal{B}}$, if $m_1 \neq m_2$. This can be observed by a slower rate for the lower bound for the respective minimax risk based on $\widetilde{\mathcal{B}}$. For this one has to replace $\vartheta(m_1, m_2)$ by some $\widetilde{\vartheta}(m_1, m_2) \;=\; \min\{\frac{m_1}{m_1+1}, \frac{m_2}{m_2+1}\}$, which is smaller than $\vartheta(m_1, m_2)$ if $m_1 \neq m_2$.

Finally it has been shown that in comparison to the rate of Theorem 3.3 we again only lose some logarithmic term if the theoretical optimal but unknown threshold is replaced by some appropriately chosen universal threshold.

3.2.2 The model of local stationarity In [Da93] Dahlhaus introduced the following model which generalizes the Cramér representation of a stationary stochastic process (see [Pr81], e.g.).

Definition 3.4 *A sequence of stochastic processes* $X_{t,T}\,(t \;=\; 1,\ldots,T)$ *is called* locally stationary *if there exists a representation*

$$X_{t,T} = \mu(\frac{t}{T}) + \int_{-\pi}^{\pi} A(\frac{t}{T}, \lambda)\; \exp(i\omega t)\, d\xi(\omega)\,, \tag{33}$$

where

(i) $\xi(\omega)$ *is a stochastic process on* $[-\pi, \pi]$ *with* $\overline{\xi(\omega)} \;=\; \xi(-\omega)$, $E\,\xi(\omega) \;=\; 0$ *and orthonormal increments, i.e.* $cov(d\xi(\omega),\; d\xi(\omega')) = \delta(\omega-\omega')d\omega\,$,
$\mathrm{cum}\{d\xi(\omega_1), \ldots, d\xi(\omega_k)\} \;=\; \eta(\sum_{j=1}^{k}\omega_j)\; h_k(\omega_1, \ldots, \omega_{k-1})\; d\omega_1\; \ldots\; d\omega_k,$

where cum$\{\ldots\}$ *denotes the cumulant of oder* k, $|h_k(\omega_1,\ldots,\omega_{k-1})| \leq const_k$ *for all* k *(with* $h_1 = 0, h_2(\omega) = 1$*) and* $\eta(\omega) = \sum_{j=-\infty}^{\infty}\delta(\omega + 2\pi j)$ *is the period* 2π *extension of the Dirac delta function.*

(ii) *$A(u,\omega)$ is a function on $[0,1] \times [-\pi,\pi]$ which is 2π-periodic in ω, with $A(u,-\omega) = \overline{A}(u,\omega)$.*

Remark 3.1

(i) In (33) t denotes a time point in the set $\{1,2,\ldots,T\}$ while u denotes a time point in the rescaled interval $[0,1]$, i.e. $u = t/T$. Note that (33) does *not* define a finer and finer discretized continuous time process as T tends to infinity. It rather means that more and more data of the same local structure, given by $A(t/T,\lambda)$, are observed with increasing T. As illustration we cite an example given in [Da93], which is

$$A(\frac{t}{T},\lambda) \;=\; I_{[\frac{1}{4},\frac{1}{2}]}(\frac{t}{T})\;\delta(\lambda-\lambda_1) \;+\; I_{(\frac{1}{2},\frac{3}{4}]}(\frac{t}{T})\;\delta(\lambda-\lambda_2) \qquad (34)$$

With increasing T more and more periods of the two harmonics $\exp(i\lambda_1 t)$ and $\exp(i\lambda_2 t)$ are observed.

(ii) In Dahlhaus, [Da93], a slightly more general definition of a locally stationary process was given. There, the representation in (33) is based on a sequence of functions $A^o_{t,T}(\omega)$ instead of the function $A(u,\omega)$, the difference of which has to fulfill $\sup_{t,\omega} |A^o_{t,T}(\omega) - A(t/T,\omega)| \le K\,T^{-1}$, for some positive constant K.
Note that with this, the class of autoregressive processes with time-varying coefficients now is included in the class of locally stationary processes.
In our work, for reasons of notational convenience, we do not want to adopt this more general definition, noting that all results will continue to hold for the broader class.

Note that, as in [Da93], for simplicity we assume that $\mu(u) = 0$, i.e. we do not treat the problem of estimating the mean of the time series.
In comparison to [Da93] and [Da94], here, our smoothness assumptions on $A(u,\omega)$ are slightly relaxed: Basically we like to impose minimal smoothness as being of bounded variation on $U \times \Pi := [0,1] \times [-\pi,\pi]$ (which is made precise in Assumption (B2)). For technical reasons, in order to facilitate proofs, we impose an additional smoothness condition on the decay of the Fourier coefficients of $A(u,\omega)$ as a function of ω, which implies continuity of A in ω.

