ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Appl. Comput. Harmon. Anal. 25 (2008) 266–275

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

Convergence of wavelet thresholding estimators of differential operators ☆

Di-Rong Chen *, Hongtao Meng

Department of Mathematics, and LMIB, Beijing University of Aeronautics and Astronautics, Beijing 100083, PR China

Available online 4 March 2008

Communicated by Gregory Beylkin on 17 December 2006

Abstract

Wavelet shrinkage is a strategy to obtain a nonlinear approximation to a given signal. The shrinkage method is applied in different areas, including data compression, signal processing and statistics. The almost everywhere convergence of resulting wavelet series has been established in [T. Tao, On the almost everywhere convergence of wavelet summation methods, Appl. Comput. Harmon. Anal. 3 (1996) 384–387] and [T. Tao, B. Vidakovic, Almost everywhere behavior of general wavelet shrinkage operators, Appl. Comput. Harmon. Anal. 9 (2000) 72–82]. With a representation of f' in terms of wavelet coefficients of f, we are interested in considering the influence of wavelet thresholding to f on its derivative f'. In this paper, for the representation of differential operators in nonstandard form, we establish the almost everywhere convergence of estimators as threshold tends to zero.

Keywords: Wavelets; Thresholding; Nonstandard form; Hardy–Littlewood maximal function

1. Introduction

The main purpose of this work is to consider the asymptotic behavior of estimators of differential operators via wavelet thresholding when the thresholds tend to zero.

The wavelet representation of a function "automatically" places significant coefficients in a neighborhood of large gradients present in the function due to the vanishing moments of wavelets. Based on this, wavelet shrinkage is a strategy to obtain a nonlinear approximation to a given signal. The shrinkage method is applied in different areas, including data compression, signal processing and statistics. When the soft or hard thresholding is applied, the resulting wavelet shrinkage estimators possess asymptotic near-minimax optimality properties [1,8,10,11]. The almost everywhere convergence and norm convergence of resulting wavelet series have been established in [16] and [17].

A question arises naturally: how does wavelet thresholding of a function f affect its derivative f'? To answer this question, we first need to represent appropriately f' by making use of the wavelet expansion of f. In this paper, the so called nonstandard form (NSF) of representation of certain operators, due to Beylkin, Coifman and Rokhlin [5], plays an important role. The NSF may be constructed by Beylkin–Coifman–Rokhlin (BCR) algorithm. As well known, for a wide classes of operators, the NSF leads to fast algorithms for matrix multiplications. We note that the NSF of many

☆ Research supported in part by NSF of China under grant 10571010.
* Corresponding author.
E-mail address: drchen@buaa.edu.cn (D.-R. Chen).

1063-5203/$ – see front matter
doi:10.1016/j.acha.2008.02.001

basic operators, among which are the differential operators, have been computed exact and explicitly [3–5]. Besides NSF, two methods for reconstruction of some operators with wavelet approach have been developed. One is based on the wavelet–vaguelette decomposition (WVD) [9] and the other is based on the vaguelette–wavelet decomposition (VWD) [2]. There are some differences among them as explained below. With NSF one may represent, using the wavelet coefficients of f, the operator with respect to the underlying wavelet basis and scale functions. However, the WVD expands f' with the underlying wavelet basis, but not the wavelet coefficients of f. On the other hand, while the VWD makes use of the wavelet coefficients of f, it expands f' with an appropriate basis other than the underlying wavelet basis. The wavelet thresholding estimators based on WVD and VWD are constructed in [9] and [2], respectively.

In this paper, it is demonstrated that the approach of wavelet thresholding can be used for the NSF as well. More precisely, we establish that, as the threshold tends to zero, the resulting NSF of f' converges to f' almost everywhere.

Before proceeding further with the main results, we introduce the notions concerning wavelet thresholding and NSF of differential operator.

1.1. Wavelet thresholding estimator

Suppose that φ is an orthogonal scaling function, i.e., it satisfies the refinement equation

$$\varphi(x)=\sum_{k=0}^{L-1}h_k\varphi(2x-k),\quad x\in\mathbb{R}, \tag{1.1}$$

and $\{\varphi(\cdot-k)\}_k$ constitutes an orthonormal set in $L_2(\mathbb{R})$. It is known that φ is supported on $[0,L-1]$. Then an orthogonal wavelet ψ is constructed by

$$\psi(x)=\sum_{k=0}^{L-1}g_k\varphi(2x-k),\quad g_k=(-1)^k h_{L-k-1}. \tag{1.2}$$

By the term *orthogonal wavelet* we mean that the set $\{\psi_{jk}=2^{j/2}\psi(2^j\cdot-k)\colon\ j,k\in\mathbb{Z}\}$ of functions is an orthonormal basis for $L_2(\mathbb{R})$.

Henceforth we assume that ψ and φ are given as above. The wavelet representation of a function f is given by

$$f=\sum_{jk}d_{jk}\psi_{jk},\quad d_{jk}=\langle f,\psi_{jk}\rangle.$$

For $f\in L^p(\mathbb{R})$, $1\leqslant p<\infty$, the above equality holds almost everywhere and in $L^p(\mathbb{R})$-norm [13,16,17].

Definition 1.1. A function $\delta(x,\lambda):\mathbb{R}\times\mathbb{R}^+\mapsto\mathbb{R}$ is called a thresholding rule, with the threshold λ, if there exist nonnegative constants a and b such that

$$\left|x-\delta(x,\lambda)\right|\leqslant a\lambda \tag{1.3a}$$

and

$$\left|\delta(x,\lambda)\right|\leqslant b|x|\boldsymbol{\chi}_{\{|x|>\lambda\}} \tag{1.3b}$$

for all $x\in\mathbb{R}$ and $\lambda\in\mathbb{R}^+$.

