ELSEVIER Computational Statistics & Data Analysis 22 (1996) 351–361

COMPUTATIONAL STATISTICS & DATA ANALYSIS

Adaptive thresholding of wavelet coefficients

Felix Abramovich[1,*], Yoav Benjamini

Department of Statistics and Operations Research, Tel Aviv University, Ramat Aviv 69978, Israel

Received August 1994; revised October 1995

Abstract

Wavelet techniques have become an attractive and efficient tool in function estimation. Given noisy data, its discrete wavelet transform is an estimator of the wavelet coefficients. It has been shown by Donoho and Johnstone (*Biometrika* **81** (1994) 425–455) that thresholding the estimated coefficients and then reconstructing an estimated function reduces the expected risk close to the possible minimum. They offered a global threshold $\lambda \sim \sigma\sqrt{2\log n}$ for $j > j_0$, while the coefficients of the first coarse j_0 levels are always included.

We demonstrate that the choice of j_0 may strongly affect the corresponding estimators. Then, we use the connection between thresholding and hypotheses testing to construct a thresholding procedure based on the false discovery rate (FDR) approach to multiple testing of Benjamini and Hochberg (*J. Roy. Statist. Soc. Ser. B* **57** (1995) 289–300). The suggested procedure controls the expected proportion of incorrectly included coefficients among those chosen for the wavelet reconstruction. The resulting procedure is inherently adaptive, and responds to the complexity of the estimated function and to the noise level. Finally, comparing the proposed FDR based procedure with the fixed global threshold by evaluating the relative mean-square-error across the various test-functions and noise levels, we find the FDR-estimator to enjoy robustness of MSE-efficiency.

Keywords: False discovery rate; Nonparametric regression; Robust smoothing; Multiple comparison procedures

1. Introduction

Suppose we are given data

$$y_i = g(t_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

*Corresponding author. E-mail: felix@math.tau.ac.il.

[1]Part of the work was done while the first author was at the School of Mathematics, University of Bristol, UK.

PII S0167-9473(96)00003-5

where $t_i = i/n$, ε_i's are i.i.d. normal variables with zero mean and variance σ^2, and we wish to estimate the unknown response function $g(\cdot)$ from the data without assuming any particular parametric form of g. There exists a variety of non-parametric regression methods based on the ideas of kernel estimation, spline smoothing, or Fourier series expansion. Recently developed wavelet techniques have become an attractive and efficient tool in adaptive function estimation. Basic wavelet theory may be found in Daubechies (1992), while applications of wavelets for various statistical problems are discussed in Donoho et al. (1995). The function g is expanded in terms of some orthonormal basis functions ψ_l in a way similar to the generalized Fourier series expansion. While for the classical Fourier series the basis functions are sine and cosine functions at different frequencies, in the case of wavelet series the basis functions ψ_{jk} are translations and dilations of a single fixed function ψ called *the mother wavelet* $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$, $j, k \in Z$. For certain choices of ψ the corresponding set of ψ_{jk} forms an orthonormal basis in $L^2(R)$ (examples of such mother wavelets are given in Daubechies, 1992). In this case,

$$g(t) = \sum_j \sum_k d_{jk}\psi_{jk}(t),$$

where the wavelet coefficients $d_{jk} = (g, \psi_{jk}) = \int g(t)\psi_{jk}(t)\,dt = 2^{j/2}\int g\ (t)\psi(2^j t - k)\,dt$.

In contrast to sines and cosines, wavelets are local in both frequency (via dialations) and in time (via translations). This localization offers an advantage, since fewer wavelet basis functions are usually needed to represent the function to a given level of approximation. Moreover, choosing the mother wavelet with corresponding regularity properties, one can generate an unconditional wavelet basis in a wide set of function spaces, such as Besov or Triebel scales. These scales include, in particular, the well-known Sobolev or Hölder scales of smooth functions, but in addition less traditional spaces, like the space of functions of bounded variation (see Meyer, 1992; Donoho et al., 1995, for precise definitions and details). Such "universality" allows a parsimonious wavelet expansion for large variety of different functions.