Before proceeding with the introduction of the evolutionary spectrum of $\{X_T\}$ we gather the assumptions that are necessary for deriving our asymptotic results:

Definition 3.5 *(Total variation on $U \times \Pi := [0,1] \times [-\pi,\pi]$):*

$$TV_{U\times\Pi}(f) := \sup \sum_i \sum_j |f(u_i,\omega_j) - f(u_i,\omega_{j-1}) - f(u_{i-1},\omega_j) + f(u_{i-1},\omega_{j-1})|\,,$$

where the supremum is to be taken over all partitions of $U \times \Pi$.

Now we impose the following assumptions:

(B2) a) $A(u,\omega)$ has bounded total variation on $U \times \Pi$, i.e. $TV_{U\times\Pi}(A) < \infty$.
b) $\sup_u TV_{[-\pi,\pi]}(A(u,.)) < \infty$ and $\sup_\omega TV_{[0,1]}(A(.,\omega)) < \infty$.
c) $\sup_{u,\omega} |A(u,\omega)| < \infty$.
d) $\inf_{u,\omega} |A(u,\omega)| \geq \kappa$ for some $\kappa > 0$.

The rather common assumption (B2) (d) (implying boundedness of the spectral density from below) is introduced to simplify the proofs of the following theorems. It implies that the variances of the empirical coefficients are bigger than $C\ T^{-1}$ for some $C > 0$. This provides a relation between variance and upper estimates of the higher order cumulants that allows to conclude asymptotic normality of the empirical coefficients. Note however that we can drop assumption (B2) (d) if we apply a slightly more involved technique as in [Ne94].

(B3) Let $\widehat{A}(u,s) := 1/(2\pi) \int A(u,\omega)\exp(i\omega s)\,d\omega$, $s \in \mathbb{Z}$, $u \in [0,1]$.
Then: $\sup_u \sum_s |\widehat{A}(u,s)| < \infty$.

(B4) Both $\phi(u)$ and $\widetilde{\phi}(\omega)$ as well as $\psi(u)$ and $\widetilde{\psi}(\omega)$ have bounded total variation on $[0,1]$ and $[-\pi,\pi]$ respectively. Further, $\sum_s |\widehat{\widetilde{\psi}}(s)| < \infty$, and the same holds for $\widehat{\widetilde{\phi}}(s)$.

(B5) $\sup_{1\leq t_1\leq T} \left\{\sum_{t_2,\ldots,t_k=1}^T |\mathrm{cum}(X_{t_1},\ldots,X_{t_k})|\right\} \leq C_2^k (k!)^{1+\gamma}$ for all $k = 2,3,\ldots,$
where $\gamma \geq 0$.

Note that these are somewhat minimal conditions part of which might be fulfilled simply by restricting $A(u,\omega)$ to be member of the specific smoothness class under consideration (anisotropic Sobolev, Hölder,...). For Sobolev restrictions, e.g., (B2) (b) and (c) and (B3) are implications of the considered Sobolev smoothness. Note also the similarity between (B5) and eq. (19) in Section 2.2.

We like to mention that this minimal smoothness of A is sufficient to ensure the locally stationary behavior of the process, in the sense that we end up with a spectrum which is uniquely defined in some L^2- rather than in an almost everywhere sense. However, for reasons of completeness, we like to also give this stronger definition of the evolutionary spectrum which, under the appropriate stronger smoothness of A, was considered by Dahlhaus, [Da93]:

Definition 3.6 *As evolutionary spectrum of $X_{t,T}$ given in (33) we define for $u \in (0,1)$*

$$f(u,\omega) = \lim_{T\to\infty} \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} cov\{X_{[uT-\frac{s}{2}],T}; X_{[uT+\frac{s}{2}],T}\}\exp(-i\omega s), \tag{35}$$

where $X_{t,T}$ is defined by $A(t/T,\omega) = A(0,\omega)$ for $t < 1$ and $A(t/T,\omega) = A(1,\omega)$ for $t > T$.

By Dahlhaus, [Da93], Theorem 2.2, if $A(u,\omega)$ is differentiable in u and ω (with uniformly bounded derivatives), then

$$f(u,\omega) = |A(u,\omega)|^2,\ u \in (0,1) \qquad \text{a.e. in } \omega. \tag{36}$$

If ever this condition on A is fulfilled we shall understand the given limit in (36) as pointwise in u and ω.

More generally, however, it is shown by Theorem 3.1 of [NvS94] that, if we turn to the L_2-limit, equation (36) still holds, in the $L_2(du, d\omega)$-sense on $U \times \Pi$:

Theorem 3.7 *Under assumptions (B2) and (B3),*

$$\lim_{T\to\infty} \int_0^1 \int_{-\pi}^{\pi} \{ \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} [\text{cov}\{X_{[uT-\frac{s}{2}],T}; X_{[uT+\frac{s}{2}],T}\} \exp(-i\omega s)] - |A(u,\omega)|^2 \}^2 \, d\omega \, du = 0.$$

An intermediate result, finally, which is in the $L^2(d\omega)$-sense, but pointwise in $u \in (0,1)$, is given by [Da94], Theorem 2.2, where uniform Lipschitz-continuity of $A(u,\omega)$ in both components with Lipschitz exponent $\alpha > 1/2$ is needed.

