Near Best Tree Approximation *

R. G. Baraniuk, R. A. DeVore,
G. Kyriazis, and X. M. Yu

November 29, 2000

Abstract

Tree approximation is a form of nonlinear wavelet approximation that appears naturally in applications such as image compression and entropy encoding. The distinction between tree approximation and the more familiar n-term wavelet approximation is that the wavelets appearing in the approximant are required to align themselves in a certain connected tree sturcture. This makes their positions easy to encode. Previous work [CDGO], [CDDD] has established upper bounds for the error of tree approximation for certain (Besov) classes of functions. The present paper, in contrast, studies tree approximation of individual functions with the aim of characterizing those functions with a prescribed approximation error. This is accomplished in the case that the approximation error is measured in L_2, or in the case $p \neq 2$, in the Besov spaces $B_p^0(L_p)$, which is close to (but not the same as) L_p. Our characterization of functions with a prescribed approximation order in these cases is given in terms of a certain maximal function applied to the wavelet coefficients.

AMS subject classification: tree approximation, nonlinear approximation, wavelets, approximation classes, adaptive approximation
41A25, 41A46, 65F99, 65N12, 65N55.
Key Words: compression, n-term approximation, encoding, approximation classes.

1 Introduction

Tree approximation is a form of nonlinear wavelet approximation that occurs in the application of wavelets to image processing and adaptive methods. The usual nonlinear wavelet approximation, known as n-term approximation, seeks to approximate a target function f by a linear combination of at most n wavelets. The approximation properties of n-term wavelet approximation are by now well-known [DJP], [CDH]. In particular, there is a characterization of the functions that can be approximated with a prescribed

*This work has been supported in part by Office of Naval Research contracts N0014–91–J1343 and N00014–99-1–0813, National Science Foundation grants MIP–9457438 and CCR–9973188, and Army Research Office contract DAAG55–98–1–0002.

rate of approximation in terms of the wavelet coefficients of the target function (and this can sometimes be restated directly in terms of the smoothness of the target function).

One deficiency in n-term approximation, which manifests itself in computations, is that the terms appearing in an n-term approximant can involve any wavelets. Thus, for example, to encode an n-term approximant, one has to not only to assign bits for the coefficients of the decomposition but also to indicate which particular wavelets appear in the decomposition. The cost of the latter can be prohibitive. Tree approximation is designed to overcome this objection.

Wavelets are naturally indexed on dyadic cubes. Dyadic cubes have a tree structure under inclusion. Tree approximation differs from n-term approximation in requiring that the wavelet functions in the approximant should correspond to a subtree of the indexing tree. This allows for efficient encoding of the positions of the wavelets used in the approximation.

Previous work [CDDD], [CDGO] has given upper bounds for the efficiency of tree approximation on certain smoothness classes and has shown the optimality of these bounds in the sense of Kolmogorov entropy. The purpose of the present paper is to go further and actually characterize the functions with a prescribed rate of tree approximation. The characterization is given in terms of a certain maximal function applied to the wavelet coefficients of the target function.

The second section of this paper introduces the wavelet setting in which we shall work and states the main results of this paper. The remaining sections prove our results.

2 Main results

In order to describe tree approximation and the results of this paper, we begin by recalling the usual setting for wavelet decompositions and nonlinear approximation. The results of this paper can be established for quite general wavelet decompositions but we shall restrict oursleves to the compactly supported biorthogonal wavelets of Cohen, Daubechies, and Feauveau [CDF]. These include, as a special case, the orthogonal wavelets of compact support of Daubechies [Da1]. A good reference for these bases and their properties is Chapter 8 of the monograph of Daubechies [Da2].

The construction of biorthogonal wavelets begins with two compactly supported univariate scaling functions ϕ and $\tilde{\phi}$ whose shifts are in duality:

$$\int_{\mathbb{R}} \phi(x-k)\tilde{\phi}(x-k')\,dx = \delta(k-k'), \quad k,k' \in \mathbb{Z},$$

with δ the Kronecker delta. Associated to each of the scaling functions are mother wavelets ψ and $\tilde{\psi}$.

These functions can be used to generate a wavelet basis for the $L_p(\mathbb{R}^d)$ spaces as follows. For any function g defined on $\mathbb{R}^d$ and any dyadic cube $I = 2^j[\mathbf{k}, \mathbf{k}+\mathbf{1}]$, $\mathbf{k} \in \mathbb{Z}^d$, we define the function

$$g_I(x) := 2^{jd/2} g(2^j(x-\mathbf{k})) \tag{2.1}$$

which is a rescaling of g to I normalized for $L_2(\mathbb{R}^d)$. We define $\psi^0 := \phi$, $\psi^1 := \psi$. Let V' denote the collection of vertices of the unit cube $[0,1]^d$ and V the nonzero vertices. For each vertex $v = (v_1, \dots, v_d) \in V'$, we define the multivariate functions

$$\psi^v(x_1, \dots, x_d) := \psi^{v_1}(x_1) \cdots \psi^{v_d}(x_d), \quad \tilde{\psi}^v(x_1, \dots, x_d) := \tilde{\psi}^{v_1}(x_1) \cdots \tilde{\psi}^{v_d}(x_d).$$