Suppose we have data at the points $t_i = i/n$ where the sample size $n = 2^{J+1}$ for some J. In the absence of random noise in the data we may find $m = n - 1 = 2^{J+1} - 1$ wavelet coefficients of the function g, d_{jk}'s, $j = 0, \ldots, J$; $k = 0, \ldots, 2^j - 1$, by performing a discrete wavelet transform (DWT) of the vector of the noiseless data $\boldsymbol{g} = (g(t_1), \ldots, g(t_n))'$. The fast DWT algorithm for finding d_{jk}'s was derived by Mallat (1989). "Noisy" data only allow us to find the MLE estimates of the coefficients, $\hat{\boldsymbol{d}}$, which is the DWT of the vector $\boldsymbol{y}$ of the real data. White noise contaminates all wavelet coefficients $\hat{d}_{jk}$'s equally (the DWT of the white noise vector $\boldsymbol{\varepsilon}$ is also a white noise). However, due to the parsimonious representation by wavelets, it is reasonable to assume that only a few $\hat{d}_{jk}$'s contain information about the real signal while others appear as a consequence of signal's corruption by random noise. The goal is to extract these significant coefficients and to ignore

others. Such an extraction may be done by thresholding the $\hat{d}_{jk}$'s:

$$\hat{d}^*_{jk} = \begin{cases} \hat{d}_{jk}, & |\hat{d}_{jk}| \geq \lambda, \\ 0, & |\hat{d}_{jk}| < \lambda, \end{cases}$$

where λ is the threshold value.

In particular, Donoho and Johnstone (1994) proposed to use the universal *global* threshold $\lambda \sim \sigma\sqrt{2\log n}$ for all j greater than some j_0. The estimator obtained by such thresholding and reconstructing the function from the remaining coefficients can be shown to have risk "close" to the minimal risk corresponding to the optimal (but unknown) thresholding rule. Practically σ is usually unknown, and is estimated from the coefficients at the finest level. It should be emphasized, though, that the threshold depends on the data only through the estimated σ, and for fixed n is otherwise the same for all samples and for all kinds of functions. Donoho and Johnstone further suggested to always include the coefficients of the first "coarse" j_0 levels, even if these coefficients do not pass the thresholding level. In their paper they used $j_0 = 5$. Intuitively, the proper choice of j_0 should depend on the smoothness of the estimated function and on the noise level. It might be argued that j_0 should be greater for oscillating functions but smaller for smooth ones. The examples considered in Section 3 illustrate the fact that the choice of j_0 may strongly affect the corresponding estimators.

From the statistical viewpoint, thresholding, as was also pointed out by Donoho and Johnstone (1994) and Fan (1994), is closely related to hypotheses testing, where each coefficient is tested whether it is zero or not. Model building involves the use of the "significantly different from zero" coefficients. If the results of hypothesis testing should guide us to the choice of the appropriate coefficients, one cannot test each hypotheses at the usual significance level, say 0.05, as if it were the only one tested. With 1023 hypotheses to be tested (for 1024 observations) about 50 would be found (1023×0.05 on the average) significant, even when the representation of the true function needs none. Hence, a stronger control of error is needed, and the most commonly used alternative is to control the probability that no truly zero coefficient enters the model. The control of such a stringent criterion is well-known to reduce power, implying that too few coefficients will enter the model. It is therefore hardly used in practice in other similar problems such as variable selection in regression, or choosing autoregressive terms in time series analysis.

Recently, Benjamini and Hochberg (1995) have suggested the false discovery error rate (FDR) criterion as an appropriate but less stringent alternative in multiple hypothesis testing problems. This paper proposes a statistical procedure for thresholding of wavelet coefficients which is based on the FDR-approach. In a way it controls the expected proportion of incorrectly included coefficients among those chosen for the model. The resulting false discovery rate of coefficients (FDRC) procedure is *inherently adaptive* due to the adaptiveness of the criterion being controlled.