3.2.3 Wavelet estimating the evolutionary spectrum Wavelet estimation of the evolutionary spectrum, a function $f(u,\omega)$ of two dimensions, frequency and time, can be done by two alternative choices of wavelet basis, namely either basis $\widetilde{\mathcal{B}}$ or basis $\mathcal{B}$, as introduced in Section 3.2.1. A first approach, using the 2-d multiresolution basis $\widetilde{\mathcal{B}}$, was developed by von Sachs and Schneider in [vSS94] where Gaussianity of the locally stationary time series was assumed. With that using techniques similar to the techniques of Gao, [G93], for the stationary situation - see Section 3.1 - appropriate thresholds can be derived by estimating the tail probabilities of the empirical wavelet coefficients. The second approach of Neumann and von Sachs, [NvS94], based on the tensor product basis $\mathcal{B}$, allows for non-Gaussian processes using techniques similar to [Ne94]. In addition, it benefits from some modification which results in a fully adaptive estimator.

Regardless to whether we assume Gaussianity of the time series (which only slightly changes the thresholding rule; cf. Section 3.1) we now describe the different approaches in detail:
The approach of [vSS94] starts from a local version of the classical periodogram over segments of length N of the tapered data $X_{t,T}, 1 \le t \le T$:

$$I_N(u,\omega) = (2\pi\ H_{2,N})^{-1} \mid \sum_{s=0}^{N-1} h(\frac{s}{N}) X_{[uT-\frac{N}{2}+s+1],T} \exp(-i\omega s)|^2, \tag{37}$$

where for the taper function h we refer back to (24) in Section 3.1.
As it is our goal to end up with a wavelet estimate of $f(u,\omega)$ by projecting $I_N(u,\omega)$ onto the 2-d wavelet basis $\widetilde{\mathcal{B}}$ we like to introduce some more necessary notation:

As in eq. (32) this basis $\widetilde{\mathcal{B}}$ of $L_2(U \times \Pi)$ is given by

$$\begin{aligned} \widetilde{\mathcal{B}} &= \{\phi_{lk_1}(u)\widetilde{\phi}_{lk_2}(\omega)\}_{k_1,k_2} \\ &\cup \bigcup_{j\ge l} \{\phi_{jk_1}(u)\widetilde{\psi}_{jk_2}(\omega), \psi_{jk_1}(u)\widetilde{\phi}_{jk_2}(\omega), \psi_{jk_1}(u)\widetilde{\psi}_{jk_2}(\omega)\}_{k_1,k_2}. \end{aligned} \tag{38}$$

For sake of notational convenience, for the second union for fixed j we introduce the following abbreviation, which uses the label $\mu = h, v, d$ in order to distinguish between "horizontal, vertical and diagonal" components in the time-frequency plane:

$$\{\Psi^{\mu}_{j\mathbf{k}}(u,\omega) : \mathbf{k} = (k_1,k_2),\ \mu = h,v,d\}$$

With that the wavelet coefficients of $f(u,\omega)$ write as

$$d^{\mu}_{j\mathbf{k}} \ = \ \int_0^1 \int_{-\pi}^{\pi} f(u,\omega)\,\Psi^{\mu}_{j\mathbf{k}}(u,\omega)\,du\ d\omega\ . \tag{39}$$

To build the empirical wavelet coefficients $\check{d}^{\mu}_{j\mathbf{k}}$, $I_N(u,\omega)$ is calculated on possibly overlapping segments of $X_{t,T}$ of length N: The shift from segment to segment is denoted by S, with $1 \le S \le N$, hence, for the number of segments M, say, we have the following relation

$$T \ = \ S\ (M-1) + \ N\,.$$

Thus, we calculate $I_N(u,\omega)$ at the midpoints of the segments, i.e. the M timepoints

$$u_i = t_i/T, \quad \text{where} \ \ t_i = S \cdot i + N/2,\ 0 \le i \le M-1.$$

In order to derive our asymptotic results, as $T \to \infty$, N and S are assumed to fulfill

$$(N\ \log T)^2/T \longrightarrow 0,\ N^4/T \longrightarrow \infty \quad \text{and} \quad S = N \text{ or } S/N \to 0.$$

In practice, as usual, the data have to be sampled on an equally spaced grid $(u_i,\omega_n), 0 \le i,n \le N-1$, with $N = 2^J$, i.e. with the finest level chosen to be $J = \log_2(N)$. Hence, for calculating the finite sample 2-d wavelet procedure, we set $M = N$ to end up with a quadratic multiresolution analysis (with same finest level J for each of the two dimensions).