Examples of thresholding rules include hard thresholding, semisoft shrinkage and hyperbole rule, etc. With a thresholding rule, the wavelet thresholding estimator of f is defined by

$$T_\lambda f(x)=\sum_{j,k}\delta(d_{jk},\lambda)\psi_{jk}(x),\quad \lambda>0. \tag{1.4}$$

The series in (1.4) converges absolutely a.e. for any $f\in L^p(\mathbb{R})$, $1\leqslant p<\infty$, and converges in $L^p(\mathbb{R})$-norm for any $f\in L^p(\mathbb{R})$, $1<p<\infty$. Further, $T_\lambda f\to f$ a.e. and in $L^p(\mathbb{R})$-norm as $\lambda\to 0$. These results have been established in [16,17].

1.2. NSF of differential operator

For the construction of NSF of differential operator d/dx, we suppose that $\varphi' \in L^p(\mathbb{R})$ and introduce the constants

$$r_l = \int_{-\infty}^{+\infty} \varphi(x-l)\varphi'(x)\,dx, \quad l \in \mathbb{Z}. \tag{1.5}$$

Clearly, $r_l = 0$ for $|l| > L-2$. We note that $r_l, l \in \mathbb{Z}$, may be computed by solving a system of linear equations, which has a unique solution with a finite number of nonzero r_l [3–5]. Moreover, let

$$\alpha_l = \int_{-\infty}^{+\infty} \psi(x-l)\psi'(x)\,dx, \qquad \beta_l = \int_{-\infty}^{+\infty} \psi(x-l)\varphi'(x)\,dx, \qquad \gamma_l = \int_{-\infty}^{+\infty} \varphi(x-l)\psi'(x)\,dx.$$

It follows from (1.1) and (1.2) that the constants $\{\alpha_l\}$, $\{\beta_l\}$ and $\{\gamma_l\}$ can be computed from $\{r_l\}$. For example $\alpha_i = \sum_{k=0}^{L-1}\sum_{k'=0}^{L-1} g_k g_{k'} r_{2i+k-k'}$.

It is well known that the assumption $\varphi' \in L^p$ implies the polynomial reproducibility of φ (see [7, p. 245])

$$\sum_l \varphi(x-l) = 1, \qquad \sum_l l\varphi(x-l) = x - c, \tag{1.6}$$

where c is a constant. By the construction we have the discrete vanishing moments

$$\sum_{|l|\leqslant L-2} r_l = \sum_{|l|\leqslant L-2} l^i \alpha_l = \sum_{|l|\leqslant L-2} l^i \beta_l = \sum_{|l|\leqslant L-2} l^i \gamma_l = 0, \quad i = 0, 1. \tag{1.7}$$

Given a function $f \in L^p(\mathbb{R})$ $(1 \leqslant p \leqslant \infty)$ with $f' \in L^q(\mathbb{R})$ $(1 \leqslant q < \infty)$, we define functions $\mathcal{S}_J f$, $J \in \mathbb{Z}$, as following

$$\mathcal{S}_J f = \sum_{j<J}\sum_k \left(\psi_{jk} 2^j \sum_l \alpha_l d_{j,k-l} + \psi_{jk} 2^j \sum_l \beta_l s_{j,k-l} + \varphi_{jk} 2^j \sum_l \gamma_l d_{j,k-l}\right), \tag{1.8}$$

where $\varphi_{jk}(x) = 2^{j/2}\varphi(2^j x - k)$, $s_{jk} = \langle f, \varphi_{jk}\rangle$. Under the condition of f, the convergence of series (1.8) both in pointwise and $L^q(\mathbb{R})$-norm will be established in Theorem 1.2.

Moreover, we prove in Theorem 1.2 that, for f with $f' \in L^q(\mathbb{R})$, $1 \leqslant q < \infty$, the sequence $\{S_J f(x)\}_J$ converges to $f'(x)$ pointwise and in $L^q(\mathbb{R})$-norm as $J \to \infty$. With this fact, the NSF of operator d/dx is the following representation

$$\mathcal{T} f = \sum_j \sum_k \left(\psi_{jk} 2^j \sum_l \alpha_l d_{j,k-l} + \psi_{jk} 2^j \sum_l \beta_l s_{j,k-l} + \varphi_{jk} 2^j \sum_l \gamma_l d_{j,k-l}\right). \tag{1.9}$$

We note that the NSF of the differential operator here is essentially the same as that given in [3–5], although it seems that they are represented in different forms.

In practice, for dealing with n data, VWD algorithm requires $O(n\log^2 n)$ operations. However, a fast algorithm for NSF needs only $O(n)$ operations provided that the constants r_l, α_l, β_l and γ_l are given.

We are interested in studying the estimation of f' by thresholding with the NSF. Recall that $T_\lambda f$ is given in (1.4). The estimator $\mathcal{T}_\lambda$ of differential operator via thresholding is given, at least formally, by $\mathcal{T}_\lambda f = \mathcal{T}(T_\lambda f)$. This can be represented formally as

$$\mathcal{T}_\lambda f = \sum_j \sum_k \left(\psi_{jk} 2^j \sum_l \alpha_l \delta(d_{j,k-l}, \lambda) + \psi_{jk} 2^j \sum_l \beta_l \hat{s}_{j,k-l} + \varphi_{jk} 2^j \sum_l \gamma_l \delta(d_{j,k-l}, \lambda)\right), \tag{1.10}$$

where $\hat{s}_{jk} = \langle T_\lambda f, \varphi_{jk}\rangle$. We will demonstrate that $\mathcal{T}_\lambda f$ is well defined. In fact, the convergence of (1.10) will be established in Theorem 1.3.

1.3. Main results

We state the main results in this subsection. The first one is about the convergence concerned with NSF of differential operator. Its proof is given in Section 2.