The following section describes the underlying approach and the corresponding testing (thresholding) procedure. Several test-cases are considered in Section 3

where the limitations of non-adaptive thresholding are demonstrated using simulations. In Section 4 the performance of the proposed FDRC-thresholding procedure is contrasted for these examples at various noise levels. Comparing the proposed FDRC-based procedure with the fixed global threshold procedure by evaluating the relative mean-square-error (MSE) across the various test-functions and noise levels, we find the FDR-estimator to enjoy robustness of MSE-efficiency. Finally, while this paper only deals with the estimation of functions on the real line, it is straightforward to extend the suggested thresholding procedure to general R^p and to image reconstruction on R^2 in particular. The details are obvious, and we do not not give them here.

2. The FDR-procedure

We consider here the problem of testing the $m = n - 1$ hypotheses H_{jk}: $d_{jk} = 0$. Of these hypotheses, m_1 are false, or equivalently the corresponding coefficients should be included in the wavelet expansion. The other $m_0 = m - m_1$ coefficients are in fact 0 and ideally should all be set to zero.

Separating the coefficients into those which are zero and those which are not zero may seem an idealization of the real situation: in practice few coefficients of a true function are identically zero, while many more of them will be merely very small. Nevertheless, if we consider a coefficient to be incorrectly included in the model either if it is truely zero and included, or it it is truly of one sign but is included in the model with the wrong sign (directional error), then the case where such coefficients are considered to be exactly zero is the extreme case that needs to be controlled (see Tukey, 1991, for a discussion of this point of view).

Adapting the general idea of Benjamini and Hochberg (1995) we analyse the performance of a thresholding procedure as follows. Let R be the number of coefficients that are not zeroed by the thresholding procedure for a given sample, and are thus included in the representation. Of these R coefficients S are correctly kept in the model and V are erroneously included, $R = V + S$. The error in such a procedure is expressed in terms of the random variable $Q = V/R$ – the proportion of the coefficients included in the representation that should have been zeroed. Naturally we define $Q = 0$ when $R = 0$ since no error of this type can be made when no coefficient is included.

The FDRC can be now defined as the expectation of Q, and thus reflects the expected proportion of erroneously included coefficients among the ones included in the representation.

Following Benjamini and Hochberg (1995) we suggest maximizing the number of included coefficients subject to controlling of the FDRC to some level q.

Two properties of the FDRC are important to note:

(a) If the data are pure noise, i.e., all true coefficients are zero, then each coefficient kept in the model has been *erroneously* included. We have $S = 0$, $V = R$ and, hence, $Q = 0$ if $V = 0$ and $Q = 1$ if $V \geq 1$, yielding $E(Q) = P(V \geq 1)$. Thus,

controlling the FDRC in this case implies the control of the probability of including incorrectly even one coefficient (Bonferroni's approach). Because of this property the traditional levels for significance testing were used, e.g., $q = 0.01$ or $q = 0.05$.

(b) The FDRC increases with an increase in the number of incorrectly chosen coefficients, and decreases as more coefficients are chosen to be included. If a large number of non-zero true coefficients are present, R will tend to be larger and, therefore, the FDRC will tend to be smaller making it easier for more estimated coefficients to be included. Thus, the error rate will respond to the complexity of the estimated function.

Applying the procedure of Benjamini and Hochberg (1995) for wavelet thresholding yields the following procedure.

FDRC thresholding procedure:

(1) For each $\hat{d}_{jk}$ calculate the corresponding two-sided p-value, p_{jk}, testing H_{jk}: $d_{jk} = 0$,

$$p_{jk} = 2(1 - \Phi(|\hat{d}_{jk}|/\sigma)).$$

(2) Order the p_{jk}'s according to their size, $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$, where each of the $p_{(i)}$'s corresponds to some coefficient d_{jk}.

(3) Let i_0 be the largest i for which $p_{(i)} \leq (i/m)q$. For this i_0 calculate $\lambda_{i_0} = \sigma\Phi^{-1}(1 - p_{(i_0)}/2)$.

(4) Threshold all coefficients at level λ_{i_0}.

Benjamini and Hochberg (1995) proved that for independent Gaussian noise in the model (1), the above procedure controls the FDRC at the (unknown) level $(m_0/m)q \leq q$. The procedure also controls the FDRC if the marginal distribution of the noise in model (1) is other than Gaussian, say F, with the only change in the above procedure being to replace Φ by F. Furthermore, as adaptive as it is, the procedure is still conservative, controlling the FDRC below q.

