# CONTINUOUS NON-NEGATIVE WAVELETS AND THEIR USE IN DENSITY ESTIMATION

GILBERT G. WALTER AND XIAOPING SHEN

ABSTRACT. It was shown by Janssen in 1993 that there are no continuous non- negative orthogonal scaling functions. For density estimation, we would like the estimators themselves to be densities, and hence non-negative. Thus the usual linear projection or the threshold estimators won't work without modification. In this paper, we introduce smooth non-negative wavelet estimators. The density estimators based on them are themselves density functions. These are applied to two numerical examples and compared to the usual wavelet estimators in these examples. An associated noise reduction method is considered.

## 1. Introduction.

In the problem of non-parametric density estimation, a continuous probability density function $f(x)$ must be estimated from with a sample $X_1, X_2, .., X_N$ of this density. A first approximation is the empiric distribution

$$f^*(x) = \frac{1}{N}\sum_{i=1}^{N} \delta(x - X_i).$$

This is the generalized derivative of the empiric distribution function

$$F^*(x) = \frac{1}{N}\sum_{i=1}^{N} H(x - X_i)$$

where $H(x)$ is the Heaviside function (unit step function). One difficulty is that $f^*(x)$ is not very smooth; in fact is not even a function. The density estimation problem is to smooth it out to get a density estimator that looks about like $f(x)$. The problem has a long history and has many solutions, the simplest of which is the Histogram. It may be constructed by beginning with the indicator function $\chi(x)$ of the unit interval [0,1) and letting

$$q(x, y) = \chi(y - [x])$$

*Date*: September 25, 1997.

1991 *Mathematics Subject Classification.* Primary 60E05, 62F10.

*Key words and phrases.* positive kernel, kernel estimation, wavelet, probability density.

where $[x]$ is the largest integer $\leq x$. We then change the scale by a multiple of 2 to get

$$q_m(x,y) = 2^m q(2^m x, 2^m y).$$

This is a *delta sequence, i.e.,* a sequence which satisfies

$$q_m(x,y) \longrightarrow \delta(x-y).$$

It is, in fact, a positive delta sequence (the best kind which leads to uniform convergence)[W-B] . The histogram with bin width $2^{-m}$. is then given by

$$\widehat{f}_m(x) = \int_{-\infty}^{\infty} q_m(x,y) f^*(y) dy = \frac{1}{N} \sum_{i=1}^{N} q_m(x, X_i),$$

which of course is not the usual way of defining it. This function $q_m(x,y)$ is the reproducing kernel of the subspace $\mathrm{V}_m$ associated with the Haar wavelet, and $\widehat{f}_m$ is the linear density estimator associated with it.

In common with other positive delta sequences, the Bias satisfies

$$E(\widehat{f}_m(x) - f(x)) = O(2^{-m}) \tag{1}$$

for continuous densities as $m \to \infty$. Hence the mean square error is

$$MSE = Var + Bias^2 = O(2^m/N) + O(2^{-2m}) = O(N^{-2/3}) \tag{2}$$

for a proper choice of $m$. One difficulty is that while this is a *bonafide* density, it is not continuous and does not converge as rapidly as it could.

Kernel estimators are almost as simple. Let $k(x)$ be a bounded density function itself and now define $q(x,y) = k(x-y)$, and as before

$$q_m(x,y) = 2^m q(2^m x, 2^m y).$$

This again is a positive delta sequence and the estimator is given by

$$\widehat{f}_m(x) = \int_{-\infty}^{\infty} q_m(x,y) f^*(y) dy = (k_m * f^*)(x) = \frac{1}{N} \sum_{i=1}^{N} q_m(x, X_i) \tag{3}$$

with the same MSE as the histogram in (1). Favorite kernels are $k(x) = \chi(x)$ and $k(x) = e^{-x^2/2}/\sqrt{2\pi}$. Again the estimator is a density function itself but now may be chosen to be smooth. But the convergence is still relatively slow.

Faster rates of convergence are obtained with orthogonal series estimators. We start from $\{\varphi_n(x)\}$, an orthounormal basis of $\mathrm{L}^2(\mathrm{I})$ where I is either an interval or the entire real line. Examples are the trigonometric system on $[0, 2\pi]$ or the Hermite system on $(-\infty, \infty)$. In this case we let

$$q_m(x,y) = \sum_{n=0}^{2^m} \varphi_n(x) \varphi_n(y).$$

Then again $\{q_m(x,y)\}$ is a delta sequence, but is no longer a positive one. This enables us to get faster convergence, but at some other cost. In fact

if, for example, $f \in C^r, r > 1$, then the estimator $\widehat{f}_m(x)$ in (3) satisfies (conservatively)

$$E(\widehat{f}_m(x) - f(x)) = O(2^{-m(r-1)}) \tag{4}$$

and the mean square error is

$$MSE = Var + Bias^2 = O(N^{-1+1/(2r-1)}). \tag{5}$$

However, the following difficulties arise:

a. $\widehat{f}_m(x)$ is not a density function since it may take negative values;

b. There are continuous densities for which $E(\widehat{f}_m(x) - f(x)) \nrightarrow 0$;

c. Gibbs' phenomenon (excessive oscillations at discontinuities) occurs for the empiric distribution function.

