On the almost everywhere convergence of wavelet summation methods

Terence Tao

Abstract

Let ψ be a rapidly decreasing one-dimensional wavelet. We show that the wavelet expansion of any L^p function converges pointwise almost everywhere under the wavelet projection, hard sampling, and soft sampling summation methods, for $1 < p < \infty$. In fact, the partial sums are uniformly dominated by the Hardy-Littlewood maximal function.

0. Introduction

In this note we will discuss the almost everywhere convergence of various summation methods for wavelet expansions. We shall not aim for maximal generality; instead, we will restrict our attention to the expansion of an L^p function f in one dimension, for $1 < p < \infty$, generated by a 0-regular orthogonal wavelet ψ. By this we mean that ψ has integral 0, is bounded and rapidly decreasing, and that the set of functions $\{\psi_{j,k}\} = \{2^{j/2}\psi(2^j \cdot - k)\}$ for $j, k \in \mathbf{Z}$, form an orthonormal basis for $L^2(\mathbf{R})$. The wavelet expansion of f is given by $f \sim \sum_{j,k} a_{j,k}\psi_{j,k}$, where $a_{j,k} = \langle f, \psi_{j,k}\rangle$. In [4] it is shown that this series converges unconditionally in L^p norm to f; however, the pointwise behaviour of this series is more dependent on the summation method used and is less well understood. The summation methods that are used in applications almost always rely on one of the following three partial summation operators:

the *wavelet projections* $\qquad P_J f = \sum_{j<J;k} a_{j,k}\psi_{j,k};$

the *hard sampling* operators $\qquad T_\lambda f = \sum_{|a_{j,k}|>\lambda} a_{j,k}\psi_{j,k};$

the *soft sampling* operators $\qquad \tilde{T}_\lambda f = \sum_{|a_{j,k}|>\lambda} \operatorname{sgn}(a_{j,k})\big(|a_{j,k}| - \lambda\big)\psi_{j,k}.$

Note that the sum in $P_J f$ converges absolutely since ψ is rapidly decreasing and $a_{j,k} = O(2^{-j/2}2^{j/p}\|f\|_p)$, by Hölder's inequality. We shall prove the absolute convergence of the other two sums in the next section (though for $p \geq 2$ one always has a finite number of terms).

Theorem 1. *If $f \in L^p(\mathbf{R})$ for some $1 < p < \infty$, then for almost every x one has*

$$\lim_{J\to\infty} P_J f(x) = \lim_{\lambda\to 0} T_\lambda f(x) = \lim_{\lambda\to 0} \tilde{T}_\lambda f(x) = f(x).$$

The almost everywhere convergence of the wavelet projection operators was already shown (under more general hypotheses) in [3]. The non-linear hard and soft sampling methods are used frequently in applications

1991 *Mathematics Subject Classification.* Primary 42C15; Secondary 40A30.
Keywords: hard sampling, soft sampling, almost-everywhere convergence.

(see [1],[2]) but this seems to be the first pointwise convergence result for these methods. In contrast, we have learned recently that T.W. Körner has found a counterexample to the pointwise convergence of the hard sampling operator for the Fourier series expansion. We also note that if one assumes some minimal regularity (e.g. right-continuity) for ψ, then one can get the pointwise convergence of the operators of Theorem 1 on the entire Lebesgue set of f (c.f. [3]).

Finally, we show that there are (somewhat artificial) wavelet summation methods which need not converge pointwise, at least for wavelets with little regularity.

Theorem 2. *Suppose that ψ is the Haar wavelet $\chi_{[0,1/2)} - \chi_{[1/2,1)}$. Fix $\alpha > 0$, $1 < p < \infty$ and consider the operator T^α_λ defined on $L^p(\mathbf{R})$ for $\lambda > 0$ by*

$$T^\alpha_\lambda f = \sum_{|a_{j,k}| \geq \alpha^j \lambda} a_{j,k} \psi_{j,k}$$

(a) If $\alpha \neq 2^{-1/2}$, then for all $f \in L^p$, one has $\lim_{\lambda \to 0} T^\alpha_\lambda f(x) = f(x)$ at every Lebesgue point x of f.