Computational note. Step (4) can be replaced by taking the i_0 coefficients corresponding to the smallest i_0 p-values. Furthermore, since a coefficient can be included in the model only if the corresponding $p_{(i)} \leq q$, it has to be at least larger (in absolute value) than $\lambda_{min} = \sigma\Phi^{-1}(1 - q/2)$. Therefore, the above steps could be performed only for $|\hat{d}_{jk}| \geq \lambda_{\min}$, making large computational savings in sorting, etc.

Note that in a specific sample thresholding is done effectively at some (adaptive) level between $\lambda_{\max} = \sigma\Phi^{-1}(1 - q/2n)$ and $\lambda_{\min} = \sigma\Phi^{-1}(1 - q/2)$. For practically used sample sizes $n = 2^{J+1}$, $J + 1 = 7,8, \ldots, 14$ and the traditional $q = 0.05$, the Donoho–Johnstone global threshold λ satisfies $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$. In fact, over this range $\lambda_{\max}$ is larger than λ by 5–15%. Fig. 1 displays some FDRC-thresholds for $n = 1024$, assuming $\sigma = 1$. While Donoho–Johnstone's $\lambda = 3.723$, if only one (the largest) coefficient enters the representation it should pass the threshold of 4.061. If exactly four coefficients are significant, the corresponding FDRC-threshold is equal to the global DJ-threshold. As more coefficients are included, the corresponding FDRC-threshold is set at lower values.

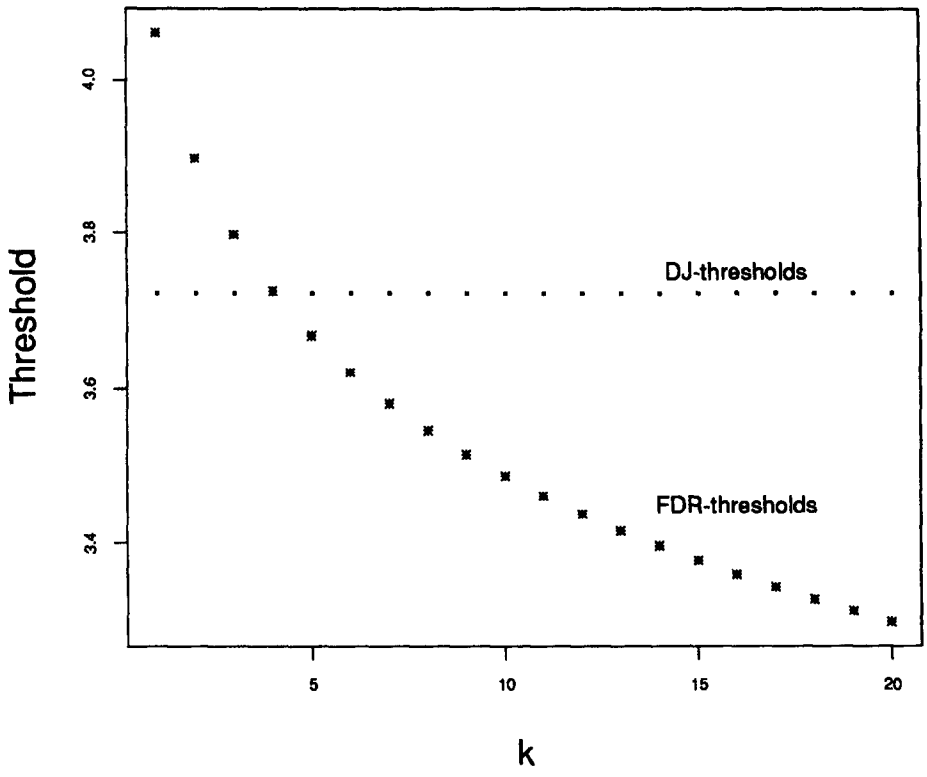

Fig. 1. Threshold for the kth largest coefficient ($n = 1024$, $q = 0.05$).