## 2. **Wavelet estimators.**

Let $\varphi(t)$ be an orthogonal scaling function, $\psi(t)$ its associated mother wavelet, $\{V_m\}$ the associated multiresolution ladder of subspaces, $q(x,y)$ the reproducing kernel of $V_0$, and

$$q_m(x,y) = 2^m q(2^m x, 2^m y) \tag{6}$$

be the reproducing kernel of $V_m$. The sequence $\{q_m(x,y)\}$ is now a quasi-positive delta sequence. Such sequences lead to uniform convergence, but may also exhibit Gibbs phenomenon [W, p.117]. These are used directly to construct the linear wavelet estimators. They are exactly as given before

$$\widehat{f}_m(x) = \int_{-\infty}^{\infty} q_m(x,y) f^*(y) dy = \frac{1}{N} \sum_{i=1}^{N} q_m(x, X_i) \tag{7}$$

with MSE as in other orthogonal series estimators$(= O(N^{-1+1/(2r-1)}))$ for sufficiently regular wavelets. But again there are difficulties:

1. $\widehat{f}_m(x)$ is itself not a density function except for the Haar wavelet case (which is exactly the histogram!), since there are no non-negative continuous wavelets [J];

2.Gibbs' phenomenon still plagues us [S-V].

On the other hand, the lack of convergence for some continuous densities does not arise since for quasipositive delta sequence, the bias converges to zero [W, p.118].

These linear estimators are essentially low pass filters. All of the fine scale coefficients are ignored. This is undesirable if the density to be estimated has irregular features which would be reflected in larger coefficients at finer scales. In order to include these coefficients, threshold estimators have been proposed [D-J]. These have been used with good results for noise elimination in approximation and regression estimation. However they must used with care in density estimation so as not to confound irregularities in the density with the irregularities in $f^*$.

These non-linear estimators have the form

$$\widehat{f}(x) = \widehat{f}_m(x) + \sum_{k=m}^{\infty} \sum_{n=-\infty}^{\infty} b_{k,n}\psi_{k,n}(x)$$

where $\widehat{f}_m(x)$ is the linear estimator at a very smooth scale, and $b_{k,n}$ is the threshold version of the wavelet coefficient of the empiric distribution

$$b^*_{k,n} = \int_{-\infty}^{\infty} f^*(y)\psi_{k,n}(y)dy$$

where $b_{k,n} = 0$ if $|b^*_{k,n}| < \delta$ or $k \geq m_1$. Other values may also be shrunken to get soft thresholding. These estimators work for both smooth and irregular densities since in the former case, the coefficients at the finer scales are all small. But there are still some undesirable properties.

a. $\widehat{f}(x)$ is still not a density;

c. We may still have Gibbs' phenomenon.

2.1. **Non-negative wavelet estimators.** In order to ensure that our estimator is itself a density, we need to restrict ourselves to non-negative estimators. This can be achieved by truncating the estimators to non-negative values and dividing the result by the integral. This, however, compromises the smoothness of the estimator and still may lead to spurious bumps.

One idea is to estimate the square root of the density with a wavelet estimator and then use its square [P-V]. This method works and gives *bonafide* density function estimators, but the construction needs to replace $f^*(x)$ with a preliminary estimator since the empiric distribution does not have a square root. Threshold methods can then be used with this estimator. Unfortunately, the procedure is quite complicated.

Another idea involves using Meyer wavelets whose Fourier transform is non-negative, and then use the linear wavelet estimators to approximate the characteristic function [W, p.204] instead of the density. These estimators will be non-negative but the convergence is slow and again the procedure is quite complicated.

2.2. **Non-negative scaling functions.** There are no continuous non-negative orthogonal scaling functions [J]. In fact, it's not even clear that there is, in general, an $f \in V_0$ which is $\geq 0$. The following construction shows there is.

2.2.1. ***A general construction.*** Let $\theta(t)$ be any continuous function on $\mathbb{R}$ with support in an interval $[M, N]$. We say that $\theta(t)$ generates a partition of unity if

$$\sum_{n=-\infty}^{\infty} \theta(t-n) \equiv 1, \quad t \in \mathbb{R}. \tag{8}$$

This is a property shared by all orthogonal scaling functions of compact support, in particular those due to Daubechies as well as the "Coiflets" of

Coifman and other "–lets". We define the Abel means of their series (1) to be

$$\rho_r(t) := \sum_{n=-\infty}^{\infty} r^{|n|}\theta(t-n), \quad 0 < r \leq 1, \quad t \in \mathbb{R}. \tag{9}$$

This series converges uniformly on [0,1] since it is locally finite. Furthermore $\rho_r(t) \longrightarrow 1$ for $t \in [0,1]$ as $r \longrightarrow 1$ by the regular summability property of Abel means. (If a series converges to $S$, it is also Abel summable to $S$).