(b) If $\alpha = 2^{-1/2}$, then there exists an $f \in L^p$ for which $\limsup_{\lambda \to 0} |T^\alpha_\lambda f(x)| = \infty$ for all $x \in \mathbf{R}$.

The series defining $T^\alpha_\lambda f$ need not converge absolutely; if this is the case, we shall define the value of the series to be the pointwise limit (where it exists) of its formal wavelet projections, so

$$T^\alpha_\lambda f(x) = \lim_{J \to \infty} \sum_{\substack{|a_{j,k}| \geq \alpha^j \lambda \\ j < J}} a_{j,k} \psi_{j,k}(x).$$

The operator T^α_λ can be thought of as a variant of the hard thresholding operator T_λ, but with a weighted threshold that depends on the resolution level. The reason that convergence fails when $\alpha = 2^{-1/2}$ can be sketched as follows. For this value of α one is effectively applying a threshold to the "heights" of the summands $a_{j,k}\psi_{j,k}$ rather than to the co-efficients $a_{j,k}$. By "fine-tuning" the magnitudes of terms with comparable heights, one can effectively sum such terms in an arbitrary order, and (with a suitable choice of ordering) this allows one to get large non-cancellatory contributions to the partial sums $T_\lambda f(x)$ for each x (the exact values of λ for which non-cancellation occurs depends on x). The ability to get significant non-cancellation seems to be due to the lack of regularity (both smoothness and moment conditions) of the Haar wavelet; it may be possible that better convergence results can be obtained for more regular wavelets.

The work in this paper originated from a question posed by Elias Stein. The author also thanks Anna Gilbert for her proofreading, the loan of several books and articles, and for many valuable discussions.

1. Proof of Theorem 1.

In the following the symbol C will denote various constants that depend only on ψ. In order to make the notion of a Lebesgue point well defined, we shall not adopt the practice of identifying functions which agree almost everywhere. Recall that a point $x \in \mathbf{R}$ is a Lebesgue point of f if $\frac{1}{2r}\int_{-r}^{r} |f(x+t) - f(x)|\, dt$ converges to 0 as r tends to 0; it is a classical theorem of Lebesgue that, for any locally integrable function f, almost every element of $\mathbf{R}$ is a Lebesgue point of f. Closely related to the concept of Lebesgue point is the Hardy-Littlewood maximal operator M; this sublinear operator is defined by

$$Mf(x) = \sup_{r>0} \frac{1}{2r} \int_{-r}^{r} |f(x+t)|\, dt.$$

Many expressions, especially ones involving approximations to the identity, can be controlled in magnitude by the maximal function. In particular, if $\phi(x)$ is a function such that $|\phi(x)| \leq C_N A(1 + \frac{|x-x_0|}{r})^{-N}$ for all $N > 0$ and some fixed A,r,x_0, then

$$|\int_{\mathbf{R}} f(x)\phi(x)\, dx| \leq CArMf(x_0).$$

Another such estimate is the simple inequality $|f(x)| \leq Mf(x) \leq \|f\|_\infty$, which holds whenever x is a Lebesgue point of f. Proofs of these statements can be found in [5].

Let f be an arbitrary function in $L^p(\mathbf{R})$. We first show that to prove Theorem 1, it suffices to show the pointwise estimates

$$\sup_J |P_J f(x)|,\ \sup_\lambda |T_\lambda f(x)|,\ \sup_\lambda |\tilde{T}_\lambda f(x)| \leq CMf(x) \tag{1}$$

