Spectral Density Estimation via Wavelet Shrinkage

Hong-Ye Gao

MathSoft, Inc.
1700 Westlake Ave. N, Suite 500
Seattle, WA 98109, USA
Tel: (206) 283-8802
Fax: (206) 283-6310
email: gao@statsci.com

June 20, 1996

Spectral Density Estimation via Wavelet Shrinkage

Hong-Ye Gao

Abstract

We study the problem of estimating the spectral density of a stationary Gaussian time series. We use an orthogonal wavelet system whose members are periodic functions and have a finite number of non-zero Fourier coefficients – periodized Meyer wavelets. We apply shrinkage rules to the empirical wavelet coefficients. We show that estimates based on thresholds $t_{j,n} = \lambda_j \log n$ for certain λ_j, with n the sample size, have near-optimal L_2 convergence rates, over any Besov class in a wide range. In some cases, which includes the Bump Algebra, wavelet shrinkage procedures significantly outperform classical linear procedures, such as window methods and AR approximation methods.

Key Words and Phrases: Non-Gaussian Model; Periodized Meyer Wavelets; Shrinkage Estimator. Spectral Density Estimation; Wavelet Transform;

1 Introduction

Suppose $\{X_t\}$ is a Gaussian stationary process with mean zero, covariance function $\gamma(k)$ and spectral density

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{ij\omega}. \tag{1}$$

We want to estimate f from a realization of n observations $X_1, ..., X_n$. For the sake of simplicity, we assume $n = 2^m$ is a dyadic number. In this paper we discuss the following 4-step wavelet shrinkage procedure:

[1] Calculate the unbiased sample covariance function

$$\tilde{\gamma}(k) = \begin{cases} \frac{1}{n-|k|} \sum_{t=1}^{n-|k|} X_t X_{t+|k|} & |k| < n \\ 0 & |k| \geq n \end{cases} \tag{2}$$

[2] Take a periodized wavelet system $\{\psi_{j,k}\}$ and calculate the empirical wavelet coefficients $\{y_{j,k}\}_{j=0,1...m-1,k=0,...2^j-1}$. by

$$y_{j,k} = \sum_l \tilde{\gamma}(l) \widehat{\psi_{j,k}}(l) \tag{3}$$

where $\widehat{\psi_{j,k}}(l)$ denotes the ℓ-th Fourier coefficient of $\psi_{j,k}$.

[3] Apply the soft threshold

$$\delta_t(x) = \mathrm{sgn}(x)(|x| - t)_+ \tag{4}$$

to the empirical wavelet coefficients $\{y_{j,k}\}$, with thresholds $t = t_{j,k,n}$ according to formula

$$t_{j,k,n} = \lambda_{j,n} \sigma_{j,k} \log n \tag{5}$$

where $\sigma_{j,k}$ is the standard deviation of $y_{j,k}$ and $\lambda_{j,n}$ are level-dependent constants.

[4] Invert the wavelet transform, producing an estimate $(\hat{f}_l^*)$ of the spectral density at the Fourier frequencies ω_l.

This procedure falls in the general category of wavelet shrinkage estimates for noisy data. Donoho and Johnstone (1992, 1994, 1995) have discussed the application of such thresholded wavelet transforms for recovering curves from noisy data, where the noise is assumed to have a Gaussian distribution, and they have proposed the level-independent threshold $t_n = \sigma\sqrt{2\log(n)}$, where σ is the noise level. Such methods have also been developed in density estimation: Johnstone et al. (1992); Donoho et al. (1993) proposed the use of level-dependent thresholds $\sqrt{j}$. In general such wavelet shrinkage methods have a number of theoretical advantages, including near-optimal mean-squared error and near-ideal spatial adaptation.

It is easy to apply wavelet shrinkage software designed for Gaussian noise to the periodogram ordinates, and in some cases acceptable reconstructions have been obtained. However, this will not always be the case. The noise in the wavelet coefficients of the periodogram ordinates has a non-Gaussian character; it reaches high levels somewhat more frequently than Gaussian theory would predict. Consequently, thresholds set based on Gaussian theory will not be high enough to completely suppress noise in the coefficients. The thresholding we propose is based on a careful analysis of the non-Gaussian character, and is somewhat larger than Gaussian theory would predict. Furthermore, as we will see later on, the asymptotic variances of wavelet coefficients are "proportional" to the unknown f. Roughly speaking, the larger the $y_{j,k}$, the larger its variance.

For a smooth function class $\mathcal{F}$ which will be specified later, our goal is to choose thresholds $t_{j,n}$ to make

$$R_n(\tilde{f}, \mathcal{F}) = \sup_{f \in \mathcal{F}} E||\tilde{f} - f||^2_{L_2}$$

as small as possible. Our main results are stated in the following theorems:

Theorem 1 *Let $\mathcal{B}^{\sigma}_{p,q}(C)$ be a Besov space (definition given in the next section). For $p \geq 1$ and $\sigma > 1/p$, by choosing threshold*

$$t_{j,k,n} = t_n = \frac{2 \log n}{\sqrt{n}}$$

we have

$$\sup_{f \in \mathcal{B}^{\sigma}_{p,q}(C)} R_n\left(\tilde{f}_{t_n}, f\right) \leq C_1 \frac{(\log n)^2}{n^r} \tag{6}$$

where $r = \frac{2\sigma}{2\sigma+1}$ and constant C_1 depends only on σ, p, q and C.

Theorem 2 *There exists a constant $C_2 > 0$, such that*

$$\inf_{\tilde{f}} \sup_{f \in \mathcal{B}^{\sigma}_{p,q}(C)} R_n\left(\tilde{f}, f\right) \geq \frac{C_2}{n^r}\left(1 + o(1)\right). \tag{7}$$

i.e. the rate in Theorem 1 is nearly optimal.

Note that the classical methods (window method, AR approximation, etc.) can only achieve the rate

$$r' = \frac{\sigma + (1/(p \vee 2) - 1/p)}{\sigma + 1/2 + (1/(p \vee 2) - 1/p)},$$

the optimal rate for linear procedures, Donoho and Johnstone (1992). It is obvious that $r' \leq r$; $r' < r$ when $p < 2$.

In section 2, we will give specifications of the function class $\mathcal{F}$ and wavelet representations of the functions. In section 3, we will discuss in detail our main results. In section 4, we give algorithms and simulation results.

2 Wavelet Representation of Spectral Density

Let $T \subset \mathbf{R_1}$ be an interval and $f, g \in L_2[T]$. Let

$$\langle f, g \rangle = \int_T f(x)g(x)dx$$

be the inner products of f and g, and

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_T e^{-ix\omega} f(x)dx$$

be the Fourier transform of f. In particular, the ℓ-th autocovariance $\gamma(l) = \hat{f}(l)$.

Since spectral densities are defined on the circle, in this paper, we are interested in $T = [-\pi, \pi]$.

2.1 Periodized Meyer Wavelet Basis

The wavelet basis we employ in this chapter is the periodized Meyer wavelet system $\{\phi_{0,0}\} \cup \{\psi_{j,k}\}_{j=0,1,\ldots,\, k=0,1,\ldots,2^j-1}$, which are defined by their Fourier coefficients:

$$\widehat{\phi_{j,k}}(l) = e^{-\frac{2\pi i l k}{2^j}} a_j(l), \qquad \widehat{\psi_{j,k}}(l) = e^{-\frac{2\pi i l(k-1/2)}{2^j}} b_j(l) \tag{8}$$

$l = 0, \pm1, \pm2, \ldots$, where $a_j(x) = 2^{-j/2}a(x/2^j)$, $b_j(x) = 2^{-j/2}b(x/2^j)$ and

$$a(x) = \begin{cases} 1 & |x| \leq \frac{1}{3} \\ \cos(\frac{\pi}{2}\nu(3|x|-1)) & \frac{1}{3} < |x| \leq \frac{2}{3} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$b(x) = \begin{cases} \sin(\frac{\pi}{2}\nu(3|x|-1)) & \frac{1}{3} < |x| \leq \frac{2}{3} \\ \cos(\frac{\pi}{2}\nu(\frac{3|x|}{2}-1)) & \frac{2}{3} < |x| \leq \frac{4}{3} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

with $\nu(\cdot)$ satisfying

$$\nu(x) = \begin{cases} 0 & x \le 0 \\ 1 & x > 1 \end{cases}$$

and

$$\nu(x) + \nu(1-x) = 1$$

Figure 1 gives functions $a(\ )$ and $b(\ )$; these are usually called the "bell" functions; and Figures 2–5 display the corresponding $\phi_{j,k}$ and $\psi_{j,k}$ for $j = 0, 1, 2, 5$ and some choices of ν_p ($p = 0, 1, 2, 3$).