The series also converges uniformly on any finite interval $[M,N]$ and therefore there is an $0 < r_0 < 1$ such that $\rho_r(t) \geq 1/2$ for $1 > r \geq r_0$, $t \in [M,N]$.

We denote by $V$ the closed linear span of $\{\theta(t-n)\}_{n\in Z}$ in $L^2(\mathbb{R})$. Then if in addition $\{\theta(t-n)\}$ is a Riesz basis of V and $a_n \in l^2$, we have

$$f(t) = \sum_{n=-\infty}^{\infty} a_n\theta(t-n) \in V.$$

This requirement will be met if $\widehat{\theta}(\omega)$ satisfies

$$0 < A \leq \sum_{k=-\infty}^{\infty} |\widehat{\theta}(\omega + 2k\pi)|^2 \leq B < \infty. \tag{10}$$

The properties of $\rho_r(t)$ are given by the next lemma.

**Lemma 2.1.** *Let $\theta(t)$ be a continuous function on $\mathbb{R}$ with compact support satisfying (1) and (3); let $V = CLS\{\theta(t-n), n \in Z\}$; then there is an $0 < r_0 < 1$, such that $\rho_r$ given by (2) for $r_0 \leq r < 1$, satisfies*

(i) $\rho_r(t) \geq 0$, $t \in \mathbb{R}$,
(ii) $\rho_r \in V$.

2.2.2. ***Back to wavelets.*** >From Lemma 1, we have an answer to the question posed at the start of this section.

**Corollary 2.2.** *Let $\{V_m\}$ be a multiresolution of $L^2(\mathbb{R})$ associated with the scaling function $\phi(t)$with compact support. Then there is an element $\rho \in V_m$ such that $\rho(t) \geq 0$.*

We can say much more than this. In fact, we have

**Theorem 2.3.** *Let $\rho_r(t) = \sum_n r^{|n|}\phi(t-n)$, where $r$ is chosen so large that $\rho_r(t) \geq 0$, for $t \in \mathbb{R}$; then $\{2^{\frac{m}{2}}\rho_r(2^m t - n)\}_{n\in Z}$ is a Riesz basis of $V_m$; its dual basis is generated by $\rho_r$, where*

$$\rho_r^*(t) = \frac{1}{2\pi(1-r^2)}[(1+r^2)\phi(t) - r\{\phi(t+1) + \phi(t-1)\}].$$

However, this biorthogonal system does not give us the positive kernel we need. Rather we use a modification which gives the desired properties. The kernel that gives us the approximation in $V_0$ to $f \in L^2$ is given by

$$k_r(t,s) = \left(\frac{1-r}{1+r}\right)^2 \sum_{n=-\infty}^{\infty} \rho_r(t-n)\rho_r(s-n), \tag{11}$$

i.e.

$$f_0^r(t) = \int_{-\infty}^{\infty} k_r(s,t) f(s) ds.$$

This kernel satisfies the conditions needed to generate a positive delta sequence $\{k_{r,m}\}$ where

$$k_{r,m}(s,t) = 2^m k_r(2^m s, 2^m t), \quad m \in \mathbb{Z}. \tag{12}$$

We summarize the properties of $k_{r,m}$ in the next proposition

**Proposition 2.4.** *Let $k_{r,m}$ be as in (4.7), let $f \in L^1 \cap L^2(\mathbb{R})$; let*

$$f_m^r(t) = \int_{-\infty}^{\infty} k_{r,m}(t,s) f(s) ds;$$

*then $f_m^r \in V_m$ and*

(i) if $M_1 \le f(t) \le M_2$ for $t \in \mathbb{R}$, then $M_1 \le f_m^r(t) \le M_2$ for $t \in \mathbb{R}, m \in \mathbb{Z}$,
(ii) if $M_3 \le f(t) \le M_4$ for $t \in [a,b]$, then for each $\epsilon > 0, \delta > 0$, there is an $m_0$, such that for $t \in (a+\delta, b-\delta), M_3 - \epsilon \le f_m^r(t) \le M_4 + \epsilon$, for $m \ge m_0$.

Clearly, (ii) guarantees that $f_m^r(t)$ will not exhibit Gibbs' phenomenon and oscillations will be controlled by the given $\epsilon$. That is, the overshoot at discontinuities can be no more than $\epsilon$, which is arbitrarily small.

The linear estimator associated with this kernel is given by the expression

$$\widehat{f_m^r}(x) = \int_{-\infty}^{\infty} k_{r,m}(x,y) f^*(y) dy = \frac{1}{N} \sum_{i=1}^{N} k_{r,m}(x, X_i) \tag{13}$$

as one might expect.

If our density $f \in H^r$, the Sobolev space, the usual linear wavelet estimator has bias satisfying $O(2^{(1-r)m})$ if the wavelet is sufficiently regular. This can be used to estimate the bias for our new estimator (surprisingly).