for almost every x. If our operators were linear, the implication would follow from standard arguments (see [5]), but only the P_J are linear and so the argument requires a little care. We shall show the argument for the $\tilde{T}_\lambda$ only, as it is the most non-linear of the three; the proof for the other two operators is analogous (and can be simplified slightly). Assume now that (1) holds for some fixed f and x. Since we only require convergence almost everywhere, we may assume that x is a Lebesgue point of f and of every $\psi_{j,k}$. Choose any $\varepsilon > 0$. By the usual method of convolving f with a smooth approximation to the identity, and using the fact that x is a Lebesgue point of f, and that f is in an L^p class for some finite p, one can find a g in $C_0^\infty(\mathbf{R})$ for which $|M(f-g)(x)| \leq C\varepsilon$. In turn, one can approximate g uniformly almost everywhere by a finite linear combination h of wavelets (for this fact, see [4]). Thus one can find such an h for which $|f(x) - h(x)| \leq |M(f-h)(x)| \leq C\varepsilon$. To conclude the argument, one observes that $\tilde{T}_\lambda f(x) = h(x) + \tilde{T}_\lambda(f-h)(x) + O_h(\lambda)$ as $\lambda \to 0$. To see this, note that the wavelet co-efficients of $\tilde{T}_\lambda f$ and $h + \tilde{T}_\lambda(f-h)$ differ by $O(\lambda)$ when the wavelet occurs in the expansion of h, and are identical otherwise. Thus $|\tilde{T}_\lambda f(x) - f(x)| = O(\varepsilon)$ for sufficiently small λ, so $\lim_{\lambda\to 0} \tilde{T}_\lambda f(x) = f(x)$, as desired.

We now show that (1) holds for the P_J. We note that this estimate was first proven in [3] under weaker decay hypotheses on ψ and f; we shall give an alternative proof below. The argument relies on two observations. The first is the crude estimate

$$|\psi_{j,k}(x)\overline{\psi_{j,k}(y)}| \leq C_N 2^{-j}(1 + 2^j|x-y|)^{-N}(1 + |2^j x - k|)^{-N} \tag{2}$$

for all $N > 0$; this is a simple consequence of the rapid decrease of ψ and the easy inequality $1 + 2^j|x - y| \leq (1 + |2^j x - k|)(1 + |2^j y - k|)$. The second is the fact that

$$\sum_{j,k} \psi_{j,k}(x)\overline{\psi_{j,k}(y)} = 0$$

for almost every $x, y \in \mathbf{R}$. Note that this sum converges absolutely for $x \neq y$ by (2). The identity follows from the fact that the $\{\psi_{j,k}\}_{j,k\in\mathbf{Z}}$ form an orthonormal basis for $L^2(\mathbf{R})$, so that the above summation converges in a distributional sense to $\delta(x - y)$. Since the series converges both distributionally and absolutely almost everywhere, the two limits must agree almost everywhere, hence the claim.

As a consequence of these two observations, we see that the kernel defined almost everywhere by

$$K(x, y) = \sum_{j<0;k} \psi_{j,k}(x)\overline{\psi_{j,k}(y)} = -\sum_{j\geq 0;k} \psi_{j,k}(x)\overline{\psi_{j,k}(y)}$$

converges absolutely and is bounded and rapidly decreasing away from the diagonal. However, we have

$$\begin{aligned} P_J f(x) &= \sum_{j<J;k} \psi_{j,k}(x) \int_{\mathbf{R}} \overline{\psi_{j,k}(y)} f(y)\ dy \\ &= \int_{\mathbf{R}} \sum_{j<J;k} \psi_{j,k}(x)\overline{\psi_{j,k}(y)} f(y)\ dy \\ &= \int_{\mathbf{R}} 2^J K(2^{-J}x, 2^{-J}y) f(y)\ dy \end{aligned}$$

from Fubini's theorem; the conditions for Fubini's theorem follow from (2), the fact that f is in an L^p class for some $p > 1$, and Hölder's inequality. Since $K(2^{-J}x, 2^{-J}y)$, as a function of y, is rapidly decreasing away from x, the desired estimate (1) now follows from the maximal function estimates mentioned at the start of this section.