This periodized Meyer system has the following properties:

[1] For each j and k, $\phi_{j,k}, \psi_{j,k}$ are trigonometric polynomials. Indeed,

$$\begin{aligned} \phi_{j,k}(x) &= \frac{1}{\sqrt{2\pi}} \sum_l e^{ilx} \widehat{\phi_{j,k}}(l) = \frac{1}{\sqrt{2\pi}} \sum_{3|l|<2^{j+1}} e^{il\left(x - \frac{2\pi k}{2^j}\right)} a_j(l) \\ \psi_{j,k}(x) &= \frac{1}{\sqrt{2\pi}} \sum_l e^{ilx} \widehat{\psi_{j,k}}(l) = \frac{1}{\sqrt{2\pi}} \sum_{3|l|<2^{j+2}} e^{il\left(x - \frac{2\pi(k-1/2)}{2^j}\right)} b_j(l) \end{aligned}$$

In particular, $\phi_{0,k}(x) \equiv 1/\sqrt{2\pi}$; $\phi_{1,k}(x) = (1 + (-1)^k \sqrt{2}\cos x)/2\sqrt{\pi}$; $\psi_{0,k}(x) = -\cos x/\sqrt{\pi}$; $\psi_{1,k}(x) = ((-1)^{k+1} \sin x - \cos 2x)/\sqrt{2\pi}$.

[2] For each j and k, the Fourier coefficients of $\phi_{j,k}$ and $\psi_{j,k}$ have compact support, i.e. only a finite number of Fourier coefficients are non-zero.

[3] The system is orthogonal, i.e., for $0 \le k < 2^j$ and $0 \le k' < 2^{j'}$,

$$\begin{aligned} \langle \psi_{j,k}, \phi_{j',k'} \rangle &= 0 \\ \langle \psi_{j,k}, \psi_{j',k'} \rangle &= \delta_{k,k'} \end{aligned}$$

Daubechies (1992), §9.3.

[4] For $f \in L_2[T]$,

$$f(\omega) = \beta_0 \phi_{0,0}(\omega) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \alpha_{j,k} \psi_{j,k}(\omega). \tag{11}$$

where $\beta_{j,k} = \langle f, \phi_{j,k} \rangle$ and $\alpha_{j,k} = \langle f, \psi_{j,k} \rangle$.

Remark 1 Meyer wavelets on the line have been studied by many authors, Meyer (1985); Daubechies (1992), pp.117 and 137; Auscher et al. (1992). It is defined by

$$\hat{\phi}(\omega) = a(\omega/2\pi)/\sqrt{2\pi} \quad \text{and} \quad \hat{\psi}(\omega) = e^{i\omega/2} b(\omega/2\pi)/\sqrt{2\pi}.$$

For Meyer wavelets, both ϕ and ψ functions are compactly supported in frequency domain and have exponential decay in time domain. (There are no wavelet bases with compact support in both time and frequency domains, e.g., see pp. 108 of Daubechies, 1992). The regularity of $\hat{\psi}$ is the same as that of ν, which therefore controls the decay of ψ. Figure 6 contains the Meyer wavelets ψ on the line for some choices of ν_p.

Remark 2 The periodized wavelet basis can be derived from wavelet on the real line, (§9.3 of Daubechies, 1992). Suppose both ϕ^* and ψ^* are the father and mother wavelets on the real line, which satisfy

$$|\phi^*(x)| + |\psi^*(x)| \le \frac{C}{(1+|x|)^{1+\varepsilon}}$$

for some positive constants C and ε. (This is true for Meyer wavelets on the line, which decay faster than any rational functions). Let

$$\phi_{j,k}(x) = \frac{1}{\sqrt{2\pi}} \sum_{s=-\infty}^{\infty} \phi^*_{j,k}\left(\frac{x}{2\pi} + s\right), \qquad \psi_{j,k}(x) = \frac{1}{\sqrt{2\pi}} \sum_{s=-\infty}^{\infty} \psi^*_{j,k}\left(\frac{x}{2\pi} + s\right).$$

It is easy to check that $\{\phi_{j,k}\}, \{\psi_{j,k}\}$ are orthogonal bases and for $\ell =$ integer

$$\begin{aligned} \widehat{\phi_{j,k}}(l) &\equiv \frac{1}{\sqrt{2\pi}} \int_T e^{-ilx} \phi_{j,k}(x)dx = \sqrt{2\pi}\widehat{\phi^*}_{j,k}(2\pi l) \\ &= \sqrt{2\pi} 2^{-j/2} e^{-\frac{i2\pi lk}{2^j}} \widehat{\phi^*}\left(\frac{2\pi l}{2^j}\right) \qquad (12) \\ \widehat{\psi_{j,k}}(l) &\equiv \frac{1}{\sqrt{2\pi}} \int_T e^{-ilx} \psi_{j,k}(x)dx = \sqrt{2\pi}\widehat{\psi^*}_{j,k}(2\pi l) \\ &= \sqrt{2\pi} 2^{-j/2} e^{-\frac{i2\pi lk}{2^j}} \widehat{\psi^*}\left(\frac{2\pi l}{2^j}\right) \qquad (13) \end{aligned}$$

and this is exactly (8) in the case of Meyer wavelet. Hence what we call the "periodized Meyer system" actually <u>does</u> arise from periodizing the Meyer wavelets.

Unlike the standard wavelet system, periodized wavelet system can not be generated from a single pair of functions, the father wavelet and mother wavelet, by dilation and translation. However, for each level j, $\psi_{j,k}$ can be generated from $\psi_{j,0}$ by translation.

Remark 3 If $\nu(\cdot)$ is chosen as a polynomial, then we can show that it must be of odd order and the general forms can be derived as

$$\nu_p(x) = x + \sum_{i=0}^{p} a_i(1-2x)^{2i+1}$$

for arbitrary $a_0, ..., a_p$ satisfying $\sum a_i = 0$.

If we further require that $\nu_p \in C^p(\mathbf{R})$, i.e. $\nu_p^{(k)}(0) = 0$, for all $k = 0, 1, ..., p$, then the solution is unique:

$$\nu_p(x) = x^{p+1} \sum_{k=0}^{p} b_k x^k \tag{14}$$

where $(b_0, b_1, ..., b_p)' = A_p^{-1}(1, 0, ..., 0)'$ with

$$A_p = (a_{i,j})_{0 \le i,j \le p} \quad \text{and} \quad a_{i,j} = \begin{pmatrix} p+j+1 \\ i \end{pmatrix}.$$

For example,

$$\begin{aligned}
\nu_0(x) &= x \\
\nu_1(x) &= x^2 (3 - 2x) \\
\nu_2(x) &= x^3 \left(10 - 15x + 6x^2\right) \\
\nu_3(x) &= x^4 \left(35 - 84x + 70x^2 - 20x^3\right) \\
\nu_4(x) &= x^5 \left(126 - 420x + 540x^2 - 315x^3 + 70x^4\right)
\end{aligned}$$

for $0 \le x \le 1$.

Another convenient choice:

$$\nu(x) = \sin^2 \left(\frac{\pi}{4} (1 - \cos(\pi x)) \right).$$

2.2 Decomposition and Reconstruction

Wavelet coefficients under the given wavelet system are

$$\alpha_{j,k} = \int_T f(\omega)\psi_{j,k}(\omega)d\omega = \sum_l \gamma(l)\widehat{\psi_{j,k}}(l)$$

For given resolution level j, notice that under the periodized Meyer wavelet, $\widehat{\psi_{j,k}}(l) = 0$ for $l \notin (2^j/3,\ 2^{j+2}/3)$, $\alpha_{j,k}$ in (15) is a linear combination of *finite* number of $\gamma(l)$'s. So it is feasible to estimate the wavelet coefficients of a spectral density from finite sample and the wavelet coefficients can be calculated by discrete cosine transform of type III (DCT-III), see §2.3 of Rao and Yip (1990):

$$\begin{aligned}
\alpha_{j,k} &= \sum_{\frac{2^j}{3} < |l| < \frac{2^{j+2}}{3}} \gamma(l)\widehat{\psi_{j,k}}(l) = 2^{-j/2} 2 \sum_{l=0}^{N-1} \cos\left(\frac{2\pi l(2k-1)}{N}\right) \gamma(l)\ b\left(\frac{l}{2^j}\right) && (15) \\
\beta_{0,0} &= \gamma(0) && (16)
\end{aligned}$$

where $N = N_j = 2^{j+1}$ and $\gamma(l)$'s are the Fourier coefficients (autocovariance function) of f, i.e. wavelet coefficients are Fourier transforms of "windowed" autocovariances. In particular, $\alpha_{0,k} = -\sqrt{2}\gamma(1)$ and $\alpha_{1,k} = -\gamma(2)$ for all k (notice that $\nu(1/2) \equiv 1/2$ and therefore $b(1/2) = b(1) = \sqrt{2}/2$). It is obvious the $\alpha_{j,k} = \alpha_{j,2^j+1-k}$.

Reconstruction can be done by invert the wavelet coefficients to autocovariances, which can done by the inverse DCT-III, §2.4 of Rao and Yip (1990):

$$\gamma(l) = \beta_0\hat{\phi}(l) + \sum_{j=0}^{\infty} 2^{-j/2}\hat{\psi}\left(\frac{l}{2^j}\right)\left\{\sum_{k=0}^{2^j-1} \alpha_{j,k} e^{-\frac{ikl}{2^j}}\right\} \tag{17}$$

followed by fold and unfold operations.

2.3 Smooth Function Classes Analyzed by Wavelet Bases

Wavelet coefficients can be used to measure the smoothness of a function, i.e. by looking at the wavelet coefficients of a function, we would be able to know how smooth this function is, (smoothness defined based on the modulus of smoothness of the function f), see Meyer (1990), DeVore and Popov (1988), §2 of Donoho and Johnstone (1992) for examples and further references.