Theorem 1.2. *Assume that the orthogonal scaling function $\varphi \in C_0^1(\mathbb{R})$, the set of compactly supported functions with continuous derivative on $\mathbb{R}$. If $f \in L^p(\mathbb{R})$, $1 \leqslant p \leqslant \infty$, $f' \in L^q(\mathbb{R})$, $1 \leqslant q < \infty$, then the series* (1.8) *converges absolutely and uniformly on $\mathbb{R}$. Moreover, $\lim_{J\to+\infty} \mathcal{S}_J f(x) = f'(x)$ holds at any Lebesgue point x of f' and holds in $L^q(\mathbb{R})$-norm. In other words, $\mathcal{T} f(x) = f'(x)$ holds at any Lebesgue point x of f' and holds in $L^q(\mathbb{R})$-norm.*

The purpose of this paper is to establish the almost convergence of the estimator $\mathcal{T}_\lambda f$ in the following result. Its proof is given in Section 3.

Theorem 1.3. *Assume that the orthogonal scaling function $\varphi \in C_0^1(\mathbb{R})$. If $f \in L^p(\mathbb{R})$, $1 \leqslant p \leqslant \infty$, $f' \in L^q(\mathbb{R})$, $1 < q < \infty$, then for any $\lambda > 0$, $x \in \mathbb{R}$, the series* (1.10) *converges. Moreover, for any Lebesgue point x of f' we have*

$$\lim_{\lambda\to 0} \mathcal{T}_\lambda f(x) = f'(x). \tag{1.11}$$

Remark 1.4. With similar arguments, we can establish the same results for operators d^n/dx^n, where $n \in N$, and fractional derivatives.

2. Convergence of NSF

This section presents a proof of Theorem 1.2.

We first introduce the Hardy–Littlewood Maximal function,

$$Mf(x) = \sup_{\tau>0} \frac{1}{2\tau} \int\limits_{|y-x|<\tau} \left|f(y)\right| dy.$$

The maximal function has an important property that $\|Mf\|_{L^p(\mathbb{R})} \leqslant C\|f\|_{L^p(\mathbb{R})}$ for $1 < p \leqslant \infty$. In this paper, the constants c, C and C' change from line to line.

If x is a Lebesgue point of f, then

$$\left|f(x)\right| \leqslant Mf(x). \tag{2.1}$$

Recall that a point $x \in \mathbb{R}$ is a Lebesgue point of a locally integrable function $f(x)$ on $\mathbb{R}$ if $\lim_{\tau\to 0} \frac{1}{2\tau} \int_{|y-x|<\tau} |f(y) - f(x)|\, dy = 0$.

We say a function $F(x, y)$ on $\mathbb{R}^2$ has a diagonal support, if there is a constant C such that $F(x, y) = 0$, $\forall x, y$ with $|x - y| \geqslant C$. If F is a bounded function on $\mathbb{R}^2$ and has a diagonal support, then for any $x \in \mathbb{R}$, there is a constant C satisfying

$$\sup_j \left| 2^j \int\limits_{\mathbb{R}} f(y) F\left(2^j x, 2^j y\right) dy \right| \leqslant C Mf(x). \tag{2.2}$$

The above results may be found in Section 2, Chapter III of [15].

For the proof of Theorem 1.2, it is convenient to introduce the expression

$$\mathcal{P}_J f(x) = \sum_k \varphi_{Jk}(x) 2^J \sum_l r_l s_{J,k-l}.$$

For any x, there is only a finite number of k such that $\varphi_{Jk}(x) \neq 0$. Further, as known, there is also only a finite number of l such that $r_l \neq 0$. The above series converges for any $x \in \mathbb{R}$. Moreover, $\mathcal{P}_J f$ is a continuous function on $\mathbb{R}$.

It is not difficult to establish the relationship that $\mathcal{P}_J f = P_J (P_J f)'$, where P_J is the projection

$$P_J f(x) = \sum_k \langle f, \varphi_{Jk} \rangle \varphi_{Jk}(x).$$

Let $\psi^0(x)=\sum_l r_l\varphi(x+l)$ and $K_0(x,y)=\sum_k \varphi(x-k)\psi^0(y-k)$. Then K_0 has a diagonal support. It is easily seen that

$$\mathcal{P}_J f(x)=2^{2J}\int_{\mathbb{R}} f(y)K_0\big(2^J x,2^J y\big)\,dy. \tag{2.3}$$

Theorem 2.1. *Suppose that the orthogonal scaling function $\varphi\in C_0^1(\mathbb{R})$. If $f\in L^p(\mathbb{R})$, $f'\in L^q(\mathbb{R})$, $1\leqslant p,q\leqslant\infty$, then at any Lebesgue point x of f', $\mathcal{P}_J f(x)$ converges to $f'(x)$ as $J\to\infty$. Further, the convergence occurs in $L^q(\mathbb{R})$-norm for $1\leqslant q<\infty$.*

Proof. The proof is a standard argument. We first give an integral expression of $\mathcal{P}_J f$. For ψ^0 given as above, $\int_{\mathbb{R}}\psi^0(x)\,dx=0$ by $\sum r_l=0$. Therefore, the function Ψ^0 given by

$$\Psi^0(x)=-\int_{-\infty}^{x}\psi^0(y)\,dy$$

is compactly supported. Let $K_1(x,y)=\sum_k\varphi(x-k)\Psi^0(y-k)$. It follows from (2.3) and integration by parts that

$$\mathcal{P}_J f(x)=2^{J}\int_{\mathbb{R}} f'(y)K_1\big(2^J x,2^J y\big)\,dy. \tag{2.4}$$