**Lemma 2.5.** *Let the scaling function $\varphi \in C^2$, $f_m(x) = E\widehat{f_m}(x)$, the usual linear estimator, let the density $f \in H^r$, for $r \ge 3$; then*

$$f_m^r(x) - f_m(x) = O(2^{-2m}).$$

The proof is based on the fact that

$$\mathcal{F} f_m^r(\omega) = \frac{(1-r)^4}{(1 - 2r\cos 2^{-m}\omega + r^2)^2} \mathcal{F} f_m(\omega)$$

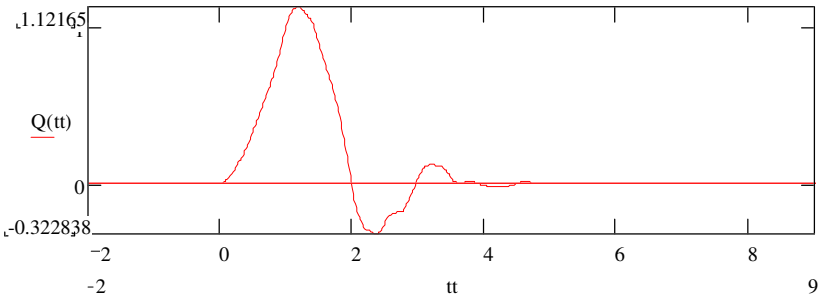


FIGURE 1. The scaling function for the Daubechies wavelet, $N = 4$.

where $\mathcal{F}$ denotes the Fourier transform. The difference between these two is $\mathrm{O}(2^{-2m}\omega^2)$.

**Theorem 2.6.** *Let $\varphi$ be as in the lemma, let $\widehat{f}_m^r(x)$ be the estimator above; let the density function $f \in H^3$, then the mean square error satisfies*

$$MSE = O(N^{-4/5})$$

*for a proper choice of $m$, uniformly on bounded sets.*

This rate is somewhat better than other non-negative estimators.

## 3. **Numerical Experiments.**

In this section, we illustrate how the smooth positive wavelet estimators are used to estimate the density functions of univariate data. The results are compared to the case of the linear projection wavelet estimators.

As in the previous section, we define two estimators based on (6) and (12), respectively, that is,

Type a. $\widehat{f}_m(x)$: the wavelet estimator directly constructed from the reproducing kernel of Daubechies wavelets as in (7), and

Type b. $\widehat{f}_m^r(x)$: the positive wavelet estimator constructed from the positive kernel as in (13), introduced in previous section.

The graphs of the Daubechies scaling functions ($N = 4, 8$) and their associated non-negative basis functions are shown in Figure 1 to Figure 4.

We consider two examples:

**Example 3.1.** *Density estimate of the Old Faithful geyser data. The data for the eruption lengths (in minutes) of* 107 *eruptions of Old Faithful geyser can be found in Table 1.*

**Example 3.2.** *Density estimate for suicide study data. The data for the lengths of treatment spells (in days) of* 86 *control patients in suicide study can be found in Table 2.*

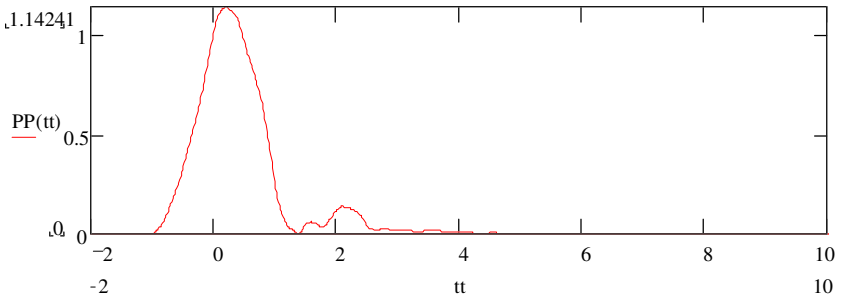


FIGURE 2. The positive basis function $\rho_r(t)$, associated with the Daubechies scaling function ${}_4\varphi, r = 0.30$.

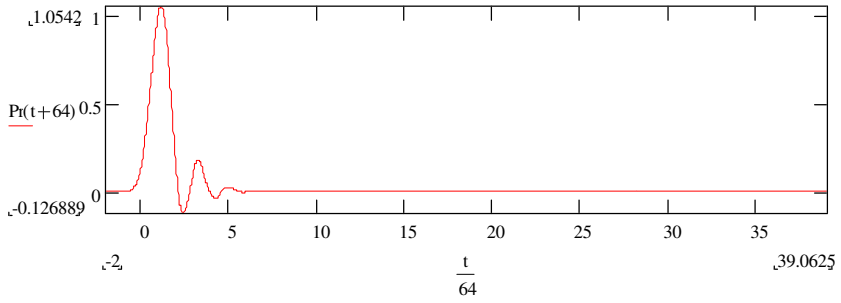


FIGURE 3. The scaling function for the Daubechies wavelet, $N = 8$.