The resulting empirical wavelet coefficients are

$$\check{d}^{\mu}_{j\mathbf{k}} = \frac{1}{N} \sum_{i=0}^{N-1} \int_{-\pi}^{\pi} I_N(u_i,\omega)\,\Psi^{\mu}_{j\mathbf{k}}(u_i,\omega)\,d\omega. \tag{40}$$

Note that in practice these coefficients are not calculated according to (40) but by using a so-called "collocation projection" described in [vSS94]. This is due to the fact, that one has to care on how to project the theoretical object $f(u,\omega)$ into the space $V^{(2)}_J$, i.e. the space generated by the 2-d scaling functions of the finest scale J. Note that in this approach one really makes use of all of the wavelet coefficients, starting from the finest scale J with $N = 2^J$.

Remark 3.2 Obviously, the adaption properties of this estimate depend on the choice of the segment length N and the shift S, respectively. Once S (or N) is chosen, the best possible resolution w.r.t. time is fixed which obviously has also consequences for the performance in frequency direction: The larger N the worse is the time resolution, but the better can low-frequency components be detected, and vice versa. This gives rise to a modification which will be presented in Section 3.2.5 below.

A second possibility, as indicated in Section 3.2.1, is to rely on the two-dimensional tensor product basis $\mathcal{B}$ of $U \times \Pi$, using two possibly different one-dimensional bases $\{\phi(u),\ \psi(u)\}$, $u \in [0,1]$, and $\{\tilde{\phi}(\omega),\ \tilde{\psi}(\omega)\}$, $\omega \in [-\pi,\pi]$, which both fulfill Assumption (B1).

The resulting true and empirical wavelet coefficients would then look like follows:

$$\alpha_{j_1k_1j_2k_2} \;=\; \int_{U\times\Pi} f(u,\omega)\,\psi_{j_1k_1}(u)\,\widetilde{\psi}_{j_2k_2}(\omega)\,du\;d\omega\,, \tag{41}$$

and

$$\widetilde{\alpha}_{j_1k_1j_2k_2} \;=\; \frac{1}{N}\sum_{i=0}^{N-1}\int_{-\pi}^{\pi} I_N(u_i,\omega)\,\psi_{j_1k_1}(u_i)\widetilde{\psi}_{j_2k_2}(\omega)\,d\omega, \tag{42}$$

where for (42) the same applies as it was mentioned for the empirical coefficients defined by (40). For both wavelet expansions, in order to apply the appropriate non-linear thresholding, we have to investigate the asymptotic properties of the empirical coefficients. This is regardless to the basis used, but rather differs according to the underlying distribution of the considered time series (Gaussian or not): Quite similarly to the distinction made in Section 3.1 for the stationary case we are able to derive asymptotic results. This is done by the very same techniques though in our investigations we have to rigorously take care of the local-stationarity which helps to manage the two-dimensional dependent error situation. We summarize the results and note that details of the proofs can be found in [vSS94] for the Gaussian, and in [NvS94] for the non-Gaussian situation.

3.2.4 Asymptotics for empirical coefficients As described in Section 3.1 we have to study the asymptotic behavior of all cumulants of the empirical wavelet coefficients. For both cases, i.e. given a locally stationary process, Gaussian or not, the quality of the order of convergence of all cumulants is the same. We end up with precisely the same rates as in the stationary case, given in Lemma 3.1. However it is not straightforward at all to derive these rates under the low regularity assumptions (B2) and (B3) on the spectrum, in particular for bias and variance (see [NvS94], proof of Lemma 3.2). Of course, the leading term of the asymptotic variance has to be slightly adapted to the two-dimensional situation, i.e. replacing the one-dimensional spectral density and wavelet function by their 2-d analogs.

For the second step, we have to distinguish between the Gaussian and non-Gaussian situation, again quite similarly to the stationary case: For the *Gaussian* case, by Proposition 5.9 in [vSS94], an analogous result to Lemma 6 in [G93] on the tail probability is delivered. Note that this makes use of Lemma 3.1(b) on the asymptotic variance. Again, as in Section 3.1, we end up with a threshold that is proportional to $T^{-1/2}\log T$. A typical formulation of a result in the near-optimal sense of Theorem 3.2 (A) can be given by the following:

Theorem 3.8 *Let $\mathcal{F}$ be an appropriate smoothness class for functions on $L_2(U\times\Pi)$ with (isotropic) degree of smoothness m. Let $\lambda_T \;=\; K\;\log T\;T^{-1/2}$, with some positive constant K, and let $\widehat{f}(u,\omega)$ be the wavelet estimator based on this universal threshold λ_T. Then*

$$\sup_{f\in\mathcal{F}}\{E\;\|\widehat{f}\;-\;f\|^2_{L_2(U\times\Pi)}\}\;=\;O\left(T^{-2m/(2m+2)}\;(\log T)^2\right). \tag{43}$$

Note that $T^{-2m/(2m+d)}$ is the classical rate of mean-squared convergence in d-dimensional function estimation problems (here with $d=2$). Note moreover, that the rate in this theorem does not depend on N, but on T merely. - Appropriate function classes can be

2-dimensional Hölder or Sobolev classes, or even Besov classes with $m > 2/p$, as shortly described in [DJKP95], Section 6.1.3.