It is easily seen that K_1 is a continuous function on $\mathbb{R}^2$ and that K_1 has a diagonal support. Then (2.2) tells us

$$\sup_J\big|\mathcal{P}_J f(x)\big|\leqslant CMf'(x). \tag{2.5}$$

Moreover, by the equality $\sum_{|l|\leqslant L-2} lr_l=-1$ [5] we have

$$\int_{\mathbb{R}} y\psi^0(y)\,dy=\sum_{|l|\leqslant L-2} r_l\int_{\mathbb{R}} y\varphi(y)\,dy-\sum_{|l|\leqslant L-2} r_l l\int_{\mathbb{R}}\varphi(y)\,dy=-\sum_{|l|\leqslant L-2} r_l l\int_{\mathbb{R}}\varphi(y)\,dy=1.$$

Consequently, by the first identity in (1.6) we have $\int_{\mathbb{R}} K_1(x,y)\,dy=1$, $x\in\mathbb{R}$. Which, together with (2.5), implies the pointwise convergence and norm convergence stated in the theorem. The arguments are standard in approximation theory (see, e.g., [13] and the references there). □

Proof of Theorem 1.2. First we show that the series in (1.8) converges absolutely and uniformly. Indeed, by Hölder's inequality, $|d_{jk}|,|s_{jk}|\leqslant C2^{j(1/2-1/p')}\|f\|_p$, where p' is the dual Hölder exponent of p. The estimations tell us that the magnitude of the summands in the series (1.8) are bounded by $C2^{j(1+1/p)}$, $j<J$. Consequently, the series (1.8) converges absolutely and uniformly, as claimed. As a byproduct, we obtain the continuity of functions $\mathcal{S}_J$ on $\mathbb{R}$.

Now the proof may proceed as in [13]. Define functions as follow

$$\psi^1(x)=\sum_{|l|\leqslant L-2}\alpha_l\psi(x+l),\qquad \psi^2(x)=\sum_{|l|\leqslant L-2}\beta_l\varphi(x+l),\qquad \psi^3(x)=\sum_{|l|\leqslant L-2}\gamma_l\psi(x+l).$$

Let

$$q_j(x,y)=\sum_k\big(\psi_{jk}(x)\psi^1_{jk}(y)+\psi_{jk}(x)\psi^2_{jk}(y)+\varphi_{jk}(x)\psi^3_{jk}(y)\big).$$

It is easily seen that

$$\mathcal{P}_{j+1}f-\mathcal{P}_j f=2^j\int_{\mathbb{R}} f(y)q_j(\cdot,y)\,dy=2^j\sum_k\bigg(\psi_{jk}\sum_l\alpha_l d_{j,k-l}+\psi_{jk}\sum_l\beta_l s_{j,k-l}+\varphi_{jk}\sum_l\gamma_l d_{j,k-l}\bigg),$$

and consequently, for any integer $M<J$,

$$\mathcal{P}_J f - \mathcal{P}_M f = \int_{\mathbb{R}} f(y) \sum_{M \leqslant j < J} 2^j q_j(\cdot, y)\, dy$$
$$= \sum_{M \leqslant j < J} \sum_k \Big(\psi_{jk} 2^j \sum_l \alpha_l d_{j,k-l} + \psi_{jk} 2^j \sum_l \beta_l s_{j,k-l} + \varphi_{jk} 2^j \sum_l \gamma_l d_{j,k-l} \Big).$$

It follows from (2.4) that, for $q < \infty$, $\lim_{M \to -\infty} \|\mathcal{P}_M f\|_{q'} \to 0$, where, as above, q' is the dual Hölder exponent of q. This means that the series in (1.8) converges to $\mathcal{P}_J f$ in $L^{q'}(\mathbb{R})$-norm. Now that its pointwise convergence has been established, it converges to $\mathcal{P}_J f$ a.e. As known, both $\mathcal{P}_J f$ and $\mathcal{S}_J$ are continuous functions, so we conclude

$$\mathcal{S}_J f(x) = \mathcal{P}_J f(x), \quad \forall x \in \mathbb{R}. \tag{2.6}$$

The proof is complete by Theorem 2.1. □

3. Convergence of wavelet thresholding estimator

This section gives a proof of Theorem 1.3.

We first make an observation for the decay of the wavelet coefficients, which is key to our study. For any x and j, there is a finite number of k such that $\varphi_{jk}(x) \neq 0$, or $\psi_{jk}(x) \neq 0$. In fact, $\varphi_{jk}(x) \neq 0$, or $\psi_{jk}(x) \neq 0$ only for those k which satisfies $2^j x - (L-1) \leqslant k \leqslant 2^j x$. Note that $\alpha_l = \beta_l = \gamma_l = 0$ for any l with $|l| \geqslant L-1$. Therefore, for any x and j, only the coefficients d_{ji} and s_{ji}, where $i \in I(x,j) := [2^j x - (2L-3), 2^j x + L - 2] \cap \mathbb{Z}$, are involved in (1.9) and (1.8). The observation is for a constant C

$$|d_{ji}| \leqslant C 2^{-3j/2} M f'(x), \quad \forall j \text{ and } i \in I(x,j). \tag{3.1}$$

The estimation follows by integrating by parts: $|d_{ji}| = 2^{-j} |\langle f', g_{ji} \rangle|$ with $g(x) = \int_{-\infty}^x \psi(t)\, dt$.