We now show that (1) holds for the T_λ. Fix $f \in L^p(\mathbf{R})$, $\lambda \in \mathbf{R}$, and $x \in \mathbf{R}$. Since we may assume that $Mf(x)$ is nonzero, we can choose a J for which $\lambda \sim 2^{-J/2} Mf(x)$; one can think of J as the finest scale for which significant terms in the summation for $T_\lambda f(x)$ can appear. We will show that $|T_\lambda f(x) - P_J f(x)| \leq CMf(x)$, which will imply (1) from our results for the P_J. We shall first need some simple estimates. For each $j, k \in \mathbf{Z}$, let $D(j, k)$ be the integer part of $2 + |2^j x - k|$. Note that for each j and for each integer $d \geq 2$ there are at most two k for which $D(j, k) = d$. From the rapid decrease of ψ and the usual estimates on approximations to the identity, we have that, for all $N > 0$,

$$\begin{aligned} |\psi_{j,k}(x)| &\leq C_N 2^{j/2} D(j, k)^{-N} \text{ and} \\ |a_{j,k}| &\leq C2^{-j/2} D(j, k) Mf(x); \text{ thus} \\ |a_{j,k}\psi_{j,k}(x)| &\leq C_N D(j, k)^{-N} Mf(x). \end{aligned}$$

Note that the second estimate implies that $|a_{j,k}|$ is greater than λ only if $j < J + C \log D(j, k)$.

The estimate for $|T_\lambda f(x) - P_J f(x)|$ is now straightforward. Indeed, we have

$$\begin{aligned}
|T_\lambda f(x) - P_J f(x)| &= \Big| \sum_{|a_{j,k}|>\lambda} a_{j,k}\psi_{j,k}(x) - \sum_{j<J;k} a_{j,k}\psi_{j,k}(x) \Big| \\
&\leq \Big| \sum_{|a_{j,k}|\leq\lambda; j<J} a_{j,k}\psi_{j,k}(x) \Big| + \Big| \sum_{|a_{j,k}|>\lambda; J\leq j} a_{j,k}\psi_{j,k}(x) \Big| \\
&\leq \sum_{j<J;k} \lambda |\psi_{j,k}(x)| + \sum_{J\leq j<J+C\log D(j,k)} |a_{j,k}\psi_{j,k}(x)| \\
&\leq \sum_{d=2}^{\infty} \sum_{j<J} C_N \lambda 2^{j/2} d^{-N} + \sum_{d=2}^{\infty} \sum_{J\leq j<J+C\log d} C_N d^{-N} Mf(x) \\
&\leq C\lambda 2^{J/2} + \sum_{d=2}^{\infty} C Mf(x) d^{-N} \log d \\
&\leq CMf(x) + CMf(x),
\end{aligned}$$

as desired. Note that this argument also shows that the sum defining $T_\lambda f(x)$ is absolutely convergent, since we know the sum for $P_J f(x)$ is.

The proof of (1) for $\tilde{T}_\lambda$ is almost identical to that for T_λ, although one must use the estimate

$$\sum_{|a_{j,k}|>\lambda; j<J} \operatorname{sgn}(a_{j,k})(|a_{j,k}| - \lambda)\psi_{j,k}(x) = \sum_{j<J;k} a_{j,k}\psi_{j,k}(x) + O\Big(\sum_{j<J;k} \lambda|\psi_{j,k}(x)|\Big)$$

in the above argument; we omit the details. Thus (1) holds for all three operators under consideration, and the theorem is proved. $\square$

We remark that these results can be extended easily to higher dimensions, to bi-orthogonal wavelets, and to decompositions involving more than one "mother wavelet". The decay conditions on ψ can also be relaxed; for instance, one can get the pointwise convergence result for the P_J alone if one merely assumes that ψ has an integrable radial majorant. The proofs of these assertions are simple applications of the ideas developed in [3].

2. Proof of Theorem 2.

The proof of (a) is an easy generalization of the corresponding argument for the T_λ in the proof of Theorem 1, although if $\alpha < 2^{-1/2}$ one has to use $I - P_J$ instead of P_J as the approximating operator. (Here J is chosen so that $\alpha^J \lambda \sim 2^{-J/2} Mf(x)$). We omit the details. Instead of using the argument in the previous section to bound P_J, one can compute $P_J f$ explicitly to achieve estimate (1) for all $x \in \mathbf{R}$, which will ensure pointwise convergence at every Lebesgue point, and not just almost everywhere. Alternatively, one can use the methods in [3].