Let $\mathcal{D}$ denote the set of all dyadic cubes in $I\!R^d$ and $\mathcal{D}_j$ the cubes in $\mathcal{D}$ that satisfy $|I| = 2^{-jd}$, where $|E|$ denotes the Euclidean measure of a set E. The collection of functions

$$\psi_I^v, \quad I \in \mathcal{D}, \ v \in V$$

are a Riesz basis for $L_2(I\!R^d)$ (in the orthogonal case they form a complete orthonormal basis for $L_2(I\!R^d)$). They are an unconditional basis for $L_p(I\!R^d)$, $1 < p < \infty$. Each function f that is locally integrable on $I\!R^d$ has the wavelet expansion

$$f = \sum_{I \in \mathcal{D}} \sum_{v \in V} a_I^v(f) \psi_I^v, \quad a_I^v(f) := \langle f, \tilde{\psi}_I^v \rangle. \tag{2.2}$$

We can start the wavelet decomposition at any dyadic level. For example, starting at dyadic level 0, we obtain

$$f = \sum_{I \in \mathcal{D}_0} \sum_{v \in V'} a_I^v(f) \psi_I^v + \sum_{j=1}^{\infty} \sum_{I \in \mathcal{D}_j} \sum_{v \in V} a_I^v(f) \psi_I^v. \tag{2.3}$$

It is sometimes convenient to choose different normalizations for the wavelets and coefficients appearing in the decompositions (2.2), (2.3). In (2.2), (2.3), we have normalized in $L_2(I\!R^d)$; we can also normalize in $L_p(I\!R^d)$, $0 < p \le \infty$, by taking

$$\psi_{I,p}^v := |I|^{-1/p+1/2} \psi_I^v, \quad I \in \mathcal{D}, \ v \in V'. \tag{2.4}$$

Then, we can rewrite (2.3) as

$$f = \sum_{I \in \mathcal{D}_0} \sum_{v \in V'} a_{I,p}^v(f) \psi_{I,p}^v + \sum_{j=1}^{\infty} \sum_{I \in \mathcal{D}_j} \sum_{v \in V} a_{I,p}^v(f) \psi_{I,p}^v, \tag{2.5}$$

where

$$a_{I,p}^v(f) := \langle f, \tilde{\psi}_{I,p'}^v \rangle$$

with $1/p + 1/p' = 1$.

For simplicity of notation, we shall combine all terms associated to a dyadic cube I in the following

$$A_I(f) := \begin{cases} \sum_{v \in V'} a_{I,p}^v(f) \psi_{I,p}^v, & I \in \mathcal{D}_0, \\ \sum_{v \in V} a_{I,p}^v(f) \psi_{I,p}^v, & I \in \mathcal{D}_j, \ j \ge 1. \end{cases} \tag{2.6}$$

Note that the definition of $A_I(f)$ does not depend on p and that

$$\|A_I(f)\|_{L_p(I\!R^d)} \asymp a_{I,p}(f) := \begin{cases} (\sum_{v \in V'} |a_{I,p}^v(f)|^p)^{1/p}, & I \in \mathcal{D}_0, \\ (\sum_{v \in V} |a_{I,p}^v(f)|^p)^{1/p}, & I \in \mathcal{D}_j, \ j \ge 1. \end{cases} \tag{2.7}$$

Here and later in this paper a statement $A \asymp B$ means that A/B is bounded from above and below by positive constants which do not depend on the variables involved.

It is easy to go from one normalization to another. For example, for any $0 < p, q \le \infty$, we have

$$\psi_{I,p} = |I|^{1/q-1/p}\psi_{I,q}, \quad a_{I,p}(f) = |I|^{1/p-1/q} a_{I,q}(f). \tag{2.8}$$

With this notation, we can rewrite the wavelet decomposition (2.3) as

$$f = \sum_{I\in\mathcal{D}_+} A_I(f) \tag{2.9}$$

with $\mathcal{D}_+ := \cup_{j=0}^{\infty}\mathcal{D}_j$.

In numerical applications of wavelets, we seek efficient approximations to f that use only a small number of terms from its wavelet decomposition. One way of accomplishing this is through what is known as *n-term approximation*. We shall consider in this paper only approximation on the whole of $I\!R^d$. However, results can be obtained for more general domains as described for example in [CDDD]. We first consider the approximation of functions from $L_p := L_p(I\!R^d)$. Let Σ_n be defined as the set of all functions

$$S = \sum_{I\in\Lambda} A_I(S), \quad \#\Lambda \le n, \tag{2.10}$$

where the sets Λ are subsets of $\mathcal{D}_+$. Given $f \in L_p(\Omega)$, $1 \le p \le \infty$, we define

$$\sigma_n(f)_p := \inf_{S\in\Sigma_n} \|f - S\|_{L_p}, \quad n = 0, 1, \ldots. \tag{2.11}$$

Note that by definition $\sigma_0(f)_p := \|f\|_{L_p}$.

One of the main accomplishments in nonlinear approximation has been the characterization of those functions f for which $\sigma_n(f)_p$ has a prescribed asymptotic behavior as $n \to \infty$. For $1 \le p \le \infty$, $0 < q \le \infty$, and $s > 0$, we define the approximation class $\mathcal{A}_q^s(L_p(\Omega))$ to be the set of all $f \in L_p(\Omega)$ such that

$$\|f\|_{\mathcal{A}_q^s(L_p(\Omega))} := \begin{cases} \left(\sum_{n=0}^{\infty}[(n+1)^s\sigma_n(f)_p]^q \frac{1}{n+1}\right)^{1/q}, & 0 < q < \infty \\ \sup_{n\ge 0}(n+1)^s\sigma_n(f)_p, & q = \infty, \end{cases} \tag{2.12}$$

is finite. From the monotonicity of $\sigma_n(f)_p$, it follows that (2.12) is equivalent to

$$\|f\|_{\mathcal{A}_q^s(L_p(\Omega))} \asymp \begin{cases} \left(\sum_{j\ge -1}[2^{js}\sigma_{2^j}(f)_p]^q\right)^{1/q}, & 0 < q < \infty, \\ \sup_{j\ge -1} 2^{js}\sigma_{2^j}(f)_p, & q = \infty, \end{cases} \tag{2.13}$$

where for the purposes of this formula $\sigma_{1/2}(f)_p := \sigma_0(f)_p$. Note that in the special case that $q = \infty$, $\mathcal{A}_\infty^\alpha(L_p(\Omega))$ consists of the functions f for which $\sigma_n(f)_p = O(n^{-\alpha})$.