The *Besov* space on the circle $\mathcal{B}^{\sigma}_{p,q}(T)$ is a set of functions f in $L_p[T]$ which have smoothness σ. The parameter q gives a finer gradation of smoothness (see the following definition). For example, $\mathcal{B}^1_{1,1}$ = Bump Algebra and $\mathcal{B}^1_{1,1} \subset BV \subset \mathcal{B}^1_{1,\infty}$.

The following facts are the key to our analysis and estimation procedure. The proofs can be developed using e.g. techniques in Donoho and Johnstone (1992), Frazier et al. (1991) and Meyer (1990).

[1] With a wavelet analysis of regularity $r > \sigma$ and $0 < p \leq \infty$,

$$f \in \mathcal{B}^{\sigma}_{p,q}(C) \cap L_p[T] \iff \exists\, A > 0 \text{ such that } \sum_{j=0}^{\infty}\left\{2^{jsq}\left(\sum_{k=0}^{2^j-1}|\alpha_{j,k}|^p\right)^{q/p}\right\} \leq A^q$$

where $\alpha_{j,k}$'s are wavelet coefficients of f under the given wavelet basis and $s = \sigma + 1/2 - 1/p$.

[2] When $\sigma > 1/p$ or $\sigma = 1/p, q \leq 1$,

$$f \in \mathcal{B}^{\sigma}_{p,q}(C) \implies \|f\|_\infty \leq C'$$

for some constant $C' > 0$ which only depends on C, p, q, σ.

3 Wavelet Shrinkage Estimation

Suppose $\mathbf{X} = (X_1, ..., X_n)'$ is a realization of a Gaussian stationary process with spectral density f and covariance function $\gamma(k)$. Let $\tilde{\gamma}(k)$ be the unbiased sample covariance function defined in (2) and $\Gamma = (\gamma(i-j))_{n\times n}$. We assume ψ is the periodized Meyer wavelet and $n = 2^m$. Let $y_{j,k}$ be the empirical wavelet coefficients defined in (3). Then

$$Ey_{j,k} = \sum_{|l|<n} \gamma(l)\hat{\psi}_{j,k}(l) = \alpha_{j,k} \tag{18}$$

for $j = 0, 1, ..., m-1$. Let

$$\sigma^2_{n,j,k} = \mathrm{var}(y_{j,k})$$

Consider the soft threshold $\delta_t(x)$, see (4), which sets to zero data x below t in absolute value, and which pulls other data towards the origin by an amount t. The wavelet coefficients $y_{j,k}$ are subjected to thresholding with a level-and-location-dependent threshold level $t_{n,j,k} = \sigma_{n,j,k}\sqrt{2}\log n$. These simple threshold nonlinearities provide near-minimax behavior when applied to independent and homoscedastic Gaussian data (white noise model, Donoho and Johnstone, 1992a, b). We will show that it works well for some non-Gaussian cases under different shrinkage rule.

Let

$$\hat{\alpha}_{j,k} = \delta_{t_{n,j,k}}(y_{j,k})$$

and consider the estimate

$$\tilde{f}(\omega) = \tilde{\gamma}(0)\phi(\omega) + \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} \hat{\alpha}_{j,k}\psi_{j,k}(\omega) \tag{19}$$

Calculate $\check{\gamma}(l)$ from (17) and consider the estimate

$$\tilde{f}(\omega) = \frac{1}{2\pi}\sum_{|l|<n} \check{\gamma}(l)e^{ij\omega}$$

Remark 1. For the weighted periodogram

$$I(\omega) = \frac{1}{\sqrt{2\pi}}\sum_{s,t=1}^{n} X_s X_t \frac{e^{-i(t-s)\omega}}{n-|s-t|}$$

we have

$$y_{j,k} = \langle I, \psi_{j,k}\rangle_L = \int_T \psi_{j,k}(\omega)I(\omega)d\omega \tag{20}$$

Remark 2. For $j \geq m+2,\ y_{j,k} \equiv 0$.

Remark 3. For $f \in \mathcal{B}^{\sigma}_{p,q}(C)$ and $\sigma > 1/p, p \geq 1$ or $\sigma = 1/p, q \leq 1$, we can show that

$$\sigma^2_{n,j,k} \leq \frac{C'}{n} \tag{21}$$

Remark 4. Gaussian assumption can be removed, as long as asymptotics

$$\frac{2I(\omega_i)}{f(\omega_i)} \text{ i.i.d. } \chi^2_2$$

holds, see Chapter 5 of Brillinger (1981), results in this paper should be true.

4 Technical Details

4.1 One-sample Non-Gaussian Model

Let X be a random variable with $EX = \theta, \mathrm{var}(X) = \sigma^2 \leq 1$ and

$$Z \equiv \frac{X-\theta}{\sigma} \sim F.$$

Suppose there exist constants $A \geq 1$ and $B > 0$, which are independent of θ and σ^2, such that

$$P\left\{|Z| \geq x\right\} = 1 - F(x) + F(-x) \leq Ae^{-Bx}$$

Let

$$\rho(t,\theta) = 2\left(t^2 \wedge \theta^2\right) + 2A\left(t^2 \vee \theta^2\right)\ e^{-B\{(t\vee|\theta|)-(t\wedge|\theta|)\}} \tag{22}$$

and $\delta_t(x)$ be the soft shrinkage defined in (4).

Lemma 1 *For all $t > 1 \vee (2/B)$ and $\theta \in \mathbf{R}$,*

$$E\left(\delta_t(X) - \theta\right)^2 \leq \rho(t,\theta) \tag{23}$$

Proof: First consider the case $\sigma^2 = 1$. For $|\theta| < t$, notice that $t > 1 \vee (2/B)$.

$$\begin{aligned}
E\left(\delta_t(X) - \theta\right)^2 &= \theta^2 + E\left\{\left((Z-t)^2 - \theta^2\right) I_{[Z>t-\theta]}\right\} + E\left\{\left((Z+t)^2 - \theta^2\right) I_{[Z<-t-\theta]}\right\} \\
&= \theta^2 + 2\int_{t-\theta}^{\infty} (z-t)\left\{1 - F(z)\right\} dz + 2\int_{t+\theta}^{\infty} (z-t)F(-z)dz \\
&\le \theta^2 + 2\int_{t-|\theta|}^{\infty} z\left\{1 - F(z) + F(-z)\right\} dz \\
&\le \theta^2 + 2A\int_{t-|\theta|}^{\infty} z\, e^{-Bz} dz \\
&= \theta^2 + 2AB^{-2}\left\{B(t-|\theta|)+1\right\} e^{-B(t-|\theta|)} \le 2\theta^2 + 2At^2 e^{-B(t-|\theta|)}
\end{aligned}$$

For $\theta > t$,

$$\begin{aligned}
E(\delta_t(X) - \theta)^2 &= E(Z-t)^2 + E\left\{\left(\theta^2 - (Z-t)^2\right) I_{[Z<t-\theta]}\right\} + \\
&\quad + E\left\{\left((Z+t)^2 - \theta^2\right) I_{[Z<-t-\theta]}\right\} \\
&= 1 + t^2 - 2\int_{-\infty}^{-t-\theta} (-t-z)F(z)dz + 2\int_{-\infty}^{t-\theta} (t-z)F(z)dz \\
&\le 1 + t^2 + 2A\int_{-\infty}^{-\theta+t} (t-z)e^{Bz}dz \\
&= 1 + t^2 + 2AB^{-2}(B\theta+1)e^{-B(\theta-t)} \le 2t^2 + 2A\theta^2 e^{-B(\theta-t)}
\end{aligned}$$

Similarly, for $\theta < -t$,

$$E(\delta_t(X) - \theta)^2 \le 2t^2 + 2A\theta^2 e^{-B(|\theta|-t)}$$

For general $\sigma^2 \le 1$, we have

$$E(\delta_t(X) - \theta)^2 = \sigma^2 E\left\{\delta_{\frac{t}{\sigma}}\left(\frac{X}{\sigma}\right) - \frac{\theta}{\sigma}\right\}^2 \le \sigma^2 \rho\left(\frac{t}{\sigma},\, \frac{\theta}{\sigma}\right) \le \rho(t,\,\theta)$$

and the proof is completed. □

Now consider the lower bound for general hard and soft shrinkage estimates: $\delta_H X$ and $\delta_S X$ where $\delta_H \in \{0,1\}$ and $\delta_S \in [0,1]$, see Donoho and Johnstone (1992).