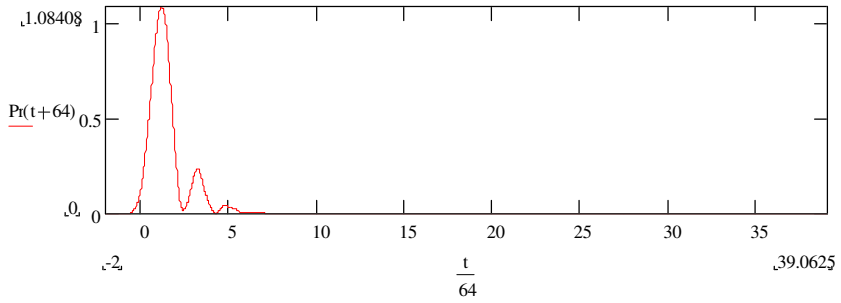


FIGURE 4. The positive basis function $\rho_r(t)$, associated with Daubechies scaling function ${}_8\varphi, r = 0.43$.

Both of the data sets can be found in [S, p. 8].

We illustrate the experiment results with the graphs of the estimators. The results show that the negative values occur for the estimators of type a in some places, while the values for the estimators of type b are always nonnegative. The latter are also smoother than the former.

The results for the first example are shown in Figure 5 through Figure 8. The results related to the second example are illustrated by Figure9 through 12 The experiments are performed with the same window widths and several different kernels (for the different values of the parameter $N$ in Daubechies wavelets) to illustrate the different convergence rates and the smoothness. The results show that, as expected, the estimators are smoother for $N$ larger. One reason for this is that the smoothness of the scaling function changed with the parameter $N$; another is that the convergence rate in series (11) is faster for larger $N$(See [W-S2].

**Table 1. Eruption Lengths (in minutes)**
**Of 107 Eruptions Of Old Faithful Geyser.**

| | | | | | |
|---|---|---|---|---|---|
| 4.37 | 3.92 | 2.33 | 3.33 | 4.18 | 4.50 |
| 4.70 | 3.20 | 4.57 | 3.73 | 4.58 | 1.80 |
| 1.68 | 4.58 | 3.58 | 1.67 | 3.50 | 3.70 |
| 1.75 | 3.50 | 3.70 | 4.63 | 4.62 | 2.50 |
| 4.35 | 3.80 | 4.25 | 1.83 | 4.03 | 2.27 |
| 1.77 | 3.80 | 3.58 | 2.03 | 1.97 | 2.93 |
| 4.25 | 1.80 | 3.67 | 2.72 | 4.60 | 4.63 |
| 4.10 | 1.95 | 1.90 | 4.03 | 4.00 | 4.00 |
| 4.05 | 1.77 | 4.13 | 1.73 | 3.75 | 1.97 |
| 1.90 | 4.28 | 4.53 | 3.10 | 4.00 | 3.93 |
| 4.00 | 4.40 | 4.10 | 4.62 | 4.33 | 4.07 |
| 4.42 | 4.65 | 4.12 | 1.88 | 1.82 | 4.50 |
| 1.83 | 4.50 | 4.00 | 3.52 | 1.67 | 2.25 |
| 1.83 | 3.43 | 4.93 | 3.77 | 3.50 | 4.25 |
| 3.95 | 2.93 | 3.68 | 3.43 | 4.20 | 4.08 |
| 4.83 | 4.33 | 1.85 | 2.00 | 4.43 | 3.92 |
| 3.87 | 4.13 | 3.83 | 3.73 | 1.90 | 4.73 |
| 1.73 | 3.72 | 1.85 | 4.60 | 4.08 | |

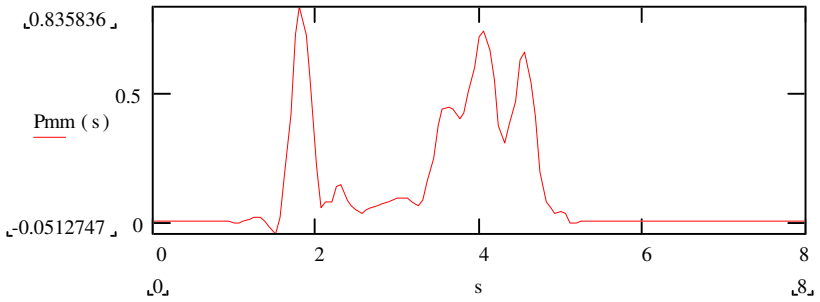


FIGURE 5. Density estimate for the Old Faithful geyser data using the reproducing kernel for ${}_4\varphi$ (Figure 1), window width $0.25(m = 2)$

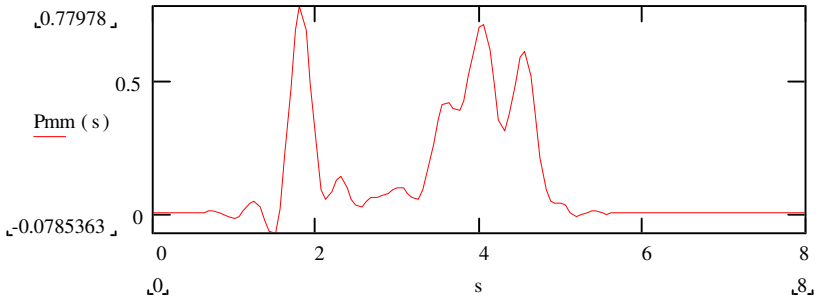


FIGURE 6. Density estimate for the Old Faithful geyser data using the reproducing kernel for ${}_8\varphi$ (Figure 5), window width $0.25(m = 2)$