The procedure can be motivated as a samplewise implementation of the "maximization subject to control" approach. If $p_{(i)}$ corresponds to a potential threshold, exactly i coefficients will pass the threshold and be included in the representation. The expected number of incorrectly included coefficients is $m_0 p_{(i)} \leq m p_{(i)}$, as for these coefficients the estimated p-values are uniformly distributed. Thus, the expected proportion of incorrectly included coefficients among the included ones can be given by $m_0 p_{(i)}/i \leq m p_{(i)}/i$, which we wish to control below q. Selecting as many as possible coefficients means choosing the largest possible i, leading to step (3) of the procedure.

3. Examples

We consider first the performance of three versions of Donoho and Johnstone estimators corresponding to three different thresholding starting levels of j_0: $j_0 = 1$ (DJ1), $j_0 = 3$ (DJ3), the default value in Nason and Silverman (1994), and $j_0 = 5$ (DJ5) used by Donoho and Johnstone. These thresholding procedures were tried with following test cases (see Fig. 2):

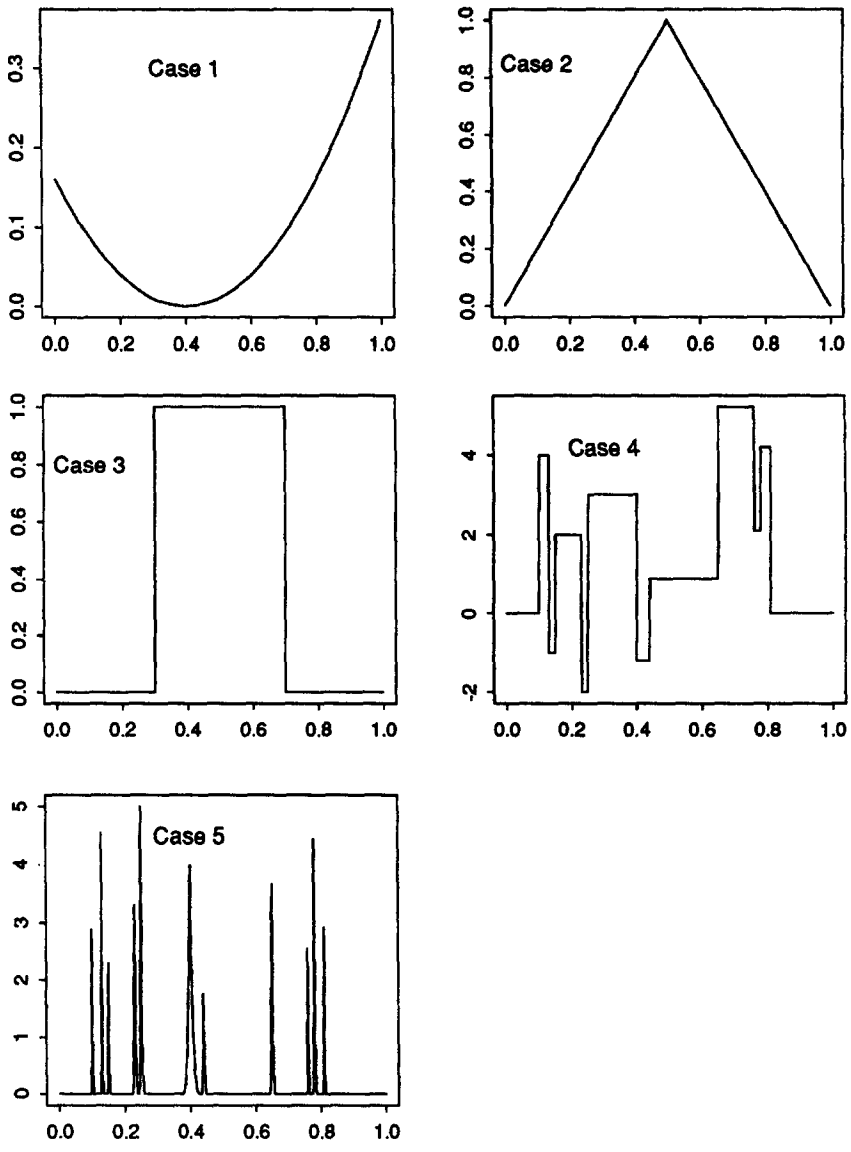

Fig. 2. Test-cases.