Now suppose that $\alpha = 2^{-1/2}$. We shall construct, for each integer $N \geq 1$, a function $f_N = \sum_{j,k} a^N_{j,k}\psi_{j,k}$ on $[0, 1)$ for which:

(i) All but a finite number of the $a_{j,k}^N$ are 0 and those that are not zero satisfy $1 < 2^{j/2} a_{j,k}^N < 2$;

(ii) $\|f_N\|_p \le C_p N^{1/2}$ for every $1 < p < \infty$;

(iii) $\sup_\lambda |T_\lambda^\alpha f_N(x)| > N/3$ for every $x \in [0,1)$.

Once we do this, one can see that, if one chooses an increasing sequence of natural numbers M_i whose successive differences are sufficiently large, the function

$$F(x) = \sum_{i=1}^{\infty} 2^{-i} \sum_{k=0}^{2^{M_i}-1} f_{3^i}(2^{M_i} x - k)$$

will be in $L^p([0,1))$ for every $1 < p < \infty$, but is such that $\limsup_\lambda |T_\lambda^\alpha F(x)| = \infty$ for every $x \in [0,1)$. The function $f(x) = \sum_{n=-\infty}^{\infty} 2^{-|n|} F(x-n)$ thus satisfies all the required conditions in (b).

We shall define f_N explicitly:

$$f_N(x) = \sum_{j=0}^{N-1} \sum_{k=0}^{2^j-1} (1 + 2^{-j}(k + \frac{1}{2}))\, \psi(2^j x - k).$$

The reason for this particular choice of co-efficients is that each $T_\lambda^\alpha f_N$ will consist only of the terms $a_{j,k}^N \psi_{j,k}$ corresponding to wavelets whose "centers" are to the right of $\lambda - 1$. At $x = \lambda - 1$, the nature of the Haar wavelet will cause this sum to be large and positive.

We now show that f_N has the properties (i-iii) stated above. The verification of (i) is immediate, and (ii) follows quickly from Khinchin's inequality (see [4]). To prove (iii), fix an $x \in [0,1)$ and observe that all but N of the factors $\psi(2^j x - k)$ in the sum defining $f_N(x)$ are zero, and the remainder are either $+1$ or -1. Furthermore, this factor is nonnegative if $1 + 2^{-j}(k + \frac{1}{2}) > 1 + x$ and nonpositive otherwise. Thus, by inspection of the definition of T_λ^α for $\lambda = 1 + x$ and $\lambda = 2$, we have

$$\begin{aligned}
T_{1+x}^\alpha f(x) &= \sum_{\psi(2^j x - k) = +1} (1 + 2^{-j}(k + \frac{1}{2})), \text{ and} \\
T_{1+x}^\alpha f(x) - T_2^\alpha f(x) &= \sum_{\psi(2^j x - k) = -1} (1 + 2^{-j}(k + \frac{1}{2})); \text{ so} \\
2T_{1+x}^\alpha f(x) - T_2^\alpha f(x) &= \sum_{\psi(2^j x - k) \ne 0} (1 + 2^{-j}(k + \frac{1}{2})) > N,
\end{aligned}$$

which implies (iii). This proves the theorem. □

References

[1] A. Averbuch, G. Beylkin, R. Coifman, M. Israeli, *Fast Adaptive Algorithm for the Elliptic equations with periodic boundary conditions*, preprint.

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

[3] S. Kelly, M.A. Kon, and L.A. Raphael, *Local convergence for wavelet expansions*, J. Funct. Anal. **126** (1994), 102-138.

[4] Y. Meyer, "Ondelettes", Hermann, Paris, 1990.

[5] E. Stein, "Singular Integrals and Differentiability Properties of Functions", Princeton University Press, 1970.

Department of Mathematics
Princeton University
Princeton, NJ 08544
tao@math.princeton.edu