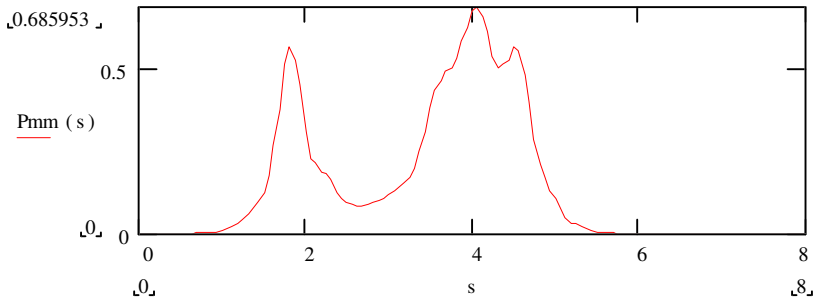


FIGURE 7. Density estimate for the Old Faithful geyser data by using the associated positive kernel, $N = 4$, window width $0.25(m = 2)$.

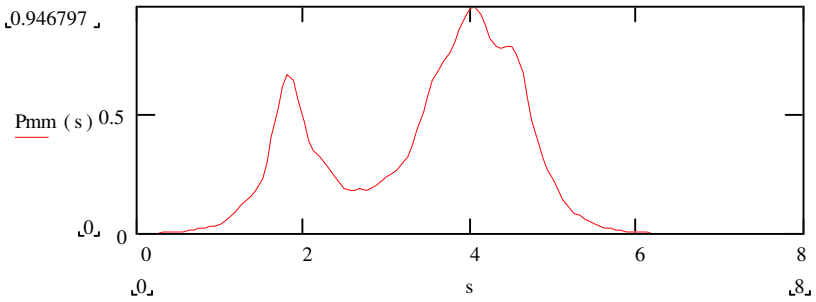


FIGURE 8. Density estimate for the Old Faithful geyser data by using the positive kernel , $N = 8$, window width $0.25(m = 2)$.

**Table 2. Lengths Of Treatment Spells (in days) Of 86 Control Patients In Suicide Study.**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 18 | 32 | 54 | 82 | 103 | 153 | 311 |
| 1 | 21 | 34 | 56 | 83 | 111 | 163 | 314 |
| 1 | 21 | 35 | 56 | 84 | 112 | 167 | 322 |
| 5 | 22 | 36 | 62 | 84 | 119 | 175 | 369 |
| 7 | 25 | 37 | 63 | 84 | 122 | 228 | 415 |
| 8 | 27 | 38 | 65 | 90 | 123 | 231 | 573 |
| 8 | 27 | 39 | 65 | 91 | 126 | 235 | 609 |
| 13 | 30 | 39 | 67 | 92 | 129 | 242 | 740 |
| 14 | 30 | 40 | 75 | 93 | 134 | 256 | 737 |
| 14 | 31 | 49 | 76 | 93 | 144 | 256 | |
| 17 | 31 | 49 | 79 | 103 | 147 | 257 | |

Comparing Figure 5 and Figure 6 with Figure 7 and Figure 8, we notice that the nonnegative wavelet estimators, in addition to being densities themselves, are smoother (but may converge slower) than the projection wavelet estimators.

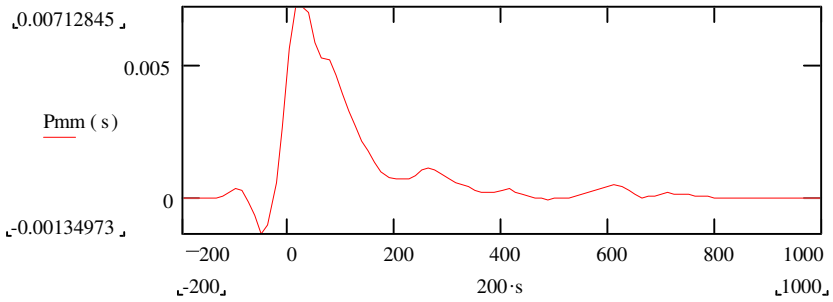


FIGURE 9. Density estimate for suicide study data by using the reproducing kernel for scaling function in Figure 1, $N = 4$, window width $0.25(m = 2)$.

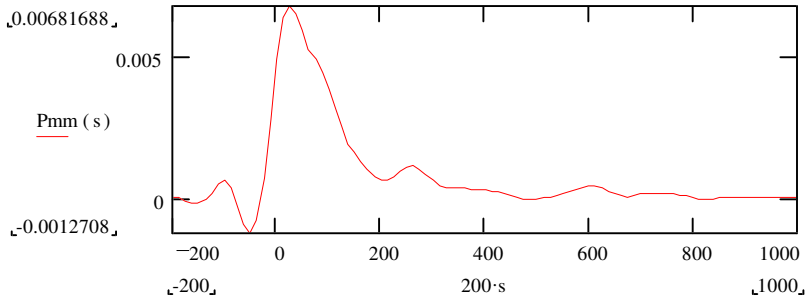


FIGURE 10. Density estimate for suicide study data by using the reproducing kernel for $N = 8$, window width $0.25(m = 2)$.