(1) $g(t) = (t - 0.4)^2$ (quadratic function).

(2) $g(t) = \min(2t, -2(t-1))$ (triangular function).

(3) $g(t) = (t - 0.3)_+ - (t - 0.7)_+$ (block function).

(4) $g(t) = \sum_j h_j K(t - t_j)$, where $K(t) = (1 + \text{sign}(t))/2$, $(t_j) = (0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81)$, $(h_j) = (4, -5{,}3, -4{,}5, -4.2, 2.1, 4.3, -3.1, 5.1, -4.2)$ (Donoho and Johnstone, 1994, blocks example).

(5) $g(t) = \sum_j h_j K((t - t_j)/w_j)$, where $K(t) = \max((1 - |t|)^4, 0)$, (t_j) are the same as in the previous example, $(h_j) = (4, 5, 3, 4, 5, 4.2, 2.1, 4.3, 3.1, 5.1, 4, 2)$, $(w_j) = (0.005,$

0.005, 0.006, 0.01, 0.01, 0.03, 0.01, 0.01, 0.005, 0.008, 0.005) (similar to Donoho and Johnstone, 1994, bumps example).

For every case we ran simulations with four different values of σ to satisfy the signal-to-noise ratio $\delta = SD(\boldsymbol{g})/\sigma = 7,4,2$ and 1 respectively. 1024 design points were taken equally spaced on [0, 1] and the data were generated for 500 replications of every combination of cases and σ's by adding to $g(t(i/n))$ independent random noise $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

To find the vector of wavelet estimates $\hat{\boldsymbol{g}}$ Mallat's (1989) well-known algorithm of decomposition–reconstruction was used. The Mallat's algorithm is even faster than Fast Fourier Transform and requires only $O(n)$ operations. On the decomposition step we found the wavelet coefficients $\hat{d}_{jk}$'s, where the wavelet basis was formed by the compactly supported mother wavelet D_4 from Daubechies' family (see Daubechies, 1992). The boundary correction is needed for DWT in non-periodic Cases 1 and 2. The correction was based on the symmetric reflection of the function beyond its boundaries. Thresholding $\hat{d}_{jk}$'s and performing the fast inverse DWT of the thresholded coefficients on the reconstruction step we derived the corresponding function estimates. The noise level σ was estimated by the standard deviation of the wavelet coefficients $\hat{d}_{Jk}$ at the finest level J and performed quite satisfactorily. All the programming was done in the statistical package S-Plus using the S WaveThresh software developed by Nason and Silverman (the description may be found in Nason and Silverman, 1994), and the built in normal random numbers generator. The goodness-of-fit of each estimator was measured by its mean squared error $\mathrm{MSE} = n^{-1}\|\boldsymbol{g} - \hat{\boldsymbol{g}}\|_{l^2}$ averaged over all 500 replications. The standard error of MSE was about 0.5–2% of its estimated mean value.

For every test-case, we found the best DJ-estimator among the three ones, i.e., the one achieving the minimum MSE. Then the relative MSE for each estimator was evaluated as $(\min \mathrm{MSE}_j)/\mathrm{MSE}_i$, $i = 1, 2, 3$. From Table 1 one sees that the relative MSE of DJ-estimators for different j_0 varies and depends on smoothness of the function. For example, for relatively smooth functions (Cases 1, 2) DJ3 is highly preferable over DJ5, while in oscillating examples (Cases 4, 5) DJ5 performs much better. Interestingly, although DJ1 is never the best among the three, its relative MSE does not get as low as the 0.676 of DJ5 (see Table 1).