The most subtle consideration is about the series $\sum_{j \geqslant J} \sum_k \psi_{jk}(x) 2^j \sum_l \beta_l \hat{s}_{j,k-l}$. For our purpose, we make use of results concerned with subdivision schemes. Let $\ell(\mathbb{Z})$ be the space of sequences of real numbers. The subdivision operator S is defined by $S: u \to y = Su$ by

$$y_k = \sum_i h_{k-2i} u_i, \quad k \in \mathbb{Z},$$

where h_k, $k = 0, 1, \ldots, L-1$, are the coefficients in refinement equation (1.1). The subdivision operator is closely related to the reconstruction stage in wavelet based fast algorithm

$$s_{j+1,k} = \frac{1}{\sqrt{2}} \sum_i h_{k-2i} s_{ji} + \frac{1}{\sqrt{2}} \sum_i g_{k-2i} d_{ji}.$$

Lemma 3.1. *For any function $f \in L^p(\mathbb{R})$, $1 \leqslant p \leqslant \infty$ and any $j_1 < j_2$, $k \in \mathbb{Z}$, we have the following implication*

$$d_{ji} = 0, \quad \forall j_1 \leqslant j < j_2,\ i \in \big[k/2^{j_2-j} - (L-1),\ k/2^{j_2-j}\big] \cap \mathbb{Z} \quad \Rightarrow \quad s_{j_2,k} = 2^{-(j_2-j_1)/2} \big(S^{j_2-j_1}(s_{j_1})\big)_k,$$

where $s_j = (s_{ji})_i$.

Proof. By the construction of subdivision operator S, it is easily seen that the output number $y_k = (Su)_k$ only depends on the input numbers u_i, $i \in [(k-L+1)/2, k/2] \cap \mathbb{Z}$. Therefore, from the reconstruction stage we deduce that, for any $k \in \mathbb{Z}$, $s_{j+1,k}$ depends on the numbers s_{ji}, d_{ji}, $i \in [(k-L+1)/2, k/2] \cap \mathbb{Z}$. Consequently,

$$d_{ji} = 0 \quad \forall i \in \big[(k-L+1)/2, k/2\big] \cap \mathbb{Z} \quad \Rightarrow \quad s_{j+1,k} = \frac{1}{\sqrt{2}} \big(S(s_j)\big)_k.$$

In this case, $s_{j+1,k}$ is determined by the numbers s_{ji}, $i \in [(k-L+1)/2, k/2] \cap \mathbb{Z}$. The proof is complete by applying the fact above iteratively to the numbers s_{ji}, where

$$j = j_2 - 1, j_2 - 2, \ldots, j_1, \qquad i \in \big[k/2^{j_2-j} - \big(1/2^{j_2-j} + \cdots + 1/2\big)(L-1), k/2^{j_2-j}\big] \cap \mathbb{Z}. \qquad \square$$

Lemma 3.2. *Suppose that the orthogonal scaling function $\varphi \in C_0^1(\mathbb{R})$, $f \in L^p(\mathbb{R})$, $1 \leqslant p \leqslant \infty$, $f' \in L^q(\mathbb{R})$, $1 < q < \infty$ and J is an integer. For any $j > J$, $k \in \mathbb{Z}$, let $\tilde{s}_{jk} = \langle P_J f, \varphi_{jk}\rangle$. Then the series $\sum_{j\geqslant J}\sum_k |\psi_{jk}(x)2^j \times \sum_l \beta_l \tilde{s}_{j,k-l}|$ converges. Furthermore, there is a constant C, independent of f, such that for any J and f with Lebesgue point x of f'*

$$\left|\sum_{j\geqslant J}\sum_k \psi_{jk}(x)2^j \sum_l \beta_l \tilde{s}_{j,k-l}\right| \leqslant CMf'(x).$$

Proof. Recall that ψ^2 is defined in the proof of Theorem 1.2. We know from (1.7) that ψ^2 has two vanishing moments: $\int_{\mathbb{R}} \psi^2(y)\,dy = \int_{\mathbb{R}} y\psi^2(y)\,dy = 0$. Therefore the compactly supported function $\Psi^2(x) = -\int_{-\infty}^{x} \psi^2(y)\,dy$ satisfies $(\Psi^2)' = \psi^2$ and has one vanishing moment $\int_{\mathbb{R}} \Psi^2(x)\,dx = 0$.

As $\varphi' \in C_0(\mathbb{R})$, it is known that φ' has a positive Hölder exponent σ (see [12]). In other words, the modulus of continuity $\omega(\varphi', h)$ of φ' satisfies $\omega(\varphi', h) = O(h^\sigma)$, $h \to 0^+$. Furthermore, for any $f \in L^p(\mathbb{R})$, it is easily seen that $(P_J f)' = \sum_k s_{Jk} 2^{3J/2}\varphi'(2^J \cdot -k)$ also has the Hölder exponent σ: $\omega((P_J f)', h) = O(h^\sigma)$, $h \to 0^+$.

On the other hand, integrating by parts we have

$$2^j \sum_l \beta_l \tilde{s}_{j,k-l} = \langle (P_J f)', \Psi_{jk}^2\rangle,$$

where $\Psi_{jk}^2 = 2^{j/2}\Psi^2(2^j \cdot -k)$. As mentioned, Ψ^2 has compact support and one vanishing moment and $\omega((P_J f)', h) = O(h^\sigma)$, we have $|\langle (P_J f)', \Psi_{jk}^2\rangle| \leqslant C' 2^{-j(1/2+\sigma)}$. Recall again that, for any j, $\psi_{jk}(x) \neq 0$ only for a fixed number of k. Therefore, the series $\sum_{j\geqslant J}\sum_k |\psi_{jk}(x)2^j \sum_l \beta_l \tilde{s}_{j,k-l}|$ is bounded by $C\sum_{j\geqslant J} 2^{-j\sigma}$ and, consequently, it converges for any $x \in \mathbb{R}$.