In the *non-Gaussian* situation, again by the general lemma of quadratic forms, we make use of the asymptotic equivalence to the case of Gaussian noise. Note that this general lemma holds regardless to the assumption of stationarity of the time series, as long as Assumption (B5) is fulfilled.
At this place we skip the presentation of the resulting theorem, which is completely analogous to the one for the stationary situation (see Theorem 3.2 (B)) with a threshold proportional to $T^{-1/2}(\log T)^{1/2}$ and the appropriate restrictions on the levels of the coefficients which are to be thresholded.

The performance of these two coordinatewise thresholded estimator based on the two different bases is presented in a simulated example which will be found in the last subsection.

3.2.5 Modification towards full adaptivity of the estimator Finally, we present a modification of the above estimator which allows to overcome the difficulty of choosing the parameter N of preliminary smoothing: This is independent of the basis used, and we decide to concentrate merely to the situation of [NvS94], i.e. for the tensor product basis $\mathcal{B}$ and the non-Gaussian situation. As particular result the estimator, which is considered for functions in an anisotropic Sobolev class (as typical example), will nearly attain the classical minimax rate for this class - cf. Section 3.2.1.

A straightforward analogy to the definition of the spectral density is the introduction of the following periodogram-like statistic $I_{t,T}$, $1 \le t \le T$, which is different to the localized periodogram defined in (37) :

$$I_{t,T}(\omega) \;=\; \frac{1}{2\pi} \sum_{s} X_{[t-s/2],T} X_{[t+s/2],T} \exp(i\omega s), \tag{44}$$

where the sum over s only exceeds over the region where $|s| \le \min\{t-1, T-t\}$. This is completely in accordance with model (33) and equation (35), as in this model it is not meaningful at all to consider the spectrum at the boundaries $u = 0$ and $u = 1$, respectively.

Note that $I_{t,T}$ can be considered as a preliminary "estimate" which is even more fluctuating than the classical periodogram is. However, in contrast to the first approach with pre-smoothing in time direction, now projection of these "rough periodograms" $I_{t,T}$ on elements of the wavelet basis $\mathcal{B}$ will do the whole task of adaptive local smoothing!

The modified empirical wavelet coefficients (as counterparts to the "true" ones in eq. (41)) are defined as follows:

$$\widetilde{\alpha}_{j_1 k_1 j_2 k_2} \;=\; \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \psi_{j_1 k_1}(u)\, du \; \left[\int_{-\pi}^{\pi} \widetilde{\psi}_{j_2 k_2}(\omega) I_{t,T}(\omega)\, d\omega \right]. \tag{45}$$

Also for those modified $\widetilde{\alpha}_{j_1 k_1 j_2 k_2}$ Lemma 3.1 holds with the original rates, now of course in its appropriately modified two-dimensional form.

By exactly the same procedure as in the previous subsection we are able to show asymptotic risk equivalence to the Gaussian white noise situation and can derive a theorem which is in the spirit of Theorem 3.3 - however, with near-optimal nonrandom thresholds which still depend on the unknown constants in the asymptotic variance of the empirical wavelet coefficients. Thus the resulting $L_2(U \times \Pi)$ - convergence rate of the thresholded estimator $\widehat{f}(u,\omega)$ is of order $O\left((\log(T)/T)^{-\vartheta(m_1,m_2)}\right)$. For details see Theorem 3.2 of [NvS94].
If, finally, we appropriately choose random thresholds (based on some consistent preliminary estimate for the asymptotic variance), then, under some technical conditions, we can derive the following theorem, where $\widehat{\widehat{f}}$ be the same estimator as $\widehat{f}$ but with these random thresholds:

Theorem 3.9 *Under the conditions of Theorem 3.3 of [NvS94],*

$$\sup_{f \in \mathcal{F}_{p_1,p_2}^{m_1,m_2}} \left\{ E\|\widehat{\widehat{f}} - f\|^2_{L_2([0,1]\times[-\pi,\pi])} \right\} = O\left((\log(T)/T)^{-\vartheta(m_1,m_2)}\right), \tag{46}$$

where $\vartheta(m_1,m_2)$ is defined by (30).

3.2.6 An illustrating simulation example Now we like to present a simulation example of a time-varying spectral density which shows the performance of the coordinatewise thresholded estimator. For simplicity our wavelet basis used is the same for both dimensions, namely orthogonal (periodized) spline wavelets of the Battle-Lemarié-family with order $m = 6$ (for details see [vSS94]), i.e. the functions are elements of $C^4(U \times \Pi)$.