Lemma 2 *Suppose F has a density. Let*

$$\eta_H = \eta_H(F) = 1 - \int_{-1}^{1} (1-z^2)dF(z) \tag{24}$$

$$\eta_S = \eta_S(F) = \left\{\int_0^1 z^2 dF(z) + 1 - F(1)\right\} \bigwedge \left\{\int_{-1}^0 z^2 dF(z) + F(-1)\right\}. \tag{25}$$

Then

$$\inf_{\delta_H \in \{0,1\}} E(\delta_H X - \theta)^2 \geq \left(\theta^2 \wedge \sigma^2\right) \eta_H \tag{26}$$

$$\inf_{\delta_S \in [0,1]} E(\delta_S X - \theta)^2 \geq \left(\theta^2 \wedge \sigma^2\right) \eta_S. \tag{27}$$

Proof: Let us first consider the hard shrinkage. Consider function on $[0, \infty)$

$$h(x) = x^2 - \int_{-x}^{x} \left(x^2 - z^2\right) dF(z)$$

and it is easy to see that $h(x)$ is an increasing function in x and $x^{-2}h(x)$ is a decreasing function in x. Notice that δ_H takes only two values and $h(1) = \eta_H$,

$$\begin{aligned} E(\delta_H X - \theta)^2 &\geq E\left\{\theta^2 \wedge (X-\theta)^2\right\} \\ &= \theta^2 + E\left\{[X(X-2\theta)] \wedge 0\right\} \\ &= \theta^2 + E\left\{(\sigma^2 Z^2 - \theta^2) \wedge 0\right\} \\ &= \sigma^2 h\left(|\theta|/\sigma\right) \geq h(1)\left(\theta^2 \wedge \sigma^2\right). \end{aligned}$$

Similarly, for the soft shrinkage, consider functions on $[0, \infty)$

$$\begin{aligned} h_+(x) &= \int_{-x}^{0} z^2 dF(z) + x^2 F(-x) \\ h_-(x) &= \int_{0}^{x} z^2 dF(z) + x^2(1 - F(x)). \end{aligned}$$

It is easy to see that $h_+(x)$ and $h_-(x)$ are increasing functions, $x^{-2}h_+(x)$ and $x^{-2}h_-(x)$ are decreasing functions and $\eta_S = h_+(1) \wedge h_-(1)$.

For $\theta > 0$, under the restriction that $0 \leq \delta \leq 1$,

$$(\delta X - \theta)^2 \geq \begin{cases} 0 & X > \theta \\ (X-\theta)^2 & 0 \leq X \leq \theta \\ \theta^2 & X < \theta \end{cases}$$

Hence,

$$\begin{aligned} E(\delta_S X - \theta)^2 &\geq E\left\{(X-\theta)^2 I_{[0 \leq X \leq \theta]}\right\} + \theta^2 P\left\{X < 0\right\} \\ &= \sigma^2 E\left\{Z^2 I_{[-\frac{\theta}{\sigma} \leq Z \leq 0]}\right\} + \theta^2 P\{X < 0\} \\ &= \sigma^2 h_+\left(\theta/\sigma\right) \geq h_+(1)\left(\theta^2 \wedge \sigma^2\right). \end{aligned}$$

Similarly, for $\theta < 0$,

$$\begin{aligned} E(\delta_S X - \theta)^2 &\geq E\left\{(X-\theta)^2 I_{[\theta \leq X \leq 0]}\right\} + \theta^2 P\{X > 0\} \\ &= \sigma^2 h_-\left(|\theta|/\sigma\right) \geq h_-(1)\left(\theta^2 \wedge \sigma^2\right) \end{aligned}$$

and the proof is completed. □

Remark 1. For all F, $0 < \eta_H < 1$; for all F which has mass at both $(-\infty, 0)$ and $(0, \infty)$, $0 < \eta_S < 1$ (notice that F has variance 1). If F is symmetric, then $\eta_H = 2\eta_S$. In particular,

$$F = \Phi, \quad \eta_H = 1 - \sqrt{\tfrac{2}{e\pi}} = 0.516; \quad \eta_S = \tfrac{1}{2} - \tfrac{1}{\sqrt{2\pi e}} = 0.258.$$

$$F = \text{Double-exponential}(\sqrt{2}), \quad \eta_H = 1 - (\sqrt{2}+1)e^{-\sqrt{2}} = 0.413; \quad \eta_S = 0.207.$$

Remark 2. The lower bounds can (only) be achieved under the help of an oracle. For the hard shrinkage, they are achieved by rules

$$\begin{aligned} \delta_H^* &= \begin{cases} 1 & |\theta| > |X-\theta| \\ 0 & |\theta| \leq |X-\theta| \end{cases} && \text{for the hard shrinkage} \\ \delta_S^* &= \begin{cases} 1 & \frac{\theta}{X} > 1 \\ \frac{\theta}{X} & 0 \leq \frac{\theta}{X} \leq 1 \\ 0 & \frac{\theta}{X} < 0 \end{cases} && \text{for the soft shrinkage.} \end{aligned}$$

(Note that they are not measurable functions of X, i.e. they are not legitimate statistics because they are also functions of unknown θ.) In particular, for the soft shrinkage

$$\delta_t(x) \equiv \operatorname{sgn}(x)(|x|-t)_+ = \begin{cases} x\left(\frac{x-t}{x}\right) & x > t \\ 0 & |x| \leq t \\ x\left(\frac{x+t}{x}\right) & x < -t \end{cases}$$

we have

$$\inf_t E(\delta_t(X) - \theta)^2 \geq \left(\theta^2 \wedge \sigma^2\right) \eta_S. \tag{28}$$

4.2 N-sample Non-Gaussian Model

The following theorem generalizes the results for Gaussian white noise models to non-Gaussian and heteroscedastic cases.

Theorem 3 *Suppose $X_1, ..., X_n$ are random variables with $EX_i = \theta_i$ and var$(X_i) = \sigma_i^2 \le \varepsilon^2$. Let*

$$Z_i = \frac{X_i - \theta_i}{\sigma_i} \sim F_i$$

and there exist constants A and B, such that

$$\sup_{1 \le i \le n} P\{|Z_i| \ge x\} \le Ae^{-Bx}.$$

Let

$$r_n(\boldsymbol{\theta}, \boldsymbol{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 + \sum_{i=1}^{n}\left(\theta_i^2 \wedge \sigma_i^2\right).$$

Take threshold $t_n = B^{-1}\log n$, then for all $\boldsymbol{\theta} \in \mathbf{R}^n$,

$$\frac{\sum_{i=1}^n E\left\{\delta_{t\sigma_i}(X_i) - \theta_i\right\}^2}{r_n(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)} \le C_1(\log n)^2 \tag{29}$$

and

$$\sum_{i=1}^{n} E\left\{\delta_{t\sigma}(X_i) - \theta_i\right\}^2 \le C_1(\log n)^2\left\{\varepsilon^2 + \sum_{i=1}^{n}\left(\theta_i^2 \wedge \varepsilon^2\right)\right\}. \tag{30}$$

I.e., the risk for diagonal shrinkage is within a factor $(\log n)^2$ of the ideal risk for the shrinkage rule, and the latter is within a constant of the over all optimal risk, Donoho and Johnstone (1994).

Furthermore, if $\inf_i \eta_S(F_i) \ge c$, then

$$\sum_{i=1}^{n} E(\delta_{t\sigma}(X_i) - \theta_i)^2 \ge c'\left\{\varepsilon^2 + \sum_{i=1}^{n}\left(\theta_i^2 \wedge \varepsilon^2\right)\right\}. \tag{31}$$

Proof: Let $\rho(t, \theta)$ be defined as (22). Since $\rho(t, \theta)$ is symmetric in θ, it suffices to consider the case $\theta \ge 0$. From

$$\frac{\partial \rho(t,\theta)}{\partial \theta} = \begin{cases} 4\theta + 2ABt^2 e^{-B(t-\theta)}, & 0 < \theta < t \\ -4A\theta(B\theta - 2)\, e^{-B(\theta - t)}, & \theta > t \end{cases}$$

we can see that $\rho(t, \theta)$ increases in θ when $\theta \in (0, t)$ and $\rho(t, \theta)$ decreases in θ when $\theta > t$ (notice that $t > 2/B$). Therefore,

$$\sup_{\theta > 0} \rho(t, \theta) = 2(A+1)t^2 \le 4At^2$$

and this implies that

$$\sup_{\theta \ge 1}\left\{\frac{\rho(t,\theta)}{n^{-1} + 1 \wedge \theta^2}\right\} \le 4At^2. \tag{32}$$

For $0 \le \theta < 1 \ \ (< t)$,

$$\rho(t, \theta) = 2\theta^2 + \rho(t, 0) e^{B\theta} \le 2\theta^2 + \rho(t, 0) e^B.$$

Choose t such that $e^B \rho(t, 0) = 2At^2 e^{-B(t-1)} \ge 2/n$, then

$$\sup_{0 \le \theta < 1} \left\{ \frac{\rho(t, \theta)}{n^{-1} + (1 \wedge \theta^2)} \right\} \le \sup_{0 \le \theta < 1} \left\{ \frac{2\theta^2 + \rho(t, 0) e^B}{n^{-1} + \theta^2} \right\} = n e^B \rho(t, 0) = 2Ae^B n t^2 e^{-Bt}. \quad (33)$$

Combining (32) and (33),

$$\Lambda_t \equiv \sup_{\theta \in \mathbf{R}} \left\{ \frac{\rho(t, \theta)}{n^{-1} + 1 \wedge \theta^2} \right\} \le 4At^2 \left\{ \left(n e^B e^{-Bt}/2 \right) \vee 1 \right\}.$$