Moreover, as $(P_J f)'$ is continuous on $\mathbb{R}$, an application of Theorem 1.2 to $P_J f$ yields $\mathcal{T}(P_J f)(x) = (P_J f)'(x)$. Note $\langle P_J f, \psi_{jk}\rangle = 0$ for any $k \in \mathbb{Z}$, $j \geqslant J$. Combining these with (2.6) we obtain

$$\sum_{j\geqslant J}\sum_k \psi_{jk}(x)2^j \sum_l \beta_l \tilde{s}_{j,k-l} = (P_J f)'(x) - \mathcal{S}_J f(x) = (P_J f)'(x) - \mathcal{P}_J f(x). \tag{3.2}$$

We consider now the function $(P_J f)'$. Let $K_2(x, y) = \sum_k \varphi'(x-k)\varphi(y-k)$. Clearly, K_2 has a diagonal support. For any x, as a compactly supported function, $K_2(x, \cdot)$ has one vanishing moment: $\int_{\mathbb{R}} K_2(x, y)\,dy = 0$. Therefore, the function $G(x, y) := \int_{-\infty}^{y} K_2(x, s)\,ds$ has a diagonal support.

By definition, we have $(P_J f)'(x) = 2^{2J}\int_{\mathbb{R}} f(y)K_2(2^J x, 2^J y)\,dy$. It follows from the definition of $G(x, y)$ that

$$(P_J f)'(x) = 2^{2J}\int_{x-c2^{-J}}^{x+c2^{-J}} \big(f(y) - f(x)\big)K_2\big(2^J x, 2^J y\big)\,dy = 2^J \int_{\mathbb{R}} f'(y)G\big(2^J x, 2^J y\big)\,dy.$$

Therefore there is a constant C satisfying $|(P_J f)'(x)| \leqslant CMf'(x)$ provided that x is a Lebesgue point of f'. The proof is complete by (2.5) and (3.2). □

The following result plays an important role in proof of Theorem 1.3.

Lemma 3.3. *Suppose that the orthogonal scaling function $\varphi \in C_0^1(\mathbb{R})$. If $f \in L^p(\mathbb{R})$, $1 \leqslant p \leqslant \infty$, $f' \in L^q(\mathbb{R})$, $1 < q < \infty$, then the series* (1.10) *converges for any $\lambda > 0$, $x \in \mathbb{R}$. Moreover, if we define a maximal function as following*

$$\mathcal{M}f(x) = \sup_{\lambda>0}\big|\mathcal{T}_\lambda f(x)\big|, \tag{3.3}$$

then there is a constant C, independent of f, satisfying

$$\mathcal{M}f(x) \leqslant CMf'(x). \tag{3.4}$$

Proof. Without loss of generality we suppose that x satisfies $Mf'(x) < \infty$. With C given as in (3.1), we choose $J > 0$ such that $\lambda/2 \leqslant C2^{-3J/2}Mf'(x) < \lambda$. Accordingly, we split the series in the right-hand side of (1.10) into two parts I_1 and I_2, where

$$I_1 = \sum_{j<J}\sum_k \left(\psi_{jk}(x)2^j \sum_l \alpha_l \delta(d_{j,k-l},\lambda) + \psi_{jk}(x)2^j \sum_l \beta_l \hat{s}_{j,k-l} + \varphi_{jk}(x)2^j \sum_l \gamma_l \delta(d_{j,k-l},\lambda)\right)$$

and

$$I_2 = \sum_{j\geqslant J}\sum_k \left(\psi_{jk}(x)2^j \sum_l \alpha_l \delta(d_{j,k-l},\lambda) + \psi_{jk}(x)2^j \sum_l \beta_l \hat{s}_{j,k-l} + \varphi_{jk}(x)2^j \sum_l \gamma_l \delta(d_{j,k-l},\lambda)\right).$$

The proof will be complete by the following two steps.

Step 1. We first prove the convergence of the series in I_1 and $|I_1| \leqslant CMf'(x)$. By construction, the series in I_1 is exactly $\mathcal{S}_J T_\lambda f(x)$. Its convergence has been obtained by Theorem 1.2. As known in (2.6), $\mathcal{S}_J T_\lambda f(x) = \mathcal{P}_J T_\lambda f(x)$. It remains to establish $|\mathcal{P}_J T_\lambda f(x)| \leqslant CMf'(x)$.

By the definition of $\mathcal{P}_j$ we know that $\mathcal{P}_j f = \mathcal{P}_j(P_j f)$ for any $f \in L^p(\mathbb{R})$. Therefore, it follows from (2.5) that

$$\|\mathcal{P}_J f\|_\infty = \big\|\mathcal{P}_J(P_J f)\big\|_\infty \leqslant C\big\|(P_J f)'\big\|_\infty. \tag{3.5}$$

We will apply the above inequality to $T_\lambda f - f$. Clearly, $\|P_J(T_\lambda f - f)\|_\infty \leqslant C2^{J/2}\lambda$ by (3.8). Then a Bernstein inequality implies that $\|(P_J(T_\lambda f - f))'\|_\infty \leqslant C2^{3J/2}\lambda$ (see, e.g., [14, Section 5 in Chapter II]). An application of (3.5) to $T_\lambda f - f$ yields $\|\mathcal{P}_J(T_\lambda f - f)\|_\infty \leqslant C2^{3J/2}\lambda \leqslant C'Mf'(x)$. Therefore, $|\mathcal{P}_J T_\lambda f(x)| \leqslant C'Mf'(x) + |\mathcal{P}_J f(x)| \leqslant CMf'(x)$, where the second inequality holds by (2.5), as desired.