It is possible to characterize the spaces $A_q^s(L_p(\Omega))$ in several ways: in terms of interpolation spaces; in terms of wavelet coefficients; and in terms of smoothness spaces (Besov spaces). See [CDH] for a discussion of these characterizations. Here, we will just state the characterization in terms of wavelet coefficients that will serve as a comparison for the results we obtain concerning tree approximation.

For this we recall the Lorentz spaces $\ell_{\tau,q}$. Given a sequence $(a_I)_{I\in\mathcal{D}_+}$, let $(\alpha_n^*)_{n\ge 1}$ denote its decreasing rearrangement. That is, α_n^* is the n-th largest of the $|a_I|$. Then, $\ell_{\tau,q}$ consists of all sequences $(a_I)_{I\in\mathcal{D}}$ such that

$$\|(a_I)_{I\in\mathcal{D}}\|_{\ell_{\tau,q}} := \|n^{1/\tau}\alpha_n^*\|_{\ell_q(w)}, \tag{2.14}$$

is finite, where w is the Haar measure on the positive integers ($w(n) := 1/n$, $n \ge 1$).

Theorem 2.1 ([DJP] *and* [CDH]) *Let* $1 < p < \infty$ *and let* $\psi, \tilde{\psi}$ *be a biorthogonal wavelet pair. For each* $\alpha > 0$, *we have that a function* f *is in* $A_q^\alpha(L_p(\Omega))$ *if and only if the sequence* $(a_{I,p}(f))_{I\in\mathcal{D}_+(\Omega)}$ *defined by* (2.7) *is in the Lorentz sequence space* $\ell_{\tau,q}$ *with* $1/\tau := \alpha + 1/p$ *and*

$$\|f\|_{\mathcal{A}_q^\alpha(L_p(\Omega))} \asymp \|f\|_{L_p(\Omega)} + \|(a_{I,p}(f))_{I\in\mathcal{D}_+(\Omega)}\|_{\ell_{\tau,q}}. \tag{2.15}$$

Results similar to (2.15) hold in the case $0 < p \le 1$ by replacing L_p by the Hardy space H_p. Also, when $q = \tau$, the space $\mathcal{A}_\tau^\alpha(L_\tau)$ can be characterized as a Besov space for a certain range of α that depends on the smoothness of ψ and the number of vanishing moments of $\tilde{\psi}$.

As noted earlier, one of the deficiencies in n-term approximation is that the terms appearing in an n-term approximant can occur in any position. The idea in tree approximation is to circumvent this difficulty by requiring that these terms be organized in a tree structure. By a *tree* $\mathcal{T}$ we shall mean a set of dyadic cubes from $\mathcal{D}_+$ with the following property: if $|I| < 1$ and $I \in \mathcal{T}$, then its parent is also in $\mathcal{T}$. The cubes $I \in \mathcal{T}$ with $|I| = 1$ are called the roots of $\mathcal{T}$.

We denote by Σ_n^t the collection of all functions

$$S = \sum_{I\in\mathcal{T}} A_I(S), \quad \#\mathcal{T} \le n, \tag{2.16}$$

with $\mathcal{T}$ a tree. Given a quasi-normed space X, we define the error of tree approximation by

$$t_n(f)_X := \inf_{S\in\Sigma_n^t} \|f - S\|_X. \tag{2.17}$$

The main point of the present paper is to obtain a characterization for tree approximation similar to Theorem 2.1. This will be accomplished by replacing the sequence of wavelet coefficients by a related sequence $(b_I(f))$ obtained by applying a certain maximal function to the sequence $(a_{I,p}(f))$ of wavelet coefficients. We shall obtain such characterizations when the approximation error is measured in L_2 or when it is measured in the Besov space B_p as defined below. This Besov space is very close to (but different from) L_p. There remains the problem of characterizing the approximation classes when the approximation error is measured in L_p, $p \ne 2$.

For $0 < p < \infty$, we define the space $B_p := B_p^0$ to consist of all functions f that are locally integrable on Ω and satisfy

$$\|f\|_{B_p} := \|(a_{I,p}(f))\|_{\ell_p(\mathcal{D}_+)} = \left(\sum_{I\in\mathcal{D}_+} a_{I,p}(f)^p\right)^{1/p} < \infty, \tag{2.18}$$

where $a_{I,p}(f)$, $I \in \mathcal{D}_+$, are the L_p normalized wavelet coefficients defined in (2.8). In the case $p = 2$, the space B_2 is the same as L_2. We note that for $p \ge 1$, $\|\cdot\|_{B_p}$ is a normed space, while for $p < 1$ instead of the triangle inequality we have that for every $f, g \in B_p$,

$$\|f + g\|_{B_p}^p \le \|f\|_{B_p}^p + \|g\|_{B_p}^p. \tag{2.19}$$

We will investigate the tree approximation of a given target function f in the metric of B_p. We will use the abbreviated notation