By Lemma 1, since $EX_i/\sigma_i = \theta_i/\sigma_i$ and $\mathrm{var}(X_i/\sigma_i) = 1$,

$$\begin{aligned}
\sum_{i=1}^n E \left\{ \delta_{t\sigma_i}(X_i) - \theta_i \right\}^2 &= \sum_{i=1}^n \sigma_i^2 E \left\{ \delta_t \left(\frac{X_i}{\sigma_i} \right) - \frac{\theta_i}{\sigma_i} \right\}^2 \le \sum_{i=1}^n \sigma_i^2 \rho\left(t, \theta_i/\sigma_i\right) \\
&= \sum_{i=1}^n \frac{\rho(t, \theta_i/\sigma_i)}{n^{-1} + 1 \wedge (\theta_i^2/\sigma_i^2)} \left\{ \frac{\sigma_i^2}{n} + (\theta_i^2 \wedge \sigma_i^2) \right\} \\
&\le \Lambda_t \left\{ n^{-1} \sum_{i=1}^n \sigma_i^2 + \sum_{i=1}^n \left(\theta_i^2 \wedge \sigma_i^2 \right) \right\}.
\end{aligned}$$

In particular, for $t = t_n = B^{-1} \log n$, $\Lambda_t \le C (\log n)^2$. The proof is completed. □

Lemma 3 *Let*

$$\begin{aligned}
P(t, \varepsilon) &= \sum_{j=0}^{\infty} \sum_{k=0}^{2^j - 1} \left(t_{j,k}^2 \wedge \varepsilon^2 \right) \\
Q_{s,p,q}(t) &= \sum_{j=0}^{\infty} 2^{jsq} \left(\sum_{k=0}^{2^j - 1} |t_{j,k}|^p \right)^{q/p} \\
J(\varepsilon, A) &= J_{\sigma,p,q}(\varepsilon, A) = \sup_{t:\, Q_{\sigma,p,q}(t) \le A^q} P(t, \varepsilon).
\end{aligned}$$

Then,

$$J(\varepsilon, A) \le A_r \varepsilon^r, \qquad \text{as } \varepsilon \to 0,$$

where $r = \frac{s + 1/p - 1/2}{s + 1/p} > 0$ *and* A_r *is a constant.*

Proof can be found in §4.4 of Donoho and Johnstone (1992).

4.3 Properties of $\sigma^2_{n,j,k}$

The variances of wavelet coefficients are the key for shrinkage estimator. When X_i's are iid $N(0,\sigma^2)$ (i.e., $f(\omega)\equiv\sigma^2$), we have

$$\sigma^2_{n,j,k} = 2\sigma^4 \sum_{|l|<n} \frac{|\hat{\psi}_{j,k}(l)|^2}{n-|l|}.$$

Lemma 4 *For fixed* $0 \le j < m$,

$$n\sigma^2_{n,j,k} \to \sigma^2_{j,k} \tag{34}$$

where $\sigma^2_{j,k} = 2\int f^2(\omega)\psi^2_{j,k}(\omega)d\omega$. *And if* $0 < M_1 \le f \le M_2 < \infty$, *then*

$$2M_1 \le \sigma^2_{j,k} \le 2M_2.$$

Proof: The first two expressions come directly from the definitions and the fact that the Meyer wavelet has compact support in the frequency domain.

From (20), $y_{j,k}$'s can be expressed as quadratic forms of independent standard Gaussian variables. Let $B_{j,k} = (b_{s,t})_{1\le s,t\le n}$ where

$$b_{s,t} = b_{s,t}(j,k) = \frac{\hat{\psi}_{j,k}(s-t)}{n-|s-t|}$$

and $\boldsymbol{\xi} = (\xi_1,...,\xi_n)'$ be i.i.d. $N(0,1)$ variates, then

$$y_{j,k} = \mathbf{X}'B_{j,k}\mathbf{X} = \boldsymbol{\xi}'\Gamma^{1/2}B_{j,k}\Gamma^{1/2}\boldsymbol{\xi} = \sum_{i=1}^n \lambda_i\xi_i^2 \tag{35}$$

where $\lambda_1,...,\lambda_n$ are eigenvalues of $\Gamma^{1/2}B_{j,k}\Gamma^{1/2}$.

Therefore,

$$\begin{aligned}
\sigma^2_{n,j,k} &= 2\sum_{i=1}^n \lambda_i^2 = 2\,tr\left\{\left(\Gamma^{1/2}B_{j,k}\Gamma^{1/2}\right)^*\left(\Gamma^{1/2}B_{j,k}\Gamma^{1/2}\right)\right\} = 2\,tr\left\{B^*_{j,k}\Gamma B_{j,k}\Gamma\right\} \\
&= 2\sum_{s,t,u,v=1}^n \frac{\gamma(s-u)\gamma(t-v)\hat{\psi}_{j,k}(s-t)\overline{\hat{\psi}_{j,k}(u-v)}}{(n-|s-t|)(n-|u-v|)} \\
&= 2\sum_{l,m=-n+1}^{n-1} \frac{\hat{\psi}_{j,k}(l)\overline{\hat{\psi}_{j,k}(m)}}{(n-|l|)(n-|m|)} \sum_{s=-n+1+l_+-m_-}^{n-1-m_++l_-} \gamma(s)\gamma(s+l-m) \\
&\qquad \{n-|l|-(m_+-l_++s)_+ + (m_--l_-+s)_-\}
\end{aligned} \tag{36}$$

where $x_- = x\wedge 0,\ x_+ = x\vee 0$.

From (36), as $n \to \infty$,

$$
\begin{aligned}
n\sigma^2_{n,j,k} &\to 2 \sum_{l,m=-\infty}^{\infty} \hat{\psi}_{j,k}(l)\overline{\hat{\psi}_{j,k}(m)} \sum_{s=-\infty}^{\infty} \gamma(s)\gamma(s+l-m) \\
&= 2\int_{-\infty}^{\infty} f^2(\omega)\psi^2_{j,k}(\omega)d\omega \equiv \sigma^2_{j,k}.
\end{aligned}
$$

□

4.4 Tail Probability of $y_{j,k}$

Lemma 5 *With above notation, for all $x \geq 0$,*

$$
\sup_{(\lambda_i):\ \sum_{i=1}^n \lambda_i^2 = 1} P\left\{ \left| \sum_{i=1}^n \lambda_i(\xi_i^2 - 1) \right| > x \right\} \leq 6e^{-x/2}. \tag{37}
$$

Consequently,

$$
\sup_{0 \leq j < m,\ 0 \leq k < 2^j} P\left\{ \left| \frac{y_{j,k} - \alpha_{j,k}}{\sigma_{n,j,k}} \right| \geq x \right\} \leq 6e^{-x/\sqrt{2}}. \tag{38}
$$

Proof: Let

$$
J(x;\lambda_1,...,\lambda_n) = P\left\{ \sum_{i=1}^n \lambda_i \left(\xi_i^2 - 1\right) \leq x \right\}.
$$

By Theorem 2 of Bakirov (1989), for $x > 2 + \sqrt{2}$,

$$
\inf_{(\lambda_i):\ \sum_{i=1}^n \lambda_i^2 = 1} J(x;\lambda_1,...,\lambda_n) = P\left\{\xi_1^2 - 1 \leq x\right\} = 2\Phi\left(\sqrt{1+x}\right) - 1.
$$

Since

$$
J\left(x;\lambda_1,...,\lambda_n\right) = 1 - J(-x;-\lambda_1,...,-\lambda_n)
$$

for $x < -2 - \sqrt{2}$,

$$
\begin{aligned}
\sup_{(\lambda_i):\ \sum_{i=1}^n \lambda_i^2 = 1} J(x;\lambda_1,...,\lambda_n) &= 1 - \inf_{(\lambda_i):\ \sum_{i=1}^n \lambda_i^2 = 1} J(-x;-\lambda_1,...,-\lambda_n) \\
&= 1 - \inf_{(\lambda_i):\ \sum_{i=1}^n \lambda_i^2 = 1} J(-x;\lambda_1,...,\lambda_n) \\
&= 2\left\{1 - \Phi\left(\sqrt{1-x}\right)\right\}.
\end{aligned}
$$

Therefore, for $x > 2+\sqrt{2}$,

$$
\begin{aligned}
&\sup_{(\lambda_i):\, \sum_{i=1}^n \lambda_i^2 = 1} P\left\{\left|\sum_{i=1}^n \lambda_i(\xi_i^2 - 1)\right| > x\right\} = \\
&= \sup_{(\lambda_i):\, \sum_{i=1}^n \lambda_i^2 = 1} \{1 - J(x;\lambda_1,...,\lambda_n) + J(-x;\lambda_1,...,\lambda_n)\} \\
&\le \ 4\left\{1-\Phi\left(\sqrt{1+x}\right)\right\} \le \frac{4\phi\left(\sqrt{1+x}\right)}{\sqrt{1+x}} = \frac{4}{\sqrt{2\pi e}}\frac{e^{-x/2}}{\sqrt{1+x}} \le 0.461\ e^{-x/2}
\end{aligned}
$$

where ϕ is temporarily the density function of the standard normal. Since $\exp(1+\sqrt{2}/2) = 5.513... < 6$, the above inequality holds for all $x \ge 0$ if we replace the constant 0.461 by 6. The Lemma is proved. □

Remark: From De Jong (1987),

$$
\sum_{i=1}^n \lambda_i \left(\xi_i^2 - 1\right) \Rightarrow\ N(0,1)
$$

iff $\max_i |\lambda_i| \to 0$. Since the Meyer wavelet ψ decays fast (exponentially), empirical coefficients at finer levels do not have good normal approximation.