Step 2. We prove the convergence of the series in I_2 and $|I_2| \leqslant CMf'(x)$. It follows from (3.1) that

$$\delta(d_{ji},\lambda) = 0, \quad \forall j \geqslant J,\ i \in I(x,j). \tag{3.6}$$

We easily obtain

$$\sum_{j\geqslant J}\sum_k \psi_{jk}(x)2^j \sum_l \alpha_l \delta(d_{j,k-l},\lambda) = \sum_{j\geqslant J}\sum_k \varphi_{jk}(x)2^j \sum_l \gamma_l \delta(d_{j,k-l},\lambda) = 0. \tag{3.7}$$

However, the estimation for $\sum_{j\geqslant J}\sum_k \psi_{jk}(x)2^j \sum_l \beta_l \hat{s}_{j,k-l}$ is much more involved. Recall that $P_J f = \sum_k s_{Jk}\varphi_{Jk}$ and $P_J(T_\lambda f) = \sum_k \hat{s}_{Jk}\varphi_{Jk}$ are the projections of f and $T_\lambda f$, respectively. By (1.3a) we have

$$\big\|P_J(T_\lambda f) - (P_J f)\big\|_\infty = \left\|\sum_{j<J}\sum_k \big(\delta(d_{j,k},\lambda) - d_{jk}\big)\psi_{jk}\right\|_\infty \leqslant \sum_{j<J} C2^{j/2}\lambda = C2^{J/2}\lambda, \tag{3.8}$$

where the inequality holds due to the compact support assumption on ψ. This together with the stability of the shifts of φ implies

$$|\hat{s}_{Jk} - s_{Jk}| \leqslant C\lambda, \quad \forall k \in \mathbb{Z}.$$

As in Lemma 3.2, denote $\tilde{s}_{ji} = \langle P_J f, \varphi_{ji}\rangle$ and $\tilde{s}_j = (\tilde{s}_{ji})_i$ for $j > J$. Note that the wavelet coefficient $\langle P_J f, \psi_{ji}\rangle = 0$ for any $j \geqslant J$, $i \in \mathbb{Z}$. An application of Lemma 3.3 to function $P_J f$ yields that $\tilde{s}_{jk} = 2^{-(j-J)/2}(S^{j-J}(s_J))_k$ for any $k \in \mathbb{Z}$.

We now consider $\hat{s}_j = (\hat{s}_{ji})_i$, $j \geqslant J$. Let $j_2 = j$, $j_1 = J$. Equality (3.6) implies that the wavelet coefficients $\langle T_\lambda f, \psi_{j'i}\rangle$ satisfy

$$\langle T_\lambda f, \psi_{j'i}\rangle = 0, \quad \forall J \leqslant j' < j,\ i \in \big[\big(2^j x - 2L + 3\big)/2^{j-j'} - (L-1), \big(2^j x + L - 1\big)/2^{j-j'}\big] \cap \mathbb{Z}.$$

Applying Lemma 3.1 to function $T_\lambda f$, we have $\hat{s}_{jk} = 2^{-(j-J)/2}(S^{j-J}(\hat{s}_J))_k$ for any $k \in I(x,j)$.

Finally, $\hat{s}_{jk} - \tilde{s}_{jk} = 2^{-(j-J)/2}(S^{j-J}(\hat{s}_J - s_J))_k$ for any $j \geqslant J$, $k \in I(x,j)$.

Since $\sum_l l^i \beta_l = 0$ for $i = 0, 1$ (see (1.7)) and φ' is continuous on $\mathbb{R}$, we conclude that

$$\begin{aligned}\left|\sum_l \beta_l(\hat{s}_{j,k-l} - \tilde{s}_{j,k-l})\right| &= \left|\sum_l \beta_l 2^{-(j-J)/2}\big(S^{j-J}(\hat{s}_J - s_J)\big)_{k-l}\right| \\ &\leqslant C2^{(-1-\tau)(j-J)}2^{-(j-J)/2}\big\|(\hat{s}_J - s_J)\big\|_\infty \\ &\leqslant C2^{(-1-\tau)(j-J)}2^{-(j-J)/2}\lambda, \quad \forall j \geqslant J,\ k \in I(x,j),\end{aligned}$$

where τ is a positive constant, see, e.g., [6]. Therefore,

$$\sum_{j\geqslant J}\sum_{k}\left|\psi_{jk}(x)2^j\sum_{l}\beta_l(\hat{s}_{j,k-l}-\tilde{s}_{j,k-l})\right|\leqslant\sum_{j\geqslant J}\sum_{k}C2^{3j/2}2^{(-1-\tau)(j-J)}2^{-(j-J)/2}\lambda\leqslant C'2^{3J/2}\lambda.$$

This together with Lemma 3.2 and (3.7) yields the convergence of the series and $|I_2|\leqslant CMf'(x)$. □

We are now in a position to prove Theorem 1.3.

Proof of Theorem 1.3. For any $\varepsilon>0$, since $f'\in L^q(\mathbb{R})$ and x is a Lebesgue point of f', we can find a function $g\in C_0^\infty(\mathbb{R})$ for which $M(f'-g')(x)\leqslant C\varepsilon$. As $g\in C_0^\infty(\mathbb{R})$, we know from [14, Section 7 in Chapter III] that P_Jg and $(P_Jg)'$ converge uniformly to g and g' on $\mathbb{R}$ respectively as $J\to\infty$. On the other hand, it is easily seen that, for some constant C,

$$\left|P_Jg(x)\right|\leqslant C2^{J/2}\|g\|_2,\qquad\left|(P_Jg)'(x)\right|\leqslant C2^{3J/2}\|g\|_2,\quad\forall x\in\mathbb{R}.$$

Consequently, both P_Jg and $(P_Jg)'$ converge uniformly to 0 on $\mathbb{R}$ as $J\to-\infty$. Therefore, there are integers $J_1<J_2$ such that the function $h:=P_{J_2+1}g-P_{J_1}g$ satisfies $\|g-h\|_\infty\leqslant\varepsilon$ and $\|(g-h)'\|_\infty\leqslant\varepsilon$. Clearly, $h=\sum_{J_1\leqslant j\leqslant J_2}\sum_{k\in\mathbb{Z}}b_{jk}\psi_{jk}$, where $b_{jk}=\langle g,\psi_{jk}\rangle$. Since g is a compactly supported function, there are only finitely many nonzero wavelet coefficients b_{jk} in above equality. Consequently, $M(g-h)'(x)\leqslant C\|(g-h)'\|_\infty\leqslant C\varepsilon$. It in turn gives $M(f-h)'(x)\leqslant C\varepsilon$.