$$t_n(f)_p := t_n(f)_{B_p} \tag{2.20}$$

We introduce a certain maximal function of the wavelet coefficients of a function f. We denote by $\mathcal{T}_I$ any finite subtree that has I as its root and define

$$\tilde{b}_I := \tilde{b}_{I,p}(f) := \sup_{\mathcal{T}_I} \left(\frac{1}{|\mathcal{T}_I|} \sum_{J \in \mathcal{T}_I} a_{J,p}(f)^p \right)^{\frac{1}{p}}, \tag{2.21}$$

where the supremum is taken over all finite subtrees $\mathcal{T}_I$ with root I (see §3 for the definition of subtrees). We modify $(\tilde{b}_I)$ in order to obtain a decreasing tree sequence:

$$b_I := b_{I,p}(f) := \inf_{J \supseteq I} \tilde{b}_J. \tag{2.22}$$

By the definition, we have $b_{I_1} \geq b_{I_2}$ if $I_1 \supset I_2$.

We can use the new sequences $(b_I(f))_{I \in \mathcal{D}_+}$ to define new function spaces. Given p, α, q, we define $\tau := (\alpha + \frac{1}{p})^{-1}$. The space $X^\alpha_{p,q}$ is by definition the set of all functions $f \in B_p$ such that

$$\|f\|_{X^\alpha_{p,q}} := \|(b_{I,p}(f))_{I \in \mathcal{D}_+}\|_{\ell_{\tau,q}} \tag{2.23}$$

is finite.

In analogy to the approximation spaces $\mathcal{A}^\alpha_q(L_p)$ we define the approximation spaces $\mathcal{A}^\alpha_q(B_p, tree)$ as in (2.12) and (2.13) with $\sigma_n(f)_p$ replaced by $t_n(f)_p$.

Our main result is the following analogue of Theorem 2.1.

Theorem 2.2 *For any $\alpha, q, p > 0$,*

$$\mathcal{A}^\alpha_q(B_p, tree) = X^\alpha_{p,q} \tag{2.24}$$

with equivalent norms.

In the process of proving Theorem 2.2 we will show that a certain thresholding procedure produces best tree approximants. Given $\epsilon > 0$, define $\Lambda_\epsilon := \Lambda_\epsilon(f) := \{I : \ b_I(f) > \epsilon\}$ and $N := N_\epsilon := \#\Lambda_\epsilon$. We define

$$T_N := T_N(f) := \sum_{I \in \Lambda_\epsilon(f)} A_I(f). \tag{2.25}$$

Theorem 2.3 *If $0 < p < \infty$ and $f \in B_p$, then for each $\epsilon > 0$ and $N = N_\epsilon$ defined as above, we have*

$$||f - T_N(f)||_{B_p} = t_N(f)_p. \tag{2.26}$$

In §4, we will also give information about best tree approximants when $N \neq N_\epsilon$.

3 Notation and properties of trees

We introduce some of the notation and terminology for trees that we shall use throughout this paper. By a *subtree* $\mathcal{T}$, we shall mean a collection of cubes from $\mathcal{D}_+$ such that whenever I and $J \subset I$ are in $\mathcal{T}$, then any cube K satisfying $J \subset K \subset I$ is also in $\mathcal{T}$. The cubes I appearing in a subtree are called its *nodes*. The cubes $I \in \mathcal{T}$ that are not contained in any other cubes from $\mathcal{T}$ are called its *roots*.

We will use the notation $\mathcal{T}_I$ to denote a subtree with a single root I.

If $\mathcal{T}$ is a subtree, then we say that the cube $J \in \mathcal{D}_+$ is a *child* of $\mathcal{T}$ if J is not in $\mathcal{T}$ but is the child of some cube from $\mathcal{T}$. We shall frequently make use of the following remark.

Remark 3.1 *Given two subtrees $\mathcal{T}$ and $\mathcal{T}_0 \subset \mathcal{T}$ with common root, we have*

$$\mathcal{T} \setminus \mathcal{T}_0 = \cup_{J \in \Lambda'} \mathcal{T}_J \tag{3.1}$$

where Λ' is a collection of children of $\mathcal{T}_0$.

4 Properties of the maximal sequences

In this section, we will derive properties of the maximal sequence (b_I). We start out by deriving related results for an arbitrary sequence $(c_I)_{I \in \mathcal{D}_+}$ of nonnegative numbers such that $c_I \to 0$, $|I| \to 0$. At the end of this section, we obtain the results we seek for the maximal sequence (b_I) by taking $c_I = a^p_{I,p}$.

We fix a sequence $(c_I)_{I \in \mathcal{D}_+}$ as described above. Given any set Λ of dyadic cubes, we let $\mathrm{Ave}(\Lambda)$ denote the average of the c_I over Λ. For each sequence $(c_I)_{I \in \mathcal{D}_+}$, we define the maximal sequence

$$c^*_I := \sup_{\mathcal{T}_I} \mathrm{Ave}(\mathcal{T}_I) = \sup_{\mathcal{T}_I} \frac{1}{\#\mathcal{T}_I} \sum_{J \in \mathcal{T}_I} c_J, \tag{4.1}$$

where the supremum is taken over all subtrees $\mathcal{T}_I$ with single root I.