This simplification is justified due to our chosen example of a spectrum $f(u,\omega)$ which can be considered as periodic in u, too, and which is also symmetric to $u = 1/2$:

$$f(u,\omega) = [1-(\frac{\omega}{\omega_m})^2]^{\frac{1}{2}}\, I_{[0,\omega_m]}(\omega) + [1-(\frac{\omega-\omega_0}{\omega_m})^2]^{\frac{1}{2}}\, I_{[\omega_0-\omega_m,\omega_0+\omega_m]}(\omega) + P_1\, \exp\{-\frac{(\omega-\omega_1)^2}{2\sigma_1^2}\} + R_0\,, \tag{47}$$

where w.l.o.g. we assume $\omega \geq 0$ due to symmetry, and where $\omega_m = \omega_m(u) = \sigma_0\ (2 + \cos 2\pi\nu_0 u)$ with $\sigma_0 = 0.03$, $\nu_0 = 4$, $\omega_0 = 0.3$, $P_1 = 0.2$, $\omega_1 = 0.45$, $\sigma_1 = 0.001$ and a constant $R_0 = 0.001$ (resulting as the spectral component of a background white noise component in the underlying process).

Whereas the narrow Gaussian bump at ω_1 might represent some narrowband interference term, the form of this example in the neighborhood of both the origin and of ω_0 can be motivated by a typical situation in the practice of mobile radio propagation. The time-dependent components of the resulting power spectrum can be derived as follows (for details, see [Ja74], and again [vSS94], Section 6):
A microwave radio signal transmitted between a fixed base station and a moving vehicle in some urban environment, exhibits extreme variations in both amplitude and apparent frequency:
From the viewpoint of an observer on the mobile unit, the received signal, a plane wave of the form

$$C \cdot \cos(2\pi\, \omega(\alpha)t \,+\, \phi)$$

may be represented as a carrier with randomly varying phase ϕ, amplitude C and frequency $\omega(\alpha)$ (with randomly varying α). Due to the Doppler shift, caused by the movement of the mobile unit with velocity V into direction α w.r.t. the sender station, the frequency

$$\omega(\alpha) = \omega_c + \omega_m \cos \alpha,$$

is to be found in a narrow band around the carrier frequency ω_c. This band is (for $\alpha = 0$) of maximum width $2\omega_m = 2V/\nu$, with ν being the wavelength of the transmitted carrier frequency.
A suitable model for the three field components of the signal (electric field E_z, magnetic field H_x and H_y) is a Gaussian random process, *stationary* as long as V (and ν) do not depend on time t. Instead of determining the statistical properties of this Gaussian random process from its moments, they are most easily obtained from the power spectrum (as the Fourier transform of the autocorrelation of the signal components).
As simplest model, the probability distribution $p(\alpha)$ of the power over the angle α is assumed to be constant. Hence, for studying the H_x-field component of the signal as typical example only, we end up with a spectrum of the following form, according to [Ja74],

$$f_{H_x}(\omega) \sim \left[1 - \left(\frac{\omega - \omega_c}{\omega_m}\right)^2\right]^{\frac{1}{2}},$$

where we used that $|d\omega| = \omega_m|-\sin\alpha||d\alpha| = (\omega_m^2 - (\omega-\omega_c)^2)^{\frac{1}{2}}|d\alpha|$ (for details, see [Ja74], Sec. 1.2.1).
As long as V is constant in time, $f_{H_x}(\omega)$ is also. But, in practice, of course, the mobile unit changes its velocity: hence, a more realistic model would be to allow for a time dependent power spectrum

$$f(t,\omega) = f_{H_x}(\omega(t)) = \left[1 - \left(\frac{\omega - \omega_c}{\omega_m(t)}\right)^2\right]^{\frac{1}{2}},$$

where $\omega_m(t) = V(t)/\nu$.
Consequently, the model for the underlying Gaussian random process becomes *instationary.* An additional modification arises if we allow for a changing environment of the transmitting channel, i.e. an explicit variation of $f(t,\omega)$ in ω, too.

Note that in addition to a time-dependent $\omega_m = \omega_m(u)$, in our chosen example (47), $f(u,\omega)$ has isolated singularities (in its derivative) in ω and a high dynamic range (in ω) of smoother and sharper component .

For the simulation of the underlying locally stationary process $X_{t,T}$ we generate $T = 2048$ data, using the following discretization (in ω) of the integral in (33):

$$X_{t,T} = \frac{2^{\frac{1}{2}}}{T_s^{\frac{1}{2}}} \sum_{k=0}^{T_s-1} A\left(\frac{t}{T}, \frac{2\pi k}{T_s}\right) \exp\left(i\ t\ \frac{2\pi k}{T_s}\right) \xi_k\ , \quad 1 \le t \le T,$$

where $T_s = 8192$ and where ξ_k, $0 \le k \le T_s - 1$, is a simulated Gaussian white noise ($\sim N(0,1)$) - generated by a standard pseudo random-number generator.