4.5 Proof of Theorem 1

Under the L_2-norm, by Parseval equation,

$$
E||\tilde{f} - f||_{L_2}^2 = E\left\{\tilde{\gamma}(0) - \gamma(0)\right\}^2 + \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} E(\hat{\alpha}_{j,k} - \alpha_{j,k})^2 + \sum_{j=m}^{\infty}\sum_{k=0}^{2^j-1}\alpha_{j,k}^2
$$

and $E(\tilde{\gamma}(0) - \gamma(0))^2 = O(n^{-1})$.

From Lemma 5, $B = 1/2$. From (21), (30) and Lemma 3,

$$
\begin{aligned}
\sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} E\left\{\delta_{t_n}(y_{j,k}) - \alpha_{j,k}\right\}^2 \ &\le\ C_1(\log n)^2\left\{\frac{C'}{n} + \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1}\left(\alpha_{j,k}^2 \wedge \frac{C'}{n}\right)\right\} \\
&\le\ C_1(\log n)^2 \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1}\left(\alpha_{j,k}^2 \wedge \frac{C'}{n}\right)(1+o(1)) \\
&\le\ C_1\frac{(\log n)^2}{n^r}(1+o(1)).
\end{aligned}
$$

□

4.6 Proof of Theorem 2

By Theorem 3,

$$\sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} E\left\{\delta_{t_j}(y_{j,k}) - \alpha_{j,k}\right\}^2 \geq \eta_S \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1}\left(\alpha_{j,k}^2 \wedge \frac{\sigma_{j,k}^2}{n}\right).$$

5 Acknowledgements

This work was supported by NASA contract NCA2-488, by NSF DMS-84-51753 and by NSF DMS-92-09130. This work is part of the author's Ph.D dissertation at the University of California, Berkeley, under the supervision of Professor David L. Donoho, whose generous guidance and encouragement are gratefully acknowledged and appreciated. Many thanks to Eric Kolaczyk, who made available to me his software tools for inverting Periodic Meyer wavelet transforms.

References

Auscher, P., Weiss, G., and Wickerhauser, M. V. (1992). Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets. In Chui, C. K., editor, *Wavelets: a tutorial in theory and applications*, pages 237–256. Academic Press, Inc., San Diego, CA.

Bakirov, N. K. (1989). Extrema of the Distributions of Quadratic Forms of Gaussian Variables. *Theory Probab. Appl.*, 34.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

Daubechies, I. (1992). *Ten lectures on wavelets.* Society for industrial and applied mathematics, Philadelphia, PA.

De Jong, P. (1987). A Central Limit Theorem for Generalized Quadratic Forms. *Probability Theory and Related Fields*, 75:261–277.

DeVore, R. and Popov, V. (1988). Interpolation of Besov Spaces. *Trans. Am. Math. Soc.*, 305:397–414.

Donoho, D. L. and Johnstone, I. M. (1992). Minimax Estimation via Wavelet Shrinkage. Technical Report 402, Stanford University.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika*, 81:425–455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1993). Density estimation by wavelet thresholding. Technical report, Stanford University.

Frazier, M., Jawerth, B., and Weiss, G. (1991). Littlewood-Paley Theory and the Study of Function Spaces. In *NSF-CBMS Regional Conf. Ser in Mathematics*, volume 79, Providence, RI. American Math. Soc.

Gao, H.-Y. (1995). Choice of Thresholds for Wavelet Shrinkage Estiamte of the Spectrum. *Journal of Time Series analysis*. Accepted pending revisions.

Johnstone, I., 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.

Kolaczyk, E. (1994). *WVD Solution of Inverse Problems*. PhD thesis, Stanford University.

Meyer, Y. (1985). *Principe d'Incertitute, Bases Hilbertiennes et Algèbres D'opérateurs*, volume No. 662. Séminaire, Bourbaki. 1985–1986.

Meyer, Y. (1990). *Ondelettes et opérateurs*. Hermann, Éditeurs des sciences et des arts, 293 rue Lecourbe, 75015 Paris.

Rao, K. R. and Yip, P. (1990). *Discrete Cosine Transform*. Academic Press, Inc., San Diego, CA.

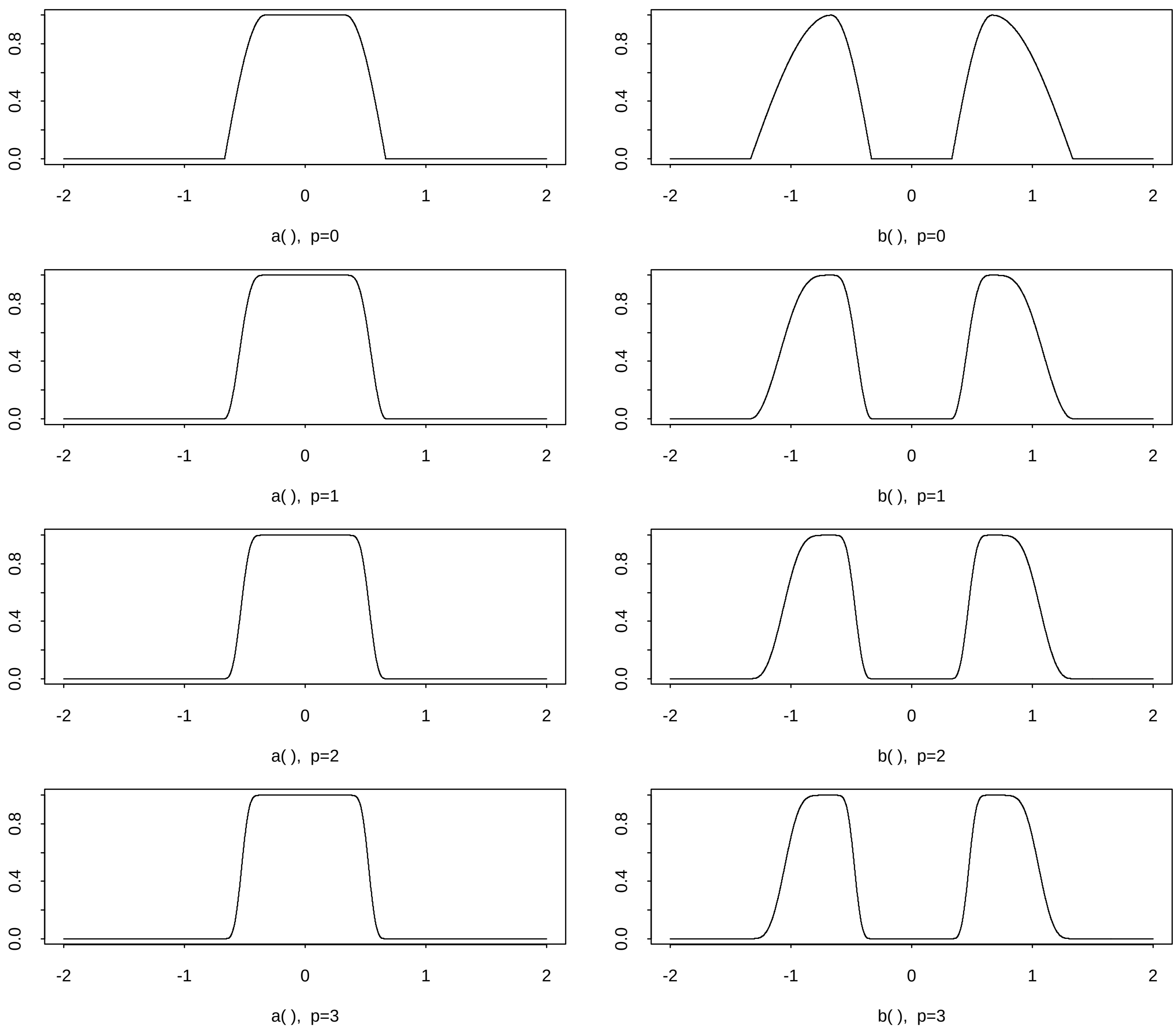

Figure 1: Bell functions for Meyer wavelets.