Table 1
Relative MSE of DJ-estimators (averaged over 500 replications), $\delta = 4$

	DJ1	DJ3	DJ5
Case 1	0.798	1.000	0.676
Case 2	0.844	1.000	0.885
Case 3	0.896	0.913	1.000
Case 4	0.912	0.920	1.000
Case 5	0.878	0.878	1.000

4. Why bother with FDRC?

As demonstrated in the previous section, the DJ-thresholding procedure with the prefixed j_0 suffers from the lack of flexibility. In the following we compare its performance with that based on FDR control. Two versions of the latter were used corresponding to FDR of $q = 0.01$ (FDR01) and $q = 0.05$ (FDR05). In order to compare the wavelet estimators we again used the minimax approach. Fixing first the signal-to-noise ratio, δ, and also the test-case, we find the best estimator among the five candidates, i.e., the one achieving minimum MSE. Then the relative MSE for each estimator was evaluated as in the previous section. The results are summarized in the Table 2. For smooth functions (Cases 1, 2) FDR01 performs slightly better than FRD05 (for such functions we would like to be more conservative in including additional coefficients in the representation) but both of them give in to DJ3 which is undoubtedly the best estimator for these cases for all δ's. However, in oscillating examples (Cases 4, 5) FDR05 is highly preferable over FDR01 and DJ3, and even somewhat better than DJ5, the best (for these cases) among DJ-estimators.

Table 2
Relative MSE (averaged over 500 replications)

	FDR05	FDR01	DJ1	DJ3	DJ5
$\delta = 7$					
Case 1	0.859	0.962	0.917	1.000	0.690
Case 2	**0.716**	**0.702**	0.714	0.818	1.000
Case 3	1.000	0.951	0.923	0.923	0.941
Case 4	1.000	0.784	0.658	0.658	0.713
Case 5	1.000	0.747	**0.602**	**0.604**	**0.680**
$\delta = 4$					
Case 1	**0.774**	0.806	0.798	1.000	**0.676**
Case 2	0.799	0.849	0.844	1.000	0.885
Case 3	1.000	0.875	0.857	0.873	0.956
Case 4	1.000	0.848	0.774	0.781	0.849
Case 5	1.000	**0.763**	**0.631**	**0.631**	0.719
$\delta = 2$					
Case 1	**0.761**	0.786	0.796	1.000	**0.527**
Case 2	0.767	0.791	0.793	1.000	0.729
Case 3	0.816	**0.752**	**0.770**	0.867	1.000
Case 4	0.992	0.934	0.913	0.922	1.000
Case 5	1.000	0.857	0.787	**0.793**	0.880
$\delta = 1$					
Case 1	0.760	0.805	0.767	1.000	**0.470**
Case 2	**0.686**	**0.687**	**0.697**	1.000	0.553
Case 3	0.722	0.694	0.718	0.793	1.000
Case 4	0.801	0.733	0.737	**0.765**	1.000
Case 5	1.000	0.857	0.845	0.878	0.959

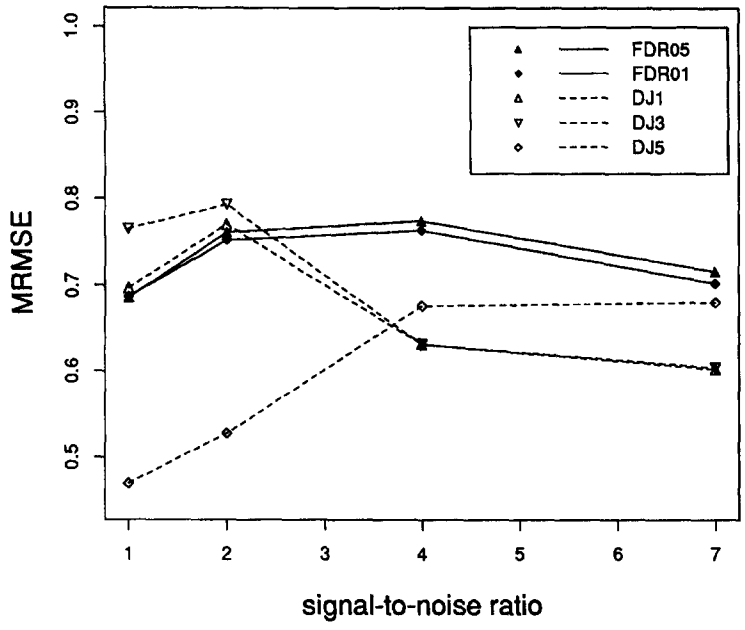

Fig. 3. Minimal relative MSE (MRMSE).