Further, we calculate the periodogram over segments of length $N = 128$, with shift $S = 15$, using a data-taper as given by (24). Note that we use the log-periodogram to benefit from the variance-stabilizing effect of taking logarithms.

First we demonstrate the performance of the estimator based on the 2-d multiresolution (MRA) basis $\widetilde{\mathcal{B}}$: Figure 1 shows the true log-spectrum log $f(u, \omega)$, the pure and two versions of the smoothed log-periodogram - globally and locally smoothed (with hard thresholding) - as grey scaled isolines in the time-frequency plane. The local smoothing was performed by hard-thresholding with threshold $\lambda_T = 1 \cdot 10^{-2}$, a value which is in accordance with the asymptotically motivated form of $\lambda_T \sim T^{-1/2} \cdot \log T$, with $T = 2048$. For comparison we add a "globally smoothed" example which simply is the wavelet series cut-off above level $J_0 - 1$ with a choice of $J_0 = 5$.

It can be clearly observed that the noise in the periodogram-estimator is suppressed by non-linear thresholding without losing local structure of $f(u, \omega)$ (e.g. the bump at ω_1), whereas with global smoothing this is not possible simultaneously.
Cuts in ω- and in u-direction (Figure 2) confirm this behaviour.

A second time-dependent spectrum arises by simply replacing the cosine function $\omega_m(u)$ in $f(u, \omega)$ by some periodically piecewise linear one:

$$\omega_m(u) \;=\; \sigma_0 \left(2 \,+\, 10/3 \cdot (frac(\nu_0 u) - 0.7)\; I_{[0.4,1)}(frac(\nu_0 u)) \;-\; I_{[0,0.4]}(frac(\nu_0 u))\right),$$

where $frac(x) := \; x - [x]$.

With this, an example is introduced which is less regular in time, and which might describe a somewhat more realistic dependence of the velocity V of the mobile unit on time. We want to use this example to compare the performance of the same thresholded estimator as above (i.e. w.r.t. the MRA basis $\widetilde{\mathcal{B}}$) with the one in the tensor product basis $\mathcal{B}$. In Figure 3, we observe the following: apparently, both soft thresholded estimators give a denoised version of the pure log-periodogram. But, as can be seen by the number of active wavelet coefficients in 2-d tableau (a representation being often used in image compression), the tensor product basis delivers a reconstruction which is at least as good as for the MRA basis but with considerably less coefficients: Using basis $\mathcal{B}$ the original $2^7 \cdot 2^7 = 16384$ coefficients have been reduced to 437 compared to 660 for the MRA basis $\widetilde{\mathcal{B}}$. Note that it is not easy to do this empirical comparison by a totally objective criterion: here we simply tried to adjust the amount of smoothing (denoising) visually to the same range, which resulted into slightly different values for the thresholds. Although these somewhat preliminary simulations are to be improved by a more detailed analysis, they seem to confirm the theory of subsection 3.2.1 on anisotropic smoothness, here in the particular case of time-dependent spectra. Note that from a theoretical point of view, a comparison of the compression ability of both bases in noisy situations has been delivered by [NvS94] where an appropriate approximation-theoretic functional was considered.

Acknowledgements. We are grateful to the anonymous referee for his careful reading which helped us to improve the presentation of our results. Many thanks also to Kai Schneider for his help with preparing the plots in the figures.

References

[BCR93] Beylkin, G., Coifman, R. and Rokhlin, V. (1991). Fast Wavelet Transforms and Numerical Algorithms I. *Comm. Pure Applied Math.* **44**, 141–183.

[Bri94] Brillinger, D. R. (1994). Some asymptotics of wavelet fits in the stationary error case. Technical report 415, Dept. of Statist., U. C. Berkeley.

[BL92] Brown, D. L. and Low, M. (1992). Asymptotic equivalence of nonparametric regression and white noise. Manuscript.

[CDV93] Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the Interval and Fast Wavelet Transform. *Appl. Comp. Harmonic Anal.* **1**, 54–81.

[Da93] Dahlhaus, R. (1993). Fitting time series models to nonstationary processes. Preprint, University of Heidelberg.

[Da94] Dahlhaus, R. (1994). On the Kullback-Leibler information divergence of locally stationary processes. Preprint, University of Heidelberg.

[DNvS95] Dahlhaus, R., Neumann, M. H., and von Sachs, R. (1995). Nonlinear wavelet estimation of time-varying autoregressive processes. Manuscript in preparation.

[Dau92] Daubechies, I. (1992). *Ten Lectures on Wavelets*, SIAM, Philadelphia.

[DeJ93] Delyon, B. and Juditsky, A. (1993). Wavelet estimators, global error measures: revisited. Technical Report No. 782, Irisa, France.