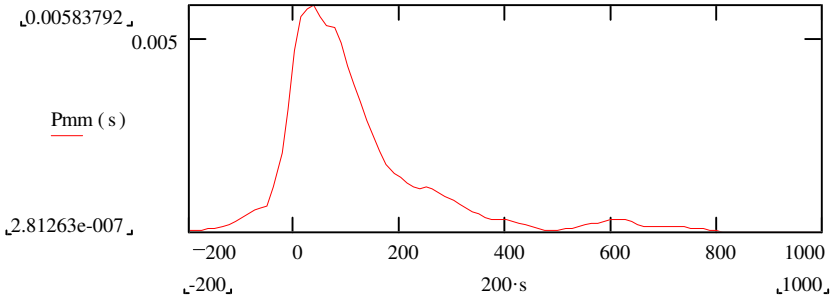


FIGURE 11. Density estimate for suicide study data by using the associated positive kernel for $N = 4$, window width $0.25(m = 2)$.

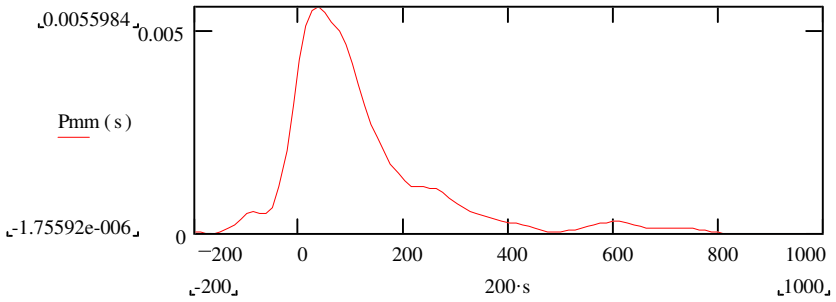


FIGURE 12. Density estimate for suicide study data by using the positive kernel for $N = 8$, window width $0.25(m = 2)$.

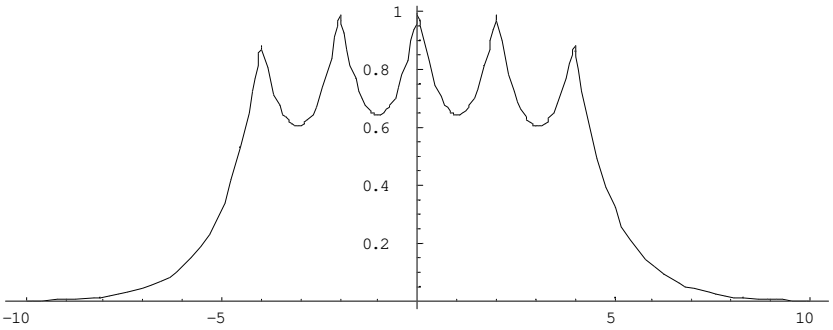


FIGURE 13. The graph of the function $f(x) = e^{-|x-4|} + e^{-|x-2|} + e^{-|x+4|}e^{-|x+2|} + e^{-|x|}$.

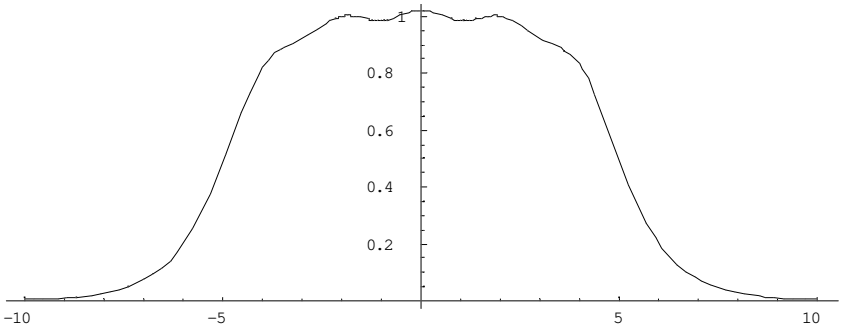


FIGURE 14. The noisy version of Figure 13 (Convolution of function $f$ in Figure 13 and the density function of a normal distribution. Both functions are normalized by their maximum values).

## 4. **Noise.**

We suppose that our sample $X_1, X_2, ..., X_N$ has been corrupted by noise. That is, we observe $Y_1, Y_2, ..., Y_N$, where

$$Y_i = X_i + Z_i$$

and $\{Z_i\}$ is IID noise with mean 0 and independent of the $X_i'$s.