Bolded numbers in Table 2 and Fig. 3 show the minimal relative MSE (MRMSE) of each estimator over all the cases. The MRMSE reflects the loss of effectiveness at the most challenging test-case for each estimator and is a natural measure of its robustness to the test-cases. This measure is further studied as a function of signal -to-noise ratio, which characterizes the robustness of the procedure in face of different noise levels. Fig. 3 implies that the proper choice of j_0 in the Donoho–Johnstone algorithm should depend on the noise level. For large noise (δ small) the optimal j_0 is small since the wavelet coefficients are strongly influenced by noise and we should threshold them starting from the very coarse levels in order to decrease the noise in the reconstruction. For moderate noise "significant" coefficients (especially those at low levels) reflect the real signal and should not be zeroed. In contrast to the behavior of the DJ-estimators, FDR-estimators are much less sensitive to the noise level due to their adaptiveness, and perform quite satisfactorily for all noise levels and test cases studied.

5. Discussion

We have demonstrated the sensitivity of the universal global threshold of Donoho and Johnstone with pre-chosen number of unthresholded coarse levels j_0 to the signal and the noise level. Unlike universal thresholding, the proposed *adaptive* FDRC-procedure has been found to be robust. Some other adaptive thresholding procedures have been suggested recently. Donoho and Johnstone's

(1995) SureShrink procedure combines an adaptive threshold, chosen to minimize the Stein unbiased risk estimator, and the universal global threshold, used when the signal is sparse (or when the signal-to-noise ratio is low). Again the low level coefficients remain untouched by this hybrid method and the number of low levels may be still crucial. Nason (1996) uses cross-validation criterion to choose the threshold that minimizes the predicted MSE. Although SureShrink has some appealing *asymptotic* properties, the current study demonstrates the importance of finite sample properties: any choice of j_0 is asymptotically equivalent. No such theoretical comparison between these methods has been done. For the time being, simulation based comparisons, as demonstrated here, are essential and a comprehensive study of the emerging adaptive methods is desirable.

In conclusion, it might be interesting to note that the FDR-approach for choosing coefficients in the wavelet representation is philosophically very different from traditional methods. Usually, we seek the most compact presentation possible and enrich the model only if some condition is met. Here we try to include coefficients as possible but subject them to a certain control rule. While this approach might be carried over to other problems of model selection, its usefulness should be demonstrated at each case separately, as it was done here.

Acknowledgements

The authors are grateful to Mark Low for a very fruitful discussion of possible applications of the FDR-approach in a wide range of statistical problems including wavelets, and to Philip Stark for helpful comments.

References

Benjamini, Y. and Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. Roy. Statist. Soc. Ser.* B, **57** (1995) 289–300.

Daubechies, I., *Ten Lectures on Wavelets* (SIAM, Philadelphia, PA, 1992).

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

Donoho, D.L. and I.M. Johnstone; Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Statist. Assoc.*, **90** (1995) 1200–1224.

Donoho, D.L., I.M. Johnstone, G. Kerkyacharian and D. Picard, Wavelet shrinkage: asymptotia? (with discussion), *J. Roy. Statist. Soc. Ser.* B, **57** (1995) 301–369.

Fan, J., Test of significance based on wavelet thresholding and Neyman's truncation, Discussion paper 9415, Institut de Statistique, Univ. Catholique de Louvian (1994).

Mallat, S.G., A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattn. Anal. Mach. Intell.*, **11** (1989) 674–693.

Meyer, Y., *Wavelets and operators* (Cambridge Univ. Press, Cambridge, 1992).

Nason, G.P., Wavelet shrinkage using cross-validation, *J. Roy. Statist. Soc., Ser. B*, **58** (1996) 463–479.

Nason, G.P. and B.W. Silverman, The discrete wavelet transform in S, *J. Comp. Graph. Statist.*, **3** (1994) 163–191.

Tukey, J.W., The philosophy of multiple comparison, *Statist. Sci.*, **6** (1991) 100–116.