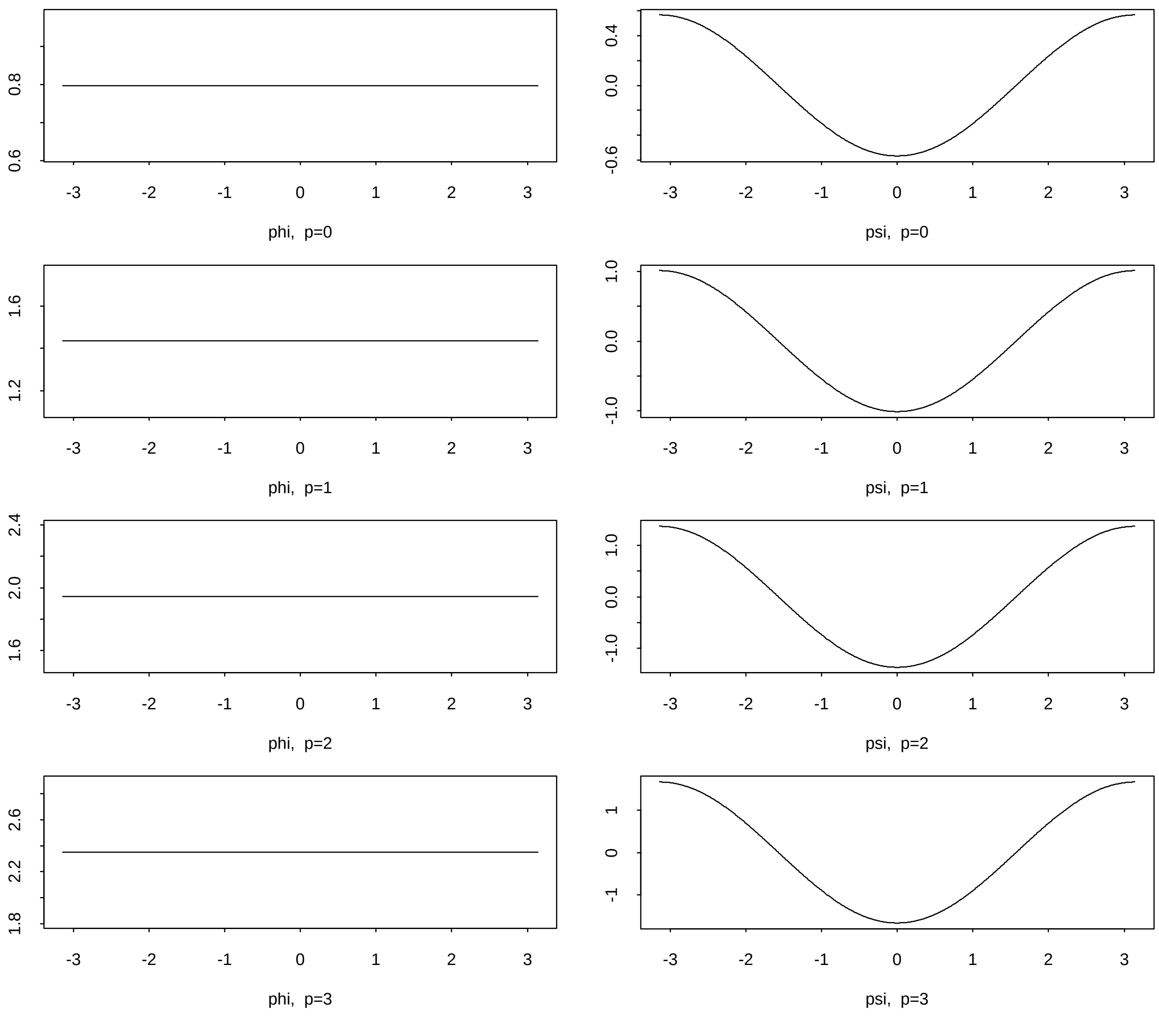

Figure 2: Periodized Meyer wavelets for $j = 0$.

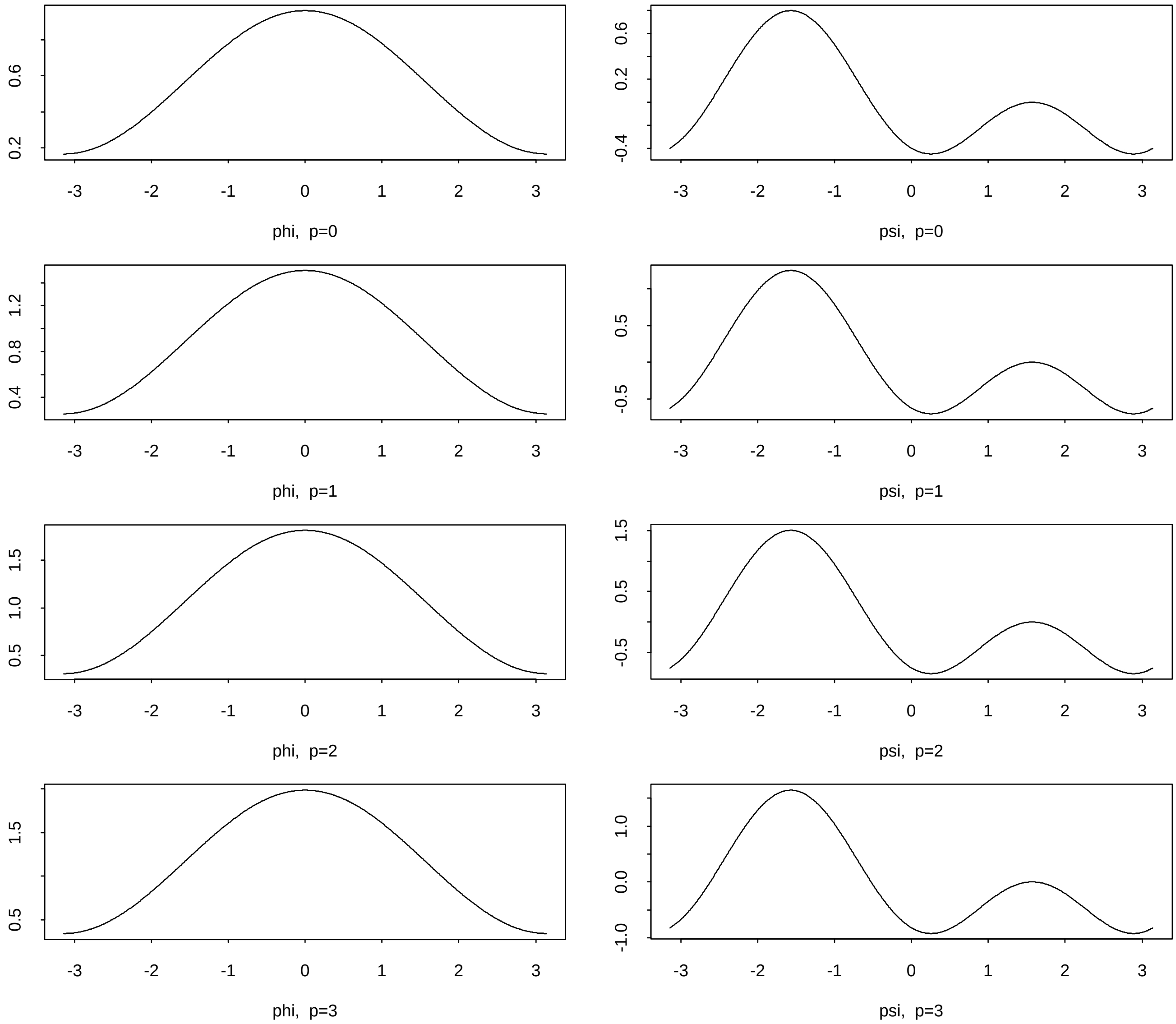

Figure 3: Periodized Meyer wavelets for $j = 1$.