We introduce some notation that we will utilize in the remainder of the paper. For any I, we shall denote by $\mathcal{T}^*_I$ the largest finite subtree with root I that satisfies

$$c^*_I = \mathrm{Ave}(\mathcal{T}^*_I). \tag{4.2}$$

The existence of such $\mathcal{T}^*_I$ follows from the fact that the c_J tend to zero as $|J| \to 0$. The following lemma gives the uniqueness of $\mathcal{T}^*_I$

Lemma 4.1 *For each $I \in \mathcal{D}_+$, there is a unique subtree $\mathcal{T}^*_I$ that satisfies* (4.1) *and has maximal cardinality.*

Proof: Suppose that $\mathcal{T}^j_I$, $j = 1, 2$, both satisfy (4.2) and that their common cardinality n is maximal. Let m be the cardinality of the tree $\mathcal{T} := \mathcal{T}^1_I \cap \mathcal{T}^2_I$ having root I. From the definition (4.1), it follows that $\alpha := \mathrm{Ave}(\mathcal{T}) \le c^*_I$. From this it follows that $\beta := \mathrm{Ave}(\mathcal{T}^j_I \setminus \mathcal{T}) \ge c^*_I$, $j = 1, 2$. The tree $\mathcal{T}' := \mathcal{T}^1_I \cup \mathcal{T}^2_I$ has cardinality $2n - m$ and root I. It satisfies $(2n - m)\mathrm{Ave}(\mathcal{T}') = m\alpha + 2(n - m)\beta$. Since $m\alpha + (n - m)\beta = nc^*_I$, we have

$$(2n - m)\mathrm{Ave}(\mathcal{T}') = 2nc^*_I - m\alpha \ge (2n - m)c^*_I.$$

In other words, $\mathrm{Ave}(\mathcal{T}') = c^*_I$, and therefore $2n - m \le n$. That is, $n = m$ and $\mathcal{T}^1_I = \mathcal{T}^2_I$.
□

Given any finite subtree $\mathcal{T}$ and any cube $J \in \mathcal{T}$, we define

$$\Gamma_J := \Gamma_J(\mathcal{T}) := \{K \in \mathcal{T} : K \subset J\}. \tag{4.3}$$

In other words, Γ_J is the largest subtree with root J that is contained in $\mathcal{T}$.

We now derive properties of the subtrees $\mathcal{T}^*_I$ and the sequence (c^*_I).

Lemma 4.2 *If $I \in \mathcal{D}_+$ and $J \in \mathcal{T}_I^*$, $J \neq I$, and $\Gamma_J = \Gamma_J(\mathcal{T}_I^*)$, then*

$$\mathrm{Ave}(\Gamma_J) \geq \mathrm{Ave}(\mathcal{T}_I^*) = c_I^*, \tag{4.4}$$

and

$$\mathrm{Ave}(\Gamma_J) \geq \mathrm{Ave}(\mathcal{T}_I^* \setminus \Gamma_J). \tag{4.5}$$

Proof: If (4.4) is not true, then $\mathrm{Ave}(\mathcal{T}_I^*) > \mathrm{Ave}(\Gamma_J)$, which implies that $\mathrm{Ave}(\mathcal{T}_I^* \setminus \Gamma_J) > \mathrm{Ave}(\mathcal{T}_I^*)$. Since $\mathcal{T}_I^* \setminus \Gamma_J$ is a tree with root I, this contradicts the definition of $\mathcal{T}_I^*$ and c_I^*. If (4.5) is not valid, then from (4.4) we find that $\mathrm{Ave}(\mathcal{T}_I^* \setminus \Gamma_J) > \mathrm{Ave}(\mathcal{T}_I^*)$, which again contradicts the definition of $\mathcal{T}_I^*$. □

The next lemma shows an ordering property for the subtrees $\mathcal{T}_I^*$.

Lemma 4.3 *If $I \in \mathcal{D}_+$ and $J \in \mathcal{T}_I^*$, then $\mathcal{T}_J^* \subset \Gamma_J(\mathcal{T}_I^*)$ and hence $\mathcal{T}_J^* \subset \mathcal{T}_I^*$.*

Proof: Let $\Gamma_J = \Gamma_J(\mathcal{T}_I^*)$ and let $\mathcal{T} := \mathcal{T}_J^* \cap \Gamma_J$, which is a tree with root J. Then,

$$c_I^* \leq \mathrm{Ave}(\Gamma_J) \leq \mathrm{Ave}(\mathcal{T}_J^*) \leq \mathrm{Ave}(\mathcal{T}_J^* \setminus \mathcal{T}), \tag{4.6}$$

where the first inequality is by (4.4), and the second inequality is because $\mathrm{Ave}(\mathcal{T}_J^*) = c_J^*$. For the last inequality note that $\mathcal{T}_J^* \setminus \mathcal{T}$ can be written as a disjoint union of trees $\Gamma_K(\mathcal{T}_J^*)$, and hence we can apply (4.4) for each of these K. Now, (4.6) guarantees that if we form the new tree $\mathcal{T}' := \mathcal{T}_I^* \cup (\mathcal{T}_J^* \setminus \mathcal{T})$ then the average over this tree will be at least as large as c_I^*. Since $(\mathcal{T}_J^* \setminus \mathcal{T}) \cap \mathcal{T}_I^* = \emptyset$, the maximality of $\mathcal{T}_I^*$ gives that $\mathcal{T}_J^* \setminus \mathcal{T} = \emptyset$. Hence, $\mathcal{T}_J^* \subset \Gamma_J$. □

Lemma 4.4 *If $I \in \mathcal{D}_+$ and $J \in \mathcal{T}_I^*$, then $c_J^* \geq c_I^*$.*

Proof: We have

$$c_J^* = \mathrm{Ave}(\mathcal{T}_J^*) \geq \mathrm{Ave}(\Gamma_J) \geq c_I^*,$$

where the last inequality is (4.4). □

From Lemma 4.3, we see that any two of the $\mathcal{T}_I^*$ are either disjoint or one is contained in another. We call $\mathcal{T}_I^*$ a *supernode* if there is no J such that $\mathcal{T}_I^* \subset \mathcal{T}_J^*$ [BJ], [Ba]. Let $\mathcal{D}^*$ denote the set of all $I \in \mathcal{D}_+$ such that $\mathcal{T}_I^*$ is a supernode. Then $\{\mathcal{T}_I^*\}_{I \in \mathcal{D}^*}$ is a partition of $\mathcal{D}_+$.