[DJ92] Donoho, D. L. and Johnstone, I. M. (1992). Minimax estimation via wavelet shrinkage. Technical Report No. 402, Department of Statistics, Stanford University.

[DJ93] Donoho, D. L. and Johnstone, I. M. (1993). Adapting to unknown smoothness via wavelet shrinkage. Technical Report, Department of Statistics, Stanford University.

[DJ94] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.

[DJKP95] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? *J. R. Statist. Soc., Ser. B* **57**, to appear.

[F94] Fan, J. (1994). Test of significance based on wavelet thresholding and Neyman's truncation. Technical Report.

[G93] Gao, H.-Ye (1993). Wavelet estimation of spectral densities in time series analysis. Ph.D. dissertation. University of California, Berkeley.

[Ja74] Jakes, W. C. (1974). *Mobile Radio Propagation.* Microwave Mobile Communications, Ch. 1, Wiley, New York.

[JK70] Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics. Continuous Univariate Distributions - 1, 2.* Wiley, New York.

[JKP92] Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1992). Estimation d'une densité de probabilité par méthode d'ondelettes. *Comptes Rendus Acad. Sciences Paris* (A) **315** 211–216.

[M91] Meyer, Y. (1991). Ondelettes sur l'intervalle. *Revista Mathematica Ibero-Americana* **7** (2), 115–133.

[Na94] Nason, G. P. (1994). Wavelet regression by cross-validation. Preprint, Department of Mathematics, University of Bristol.

[N85] Nemirovskii, A. S. (1985). Nonparametric estimation of smooth regression functions. *Izv. Akad. Nauk SSR Techn. Kibernet.* **3**, 50–60 (in Russian).

[NTP85] Nemirovskii, A. S., Polyak, B. T. and Tsybakov, A. B. (1985). Rate of convergence of nonparametric estimates of maximum-likelihood type. *Problems of Information Transmission* **21**, 258–272.

[Ne94] Neumann, M. H. (1994). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. Manuscript.

[NS95] Neumann, M. H. and Spokoiny, V. G. (1995). On the efficiency of wavelet estimators under arbitrary error distributions. *Math. Methods of Statist.*, to appear.

[NvS94] Neumann, M. H. and von Sachs, R. (1994). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. Discussion Paper No. 24/95, SFB 373, Humboldt University, Berlin, and Technical Report "Berichte der AG Technomathematik" No. 132, University of Kaiserslautern.

[Ni75] Nikol'skii, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems.* Springer, New York.

[Nu94] Nussbaum, M. (1994). Asymptopic equivalence of density estimation and white noise. to appear *Ann. Statist.*

[P94] Patil, P. (1994). Nonparametric hazard rate estimation. Statistics Research Report No. SRR 018-94, Centre for Mathematics and its Applications, Australian National University.

[Pr81] Priestley, M. B. (1981). *Spectral Analysis and Time Series.* Vol. 2, Academic Press, London.

[RSS78] Rudzkis, R., Saulis, L. and Statulevicius, V. (1978). A general lemma on probabilities of large deviations. *Lithuanian Math. J.* **18**, 226–238.

[vSS94] von Sachs, R. and Schneider, K. (1994). Wavelet smoothing of evolutionary spectra by non-linear thresholding. Technical Report "Berichte der AG Technomathematik" No. 106, University of Kaiserslautern.

[S91] Sakhanenko, A. I. (1991). On the accuracy of normal approximation in the invariance principle. *Siberian Advances in Mathematics* **1**, 58–91.

[SW86] Shorack, G. R. and Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. Wiley, New York.

[SJ88] Statulevicius, V. and Jakimavicius, D. (1988). Estimates of semiinvariants and centered moments of stochastic processes with mixing I. *Lithuanian Math. J.* **28**, 67–80.

Captions:

Figure 1: Grey scaled isolines of first example in the time-frequency plane:
$\log f(u,\omega)$,
$\log I_N(u,\omega), N = 128, S = 15,$
globally smoothed log-periodogram ($J_0 = 5$),
locally smoothed log-periodogram (hard thresholding, $\lambda_T = 1 \cdot 10^{-2}$).

Figure 2: cuts of Fig. 1 in frequency direction at $u = 0.54$
and in time direction at $\lambda = 0.24$.

Figure 3: Grey scaled isolines of the second example:
true log-spectrum,
log-periodogram (same parameters as above),
soft-thresholded log-periodogram (MRA basis $\widetilde{\mathcal{B}}$) with $\lambda_T = 5 \cdot 10^{-3}$,
soft-thresholded log-periodogram (tensor product basis $\mathcal{B}$) with
$\lambda_T = 1 \cdot 10^{-2}$,
active wavelet coefficients of these two estimators:
660 for MRA basis $\widetilde{\mathcal{B}}$ vs. 437 for tensor product basis $\mathcal{B}$.