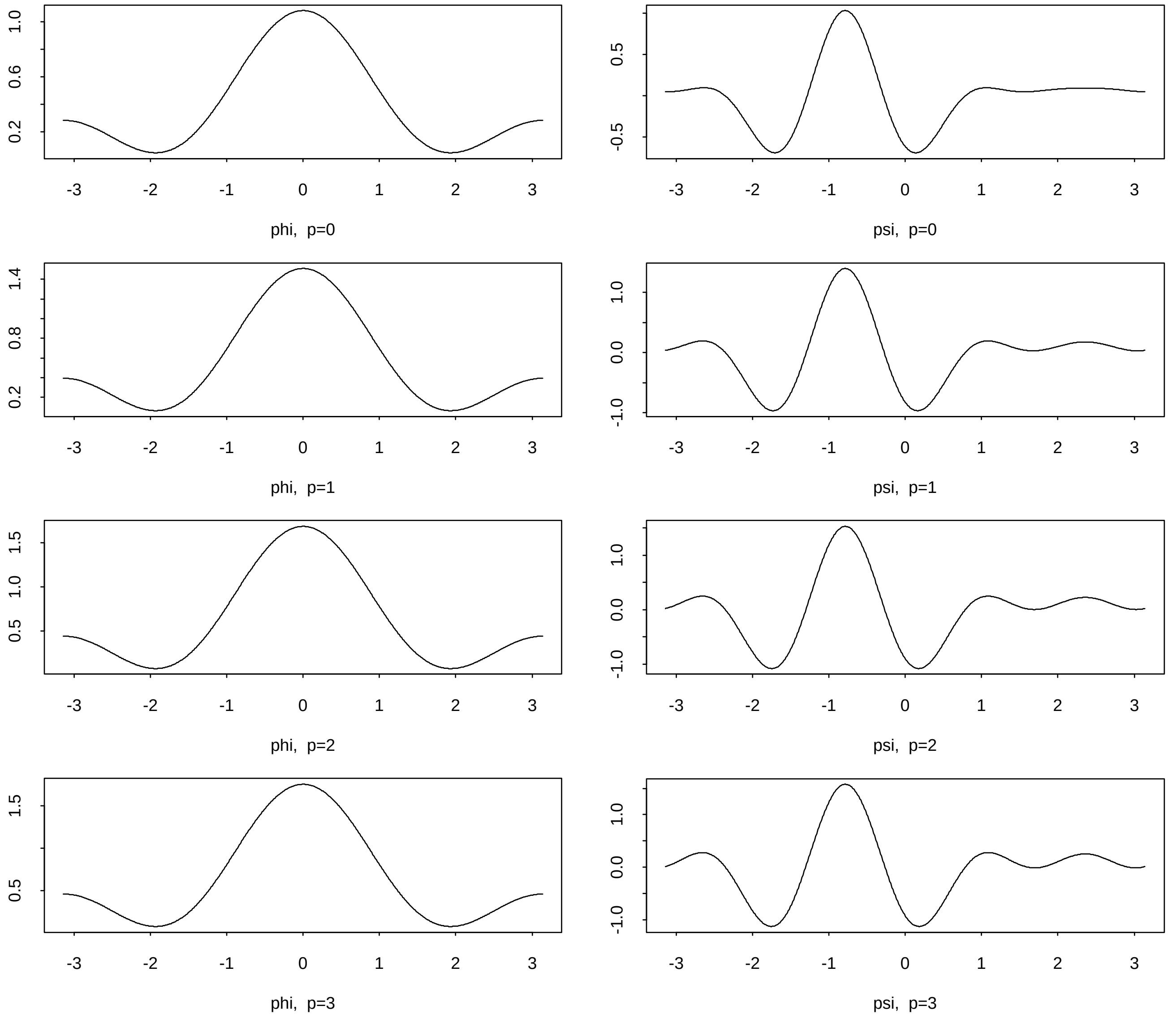

Figure 4: Periodized Meyer wavelets for $j = 2$.

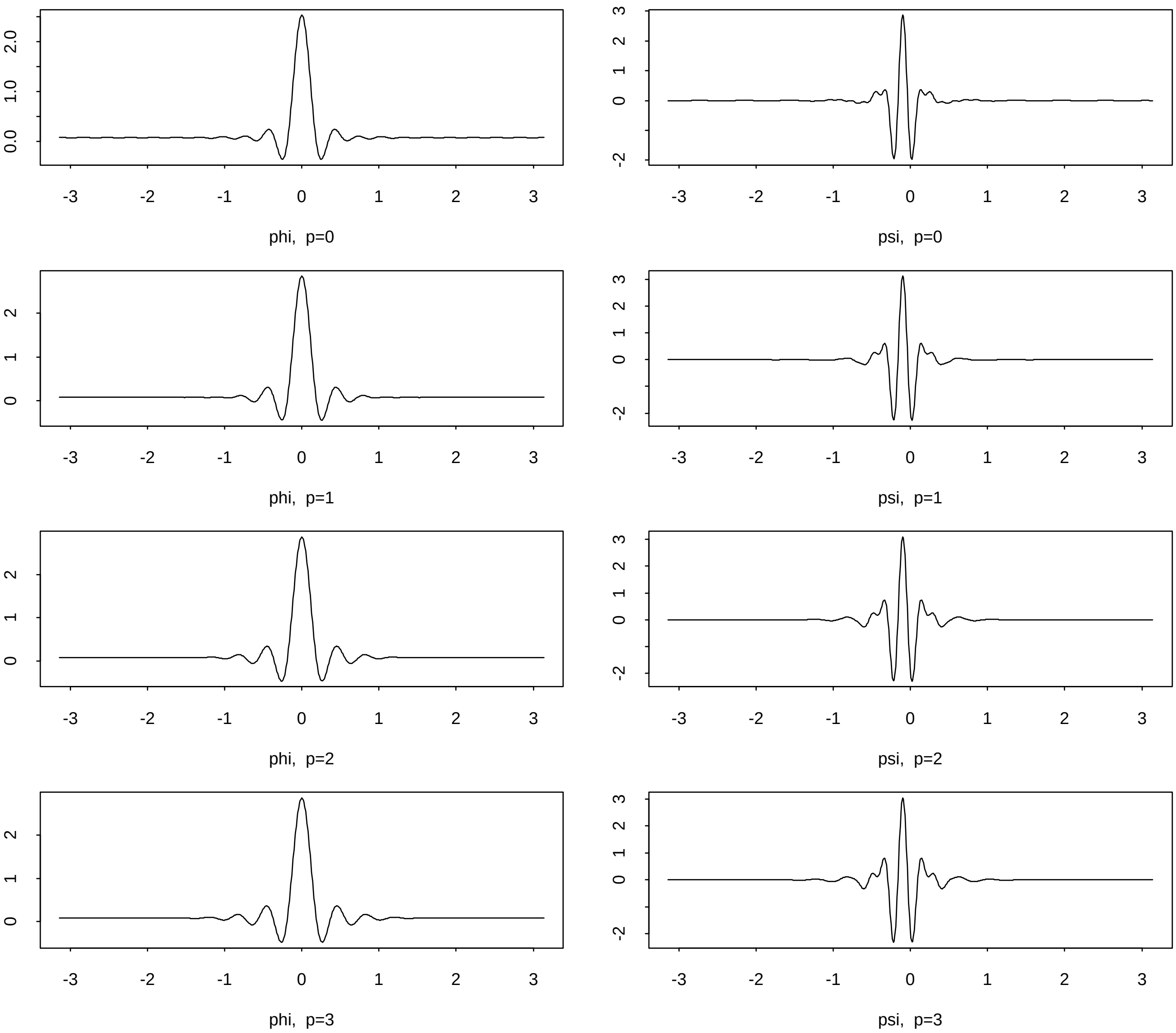

Figure 5: Periodized Meyer wavelets for $j = 5$.

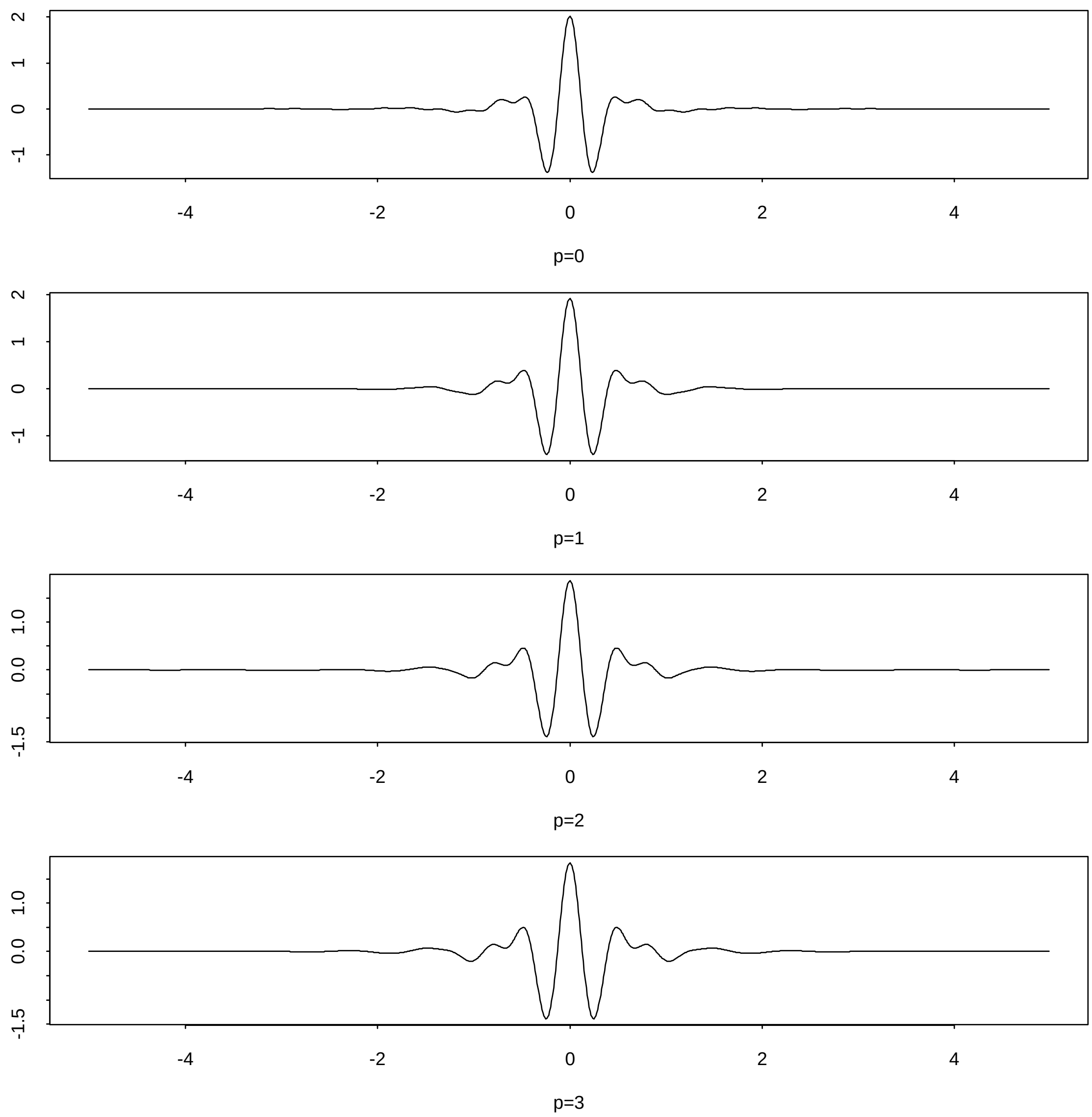

Figure 6: Meyer wavelets on the real line.