The set of supernodes has a natural ordering. We say $\mathcal{T}_J^* \leq \mathcal{T}_I^*$ if $J \subset I$. We also say that the supernode $\mathcal{T}_J^*$ is a *child* of the supernode $\mathcal{T}_I^*$ if J is the child of some cube $K \in \mathcal{T}_I^*$.

Lemma 4.5 *If the supernode $\mathcal{T}_J^*$ is a child of the supernode $\mathcal{T}_I^*$, then $c_J^* < c_I^*$.*

Proof: If $c_J^* \geq c_I^*$, then $\mathcal{T}_I^* \cup \mathcal{T}_J^*$ would be a tree with a single root I on which the average of the (c_K) is $\geq c_I^*$. This contradicts the maximality of $\mathcal{T}_I^*$. □

We shall apply the above results to the sequence b_I defined in (2.22). If we fix an $f \in B_p$ and take $c_I := a_{I,p}(f)^p$, then in view of the definitions (2.21) and (2.22), we have $\tilde{b}_I^p = c_I^*$ and

$$b_I^p = \inf_{J \supseteq I} c_J^*. \tag{4.7}$$

It follows from the above that we have the following properties for the sequences $(\widetilde{b}_I)$ and (b_I) [BJ], [Ba].

Property 1: *For each* $I \in \mathcal{D}_+$, $\widetilde{b}_J \geq \widetilde{b}_I$ *whenever* $J \in \mathcal{T}_I^*$.
Indeed, this follows from Lemma 4.4.

Property 2: *The sequence* $(\widetilde{b}_I)_{I\in\mathcal{D}^*}$ *is a decreasing sequence. That is, if* $I, J \in \mathcal{D}^*$, *and* $J \subset I$, $J \neq I$, *then* $\widetilde{b}_J < \widetilde{b}_I$.
This follows from Lemma 4.5

Moreover, from these two properties, we conclude that

Property 3: *For each* $I \in \mathcal{D}^*$ *and* $J \in \mathcal{T}_I^*$, *we have*

$$b_J = b_I = \widetilde{b}_I = \left(\frac{1}{\#\mathcal{T}_I^*}\sum_{K\in\mathcal{T}_I^*} a_K^p\right)^{\frac{1}{p}}.$$

That is, (b_J) *is constant on each supernode.*

Also from the definition (4.7) of the b_I, we have

Property 4: *If* $J \subset I$, *then* $b_J \leq b_I$.

It also follows from Property 3 that

$$\sum_{J\in\mathcal{T}_I^*} a_{J,p}(f)^p = b_I^p \#\mathcal{T}_I^* = \sum_{J\in\mathcal{T}_I^*} b_J^p. \tag{4.8}$$

Since the supernodes $\mathcal{T}_I^*$, $I \in \mathcal{D}^*$ form a partition of $\mathcal{D}$, we have

$$\sum_{J\in\mathcal{D}} a_{J,p}(f)^p = \sum_{J\in\mathcal{D}} b_J^p. \tag{4.9}$$

Finally, we have

Property 5: *If* $\mathcal{T}_I^*$, $I \in \mathcal{D}^*$, *is any supernode, then for any tree* $\mathcal{T}_I \subset \mathcal{T}_I^*$ *with root* I, *we have*

$$\sum_{J\in\mathcal{T}_I} a_{J,p}^p \leq (\#\mathcal{T}_I) b_I^p. \tag{4.10}$$

Indeed, the left side of (4.10) does not exceed $\widetilde{b}_I^p(\#\mathcal{T}_I) = b_I^p(\#\mathcal{T}_I)$, because of the definition of $\widetilde{b}_I$ and Property 3.

5 Best tree approximation

There are certain values of N for which we can find the best tree approximation to a function $f \in B_p$ with N nodes. To describe these situations, we fix $0 < p < \infty$, let $\epsilon > 0$ be any given positive number, and consider the set $\Lambda_\epsilon := \Lambda_\epsilon(f) = \{I : b_I(f) > \epsilon\}$. It follows from Properties 2 and 3 of the last section that Λ_ϵ is a union of supernodes: $\Lambda_\epsilon = \cup_{I\in\mathcal{D}_\epsilon^*}\mathcal{T}_I^*$ with $\mathcal{D}_\epsilon^* \subset \mathcal{D}^*$. From Property 4 it follows that the set Λ_ϵ forms a tree. From the fact that $f \in B_p$, it follows that Λ_ϵ has finite cardinality $N_\epsilon = \#\Lambda_\epsilon$. For such values of $N = N_\epsilon$, we define

$$T_N(f) := \sum_{I\in\Lambda_\epsilon(f)} A_I(f), \tag{5.1}$$

which is a tree approximation to f with N nodes.

Theorem 5.1 *For any $f \in B_p$ and any $\epsilon > 0$, the function $T_N(f)$, $N = N_\epsilon$, is the best tree approximation to f with $\leq N$ nodes, that is,*

$$\|f - T_N(f)\|_{B_p} = t_N(f)_p \tag{5.2}$$

Proof: Let $\mathcal{T}$ be any tree with cardinality N and let $T := \sum_{I\in\mathcal{T}} A_I(f)$. We shall show that T does not approximate f as well as T_N.