By $h\in C_0^1(\mathbb{R})$ and Theorem 1.2, $\mathcal{T}h(y)=h'(y)$ for any $y\in\mathbb{R}$. Therefore,

$$\left|\mathcal{T}_\lambda f(x)-f'(x)\right|\leqslant\left|\mathcal{T}_\lambda f(x)-\mathcal{T}_\lambda(f-h)(x)-\mathcal{T}h(x)\right|+\left|\mathcal{T}_\lambda(f-h)(x)\right|+\left|f'(x)-h'(x)\right|.$$

For any $\lambda>0$, $|\mathcal{T}_\lambda(f-h)(x)|\leqslant CM(f-h)'(x)\leqslant C\varepsilon$ by Lemma 3.3. For the third term, we have $|f'(x)-h'(x)|\leqslant M(f-h)'(x)\leqslant C\varepsilon$ by (2.1). For the estimate of the first term, let $\bar{f}_\lambda:=T_\lambda f-T_\lambda(f-h)-h$. As $(\bar{f}_\lambda)'$ is continuous on $\mathbb{R}$, and x is a Lebesgue point of both f and $f-h$, it follows from Theorem 1.2 and Lemma 3.3 that

$$\mathcal{T}\bar{f}_\lambda(x)=\mathcal{T}_\lambda f(x)-\mathcal{T}_\lambda(f-h)(x)-\mathcal{T}h(x).$$

We now bound $\mathcal{T}\bar{f}_\lambda(x)$. For any $j>J_2$, we have by $b_{jk}=0$ ($\forall k\in\mathbb{Z}$) that

$$\langle\bar{f}_\lambda,\psi_{jk}\rangle=\delta(d_{jk},\lambda)-\delta(d_{jk}-b_{jk},\lambda)-b_{jk}=\delta(d_{jk},\lambda)-\delta(d_{jk},\lambda)=0,\quad\forall k\in\mathbb{Z}.$$

This tells us that $\bar{f}_\lambda=P_{J_2}\bar{f}_\lambda$, and which in turn, together with (1.3a) implies that, for any $y\in\mathbb{R}$,

$$\begin{aligned}\left|\bar{f}_\lambda(y)\right|&=\left|\sum_{J_1\leqslant j\leqslant J_2}\sum_{k}\Big(\big(\delta(d_{jk},\lambda)-d_{jk}\big)-\big(\delta(d_{jk}-b_{jk},\lambda)-(d_{jk}-b_{jk})\big)\Big)\psi_{jk}(y)\right|\\&\leqslant C'\lambda\sum_{J_1\leqslant j\leqslant J_2}\sum_{k}\left|\psi_{jk}(y)\right|\leqslant C\lambda,\end{aligned}$$

where C is independent of y.

As in the proof of Lemma 3.3, a Bernstein inequality yields that $\|(\bar{f}_\lambda)'\|_\infty\leqslant C2^{J_2}\lambda$. It follows from Lemma 3.3 that $|\mathcal{T}\bar{f}_\lambda(x)|\leqslant C2^{J_2}\lambda$ (see also, e.g., [14, Section 5 in Chapter II]), which can be made less than any positive number by choosing λ sufficiently small. The proof of the theorem is complete. □

References

[1] F. Abramovich, U. Amato, C. Angelini, On optimality of Bayesian wavelet estimators, Scand. J. Statist. 31 (2004) 217–234.
[2] F. Abramovich, B.W. Silverman, Wavelet decomposition approaches to statistical inverse problems, Biometrika 85 (1998) 115–129.
[3] G. Beylkin, On the representation of operators in bases of compactly supported wavelets, SIAM J. Numer. Anal. 29 (1992) 1716–1740.
[4] G. Beylkin, Wavelets and fast numerical algorithms, in: Proceedings of Symposia in Applied Mathematics, vol. 47, 1993, pp. 89–117.
[5] G. Beylkin, R.R. Coifman, V. Rokhlin, Fast wavelet transforms and numerical algorithms I, Comm. Pure Appl. Math. 44 (1991) 141–183.
[6] D.R. Chen, R.Q. Jia, S. Riemenschneider, Convergence of vector subdivision schemes in Sobolev spaces, Appl. Comput. Harmon. Anal. 12 (2002) 128–149.
[7] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.
[8] D.L. Donoho, Denoising via soft thresholding, IEEE Trans. Inform. Theory 41 (1995) 613–627.

[9] D.L. Donoho, Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition, Appl. Comput. Harmon. Anal. 2 (1995) 101–126.
[10] D.L. Donoho, I.M. Johnstone, Ideal spatial adaption via wavelet shrinkage, Biometrika 81 (1994) 425–455.
[11] D.L. Donoho, I.M. Johnstone, Minimax estimation via wavelet shrinkage, Ann. Statist. 26 (1998) 879–921.
[12] R.Q. Jia, S.D. Riemenschneider, D.X. Zhou, Smoothness of multiple refinable functions and multiple wavelets, SIAM J. Matrix Anal. Appl. 21 (1999) 1–28.
[13] S. Kelly, M.A. Kon, L.A. Raphael, Local convergence for wavelet expansions, J. Funct. Anal. 126 (1994) 102–138.
[14] Y. Meyer, Ondelettes et Opérateurs, Hermann, Paris, 1990.
[15] E. Stein, Sigular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, NJ, 1970.
[16] T. Tao, On the almost everywhere convergence of wavelet summation methods, Appl. Comput. Harmon. Anal. 3 (1996) 384–387.
[17] T. Tao, B. Vidakovic, Almost everywhere behavior of general wavelet shrinkage operators, Appl. Comput. Harmon. Anal. 9 (2000) 72–82.