We suppose that $Z_i$ has density $k(z)$. Then the density of $Y_i$ is given by

$$g(y) = (k * f)(y) = \int_{-\infty}^{\infty} k(y-x)f(x)dx.$$

One procedure is first to estimate g and then solve the convolution equation; but this is an ill-posed problem!! (Can be converted to well-posed with wavelets, see [W-S3]). In Figure 17 and Figure 18, we illustrate how the addition of noise makes the density function smoother. The smoothness will depend on the particular distribution of the noise, with the case of the normal distribution giving the worst behavior. (See for example, [W-S3].)

Another procedure is to treat the problem as a problem in empirical Bayes (EB) estimation. EB problems of this type have been attacked by using orthogonal polynomials when the kernel $k(z)$ is from a natural exponential family of a certain kind (with quadratic variance function)[W-H]. But estimators of $f(x)$ based on these polynomials are not necessarily themselves densities.

In order to use wavelets to attack the EB problem, we begin with a scaling function $\varphi$ in the domain of the convolution operator

$$(Kf)(y) = (k * f)(y)$$

and the multiresolution $\{V_m\}$, We denote by $U_m = KV_m$, and suppose the Fourier transform of $k$ has support at least in the interval $[-2^m\pi, 2^m\pi]$. We denote by $\varphi_m(t) = 2^{m/2}\varphi(2^m t)$; and by $\theta_m = K\varphi_m$. Then $\{\theta_m(t - n2^{-m}\}$ is a Riesz basis of $U_m$ [W-S3].

The kernel used to construct our estimator is

$$h_m(t,s) = \sum_{n=-\infty}^{\infty} \theta_m(t - n2^{-m})\theta_m(s - n2^{-m}).$$

If the Fourier transform of $k$ never vanishes, then $\{h_m(t,s)\}$ is a quasi-positive delta sequence. The usual estimators can be used to get an estimate of $g(y) = (k*f)(y)$, but what we want is an estimate of $f(x)$. This is obtained by the estimator

$$\widehat{f}_m(x) = \frac{1}{N}\sum_{i=1}^{N} q_m(x, Y_i)$$

where

$$q_m(x,y) = \sum_{n=-\infty}^{\infty} \varphi_m(x - n2^{-m})\theta_m(y - n2^{-m}).$$

However, the convergence is very slow. Another approach which gives faster convergence is found in [P].

By modifying this kernel as in the previous section, we can obtain a non-negative estimator as well. One needs only to replace the $\varphi(t)$'s in the above formulae by the $\rho_r(t)$. Also threshold estimators can be adapted to this procedure as well by adding terms involving the image of the wavelets $K\psi_{m,n}$.

Note that the $\theta_m$ and hence the kernels will also be non-negative in such a case. A complete study of this remains to be done.

## References


[D] I. Daubechies, Ten Lectures On Wavelets, CBMS-NSF Series in Appl. Math., SIAM Publ., 1992.

[J] A.J.E.M. Janssen, The Smith - Barnwell Condition And Non - Negative Scaling Functions, IEEE Trans. Info. The., 38, 884-885 (1994).

[K-R-K] S. E. Kelly, M. A. Kon and L. A. Raphael, Local Convergence For Wavelet Expansions, J. Funct. Anal 126(1994), 102-138.

[S] B. W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, 1986.

[W] G. G. Walter, Wavelets And Other Orthogonal Wavelets With Applications, CRC Press, 1994.

[W-S] G. G. Walter and X. Shen, Positive Estimation With Wavelets, preprint, 1997.

[W-S2] G. G. Walter and X. Shen, Summability For Wavelet, preprint, 1997.

[W-S3] G. G. Walter and X. Shen, Deconvolution Using Meyer Wavelets, preprint, 1997.

[W-Sh] G. G. Walter and H. T. Shim, Gibbs' Phenomenon For Sampling Series And What To Do About It. Preprint, 1996.

[W-B] G. G. Walter and J. Blum, Probability Density Estimation using Delta sequences,Ann. Statist. 7(1979),328-340.

[P] M. Pensky, Wavelet Empirical Bayes Estimation of a Location or a Scale Parameter, preprint, 1997.

[D-J-K-P] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Density Estimation by Wavelet Thresholding, Ann Statist. 24(1996), 508-539.

[S-V] H-T. Shim and H. Volkmer, On Gibbs' Phenomenon for Wavelet Expansions, J. Approx. Th. 84(1996(, 74-95.

[W-H] G. G. Walter and G. G. Hamedani, Bayes Emperical Bayes Estimation for Natural Exponential Families with Quadratic Variance Functions, Ann. Statist. 19(1991), 1191-1224.

[P-V] A. Pinheiro and B. Vidakovic, Estimating the Square Root of a Density via Compactly Supported Wavelets, Technical Report, ISDS, Duke Univ. (1995).

*Current address*: Mathematics Department, University of Wisconsin-Milwaukee, Milwaukee, WI 53201.

*E-mail address*: ggw@csd.uwm.edu

*Current address*: Mathematics Department, University of Wisconsin-Oshkosh, Oshkosh, WI 54901.

*E-mail address*: shen@vaxa.cis.uwosh.edu