In the case where $\Lambda_\epsilon \cap \mathcal{T} = \emptyset$, that is, when $\mathcal{T}$ and Λ_ϵ do not have common roots, then $\mathcal{T}$ can be written as the disjoint union of subtrees $\mathcal{T}_J$, with roots $J \in \mathcal{D}^* \setminus \mathcal{D}^*_\epsilon$, and such that $\mathcal{T}_J \subset \mathcal{T}^*_J$. It follows from Property 5 of the last section that

$$\sum_{K\in\mathcal{T}_J} a_{K,p}(f)^p \leq \tilde{b}^p_J \#(\mathcal{T}_J) = b^p_J \#(\mathcal{T}_J) < \epsilon^p \#(\mathcal{T}_J), \tag{5.3}$$

where the last inequality follows from the fact that J is not in $\mathcal{D}^*_\epsilon$.

On the other hand, since Λ_ϵ is a union of supernodes, for every $J \in \mathcal{D}^*_\epsilon$

$$\sum_{K\in\mathcal{T}^*_J} a_{K,p}(f)^p = \tilde{b}^p_J \#(\mathcal{T}^*_J) = b^p_J \#(\mathcal{T}^*_J) > \epsilon^p \#(\mathcal{T}^*_J). \tag{5.4}$$

From (6.14) and (6.15) and taking into account the fact that $\mathcal{T}$ and Λ_ϵ have the same cardinality, we have that

$$\sum_{K\in\mathcal{T}} a^p_{K,p} < \sum_{K\in\Lambda_\epsilon} a^p_{K,p}. \tag{5.5}$$

This gives

$$\|f - T_N\|^p_p = \|f\|^p_p - \sum_{K\in\Lambda_\epsilon} a^p_{K,p} < \|f\|^p_p - \sum_{K\in\mathcal{T}} a^p_{K,p} = \|f - T\|^p_p, \tag{5.6}$$

as desired.

For the rest of the proof for any tree T we shall denote by $R(T)$ the set of its roots. Let us now assume that $\mathcal{T}$ and Λ_ϵ have common roots, in other words $R(\mathcal{T} \cap \Lambda_\epsilon) = R(\mathcal{T}) \cap R(\Lambda_\epsilon) \neq \emptyset$. If $\mathcal{T}_0 := \mathcal{T} \cap \Lambda_\epsilon$, then we can write the set $\mathcal{T}' := \mathcal{T} \setminus \mathcal{T}_0$ as the union $\mathcal{T}' = \mathcal{T}_1 \cup \mathcal{T}_2$ where $\mathcal{T}_1$ has all of its roots in $R(\mathcal{T}) \setminus R(\mathcal{T}_0)$ and (using Remark 3.1) $\mathcal{T}_2 = \cup_{J\in\Lambda'} \mathcal{T}'_J$ where each $\mathcal{T}'_J$, $J \in \Lambda'$, is a child of Λ_ϵ and hence it is the root of a supernode. As in (6.14, we have

$$\sum_{K\in\mathcal{T}_1} a_{K,p}(f)^p \leq \epsilon^p \#(\mathcal{T}_1). \tag{5.7}$$

Similarly, using Property 5 of the last section, we obtain

$$\sum_{K\in\mathcal{T}_2} = \sum_{J\in\Lambda'} \sum_{K\in\mathcal{T}'_J} a_{K,p}(f)^p \leq \sum_{J\in\Lambda'} \tilde{b}^p_J \#(\mathcal{T}'_J) = \sum_{J\in\Lambda'} b^p_J \#(\mathcal{T}'_J) < \epsilon^p \#(\mathcal{T}_2). \tag{5.8}$$

where the last inequality follows from the fact that each $J \in \Lambda'$ is not in $\mathcal{D}^*_\epsilon$.

On the other hand, consider the set $\mathcal{T}'' := \Lambda_\epsilon \setminus \mathcal{T}_0$. This set can be written as a union of subtrees $\mathcal{T}''_J$, $J \in \Lambda''$ where $\mathcal{T}''_J = \Gamma_J(\mathcal{T}^*_I)$ for some supernode $\mathcal{T}^*_I$, $I \in \mathcal{D}^*_\epsilon$ and some $J \in \mathcal{T}^*_I$. It follows from Lemma 4.2 that

$$\sum_{K\in\mathcal{T}''_J} a_{K,p}(f)^p \geq \epsilon^p \#(\mathcal{T}''_J). \tag{5.9}$$

Now, $\mathcal{T} \setminus \mathcal{T}_0$ and $\Lambda_\epsilon \setminus \mathcal{T}_0$ have the same cardinality and therefore, if these sets are not empty, we will have

$$\sum_{K \in \mathcal{T}} a_{K,p}^p < \sum_{K \in \Lambda_\epsilon} a_{K,p}^p. \tag{5.10}$$

Again, this gives

$$\|f - T_N\|_p^p < \|f - T\|_p^p, \tag{5.11}$$

which completes the proof of the Theorem. □

6 Proof of Theorem 2.2

In this section, we shall prove Theorem 2.2. We continue to use the notation of the last section for the sets Λ_ϵ. We denote by $b_n := b_n(f)$ the n-th largest of the numbers $b_I(f)$, $I \in D_+$. Since $(b_I(f))_{I \in D_+}$ is a decreasing sequence with respect to I, and b_I is constant on each supernode, the reordering runs from one family of supernodes to another, where on each family of supernodes the b_n's have the same value. In other words, we have

$$\underbrace{b_{n_1} = b_1, \cdots,}_{\text{1st family of supernodes}} \quad , \quad \underbrace{b_{n_2}, b_{n_2+1} \cdots}_{\text{2nd family of supernodes}} \quad , b_{n_3}, \cdots,$$

where $b_{n_k} \geq b_{n_{k+1}}$ and $b_i = b_j$ if $n_k \leq i, j < n_{k+1}$, $k = 1, 2, \ldots$

We can identify each each b_n with a node $I \in \mathcal{D}_+$ such that $b_I = b_n$. It may happen that many b_I take the same value b_n. In this case we order the b_I as follows. The first priority is given to the supernode (in its natural order) and then the priority within the supernode goes according to tree structure within that supernode. In this way, to each positive integer N, we have associated a tree $\mathcal{T}_N$ with N nodes and the sets $\{b_I\}_{I \in \mathcal{T}_N}$ and $\{b_n\}_{n=1}^N$ are identical. The tree $\mathcal{T}_{N+1}$ is obtained from $\mathcal{T}_N$ by adding one node (growing $\mathcal{T}_N$). We define

$$T_N := T_N(f) := \sum_{I \in \mathcal{T}_N} A_I(f). \tag{6.1}$$

We begin by deriving the inclusion $X_{p,q}^\alpha \subset \mathcal{A}_q^\alpha(B_p, tree)$.

Theorem 6.1 *Let $0 < p < \infty$, $0 < q \leq \infty$, and $\alpha > 0$. If $f \in X_{p,q}^\alpha$, then $f \in \mathcal{A}_q^\alpha(B_p, tree)$ and*

$$|f|_{\mathcal{A}_q^\alpha(B_p,tree)} \leq C \left(\sum_{n \geq -1} [2^{n\alpha} t_{2^n}(f)_p]^q \right)^{1/q} \leq C \|f\|_{X_{p,q}^\alpha}, \tag{6.2}$$

where $t_{1/2}(f) := \|f\|_p$, $\frac{1}{q} = \alpha + \frac{1}{p}$ and C depends only on α and p.

Proof: We consider only the case $q < \infty$. The case $q = \infty$ requires some trivial modifications. The left inequality follows directly from the definition of the norm on $\mathcal{A}_q^\alpha(B_p, tree)$ and the monotonicity of the sequence (t_n). In order to establish the right inequality in (6.2), it suffices to prove that for each $n \geq -1$, the following weak-type inequality holds

$$t_{2^n}(f)_p \leq C \left(\sum_{j \geq n} (2^{\frac{j}{p}} b_{2^j})^p \right)^{\frac{1}{p}}, \tag{6.3}$$

where for the purposes of this formula we define $b_{1/2} := 0$. Indeed, if (6.3) holds, then from Hardy's inequality (see p. 27 of [DL]), we find

$$\left(\sum_{n=-1}^{\infty}(2^{n\alpha}t_{2^n}(f)_p)^q\right)^{\frac{1}{q}} \leq C\left(\sum_{n=0}^{\infty}(2^{n\alpha}2^{\frac{n}{p}}b_{2^n})^q\right)^{\frac{1}{q}} = C\left(\sum_{n=0}^{\infty}(2^{\frac{n}{\tau}}b_{2^n})^q\right)^{\frac{1}{q}} \leq C\|f\|_{X^{\alpha}_{p,q}}.$$

To prove (6.3), we let $N = 2^n$ and assume that $n_k \leq N < n_{k+1}$. Then

$$\|f - T_N\|_p^p \leq \sum_{I\in\mathcal{T}_{n_{k+1}-1}\setminus\mathcal{T}_N} a_{I,p}(f)^p + \sum_{I\notin\mathcal{T}_{n_{k+1}-1}} a_{I,p}(f)^p =: S_1 + S_2. \tag{6.4}$$

From (4.8), we have

$$S_2 = \sum_{j=n_{k+1}}^{\infty} b_j^p,$$

and therefore, using the monotonicity of the coefficients, we easily obtain

$$S_2 \leq \sum_{j=2^n}^{\infty} b_j^p \leq \sum_{j=n}^{\infty} 2^j b_{2^j}^p. \tag{6.5}$$

To estimate S_1, note that

$$S_1 \leq \sum_{I\in\mathcal{T}_{n_{k+1}-1}\setminus\mathcal{T}_{n_k-1}} a_{I,p}(f)^p = \sum_{j=n_k}^{n_{k+1}-1} b_j^p = (n_{k+1}-1)b_{n_{k+1}-1}^p,$$

where the first equality uses (4.9) and the last inequality uses that b_j is constant in the range of the right sum. The right side does not exceed $2\cdot 2^m b_{2^m}^p$ where $m \geq n$ is the largest integer such that $2^m \leq n_{k+1}-1$. Hence,

$$S_1 \leq C\sum_{j=n}^{\infty} 2^j b_{2^j}^p. \tag{6.6}$$

Putting together (6.4), (6.5) and (6.6), we obtain (6.3), which conludes the proof of the Theorem. □

Next we shall prove the following converse to Theorem 6.1

Theorem 6.2 *Let* $0 < p < \infty$, $0 < q \leq \infty$, *and* $\alpha > 0$. *If* $f \in \mathcal{A}_q^{\alpha}(B_p, tree)$ *then* $f \in X^{\alpha}_{p,q}$ *and*

$$\|f\|_{X^{\alpha}_{p,q}} \leq C\|f\|_{\mathcal{A}_q^{\alpha}(B_p,tree)}, \tag{6.7}$$

with C *depending only on* α *and* p.

Proof: Again we shall treat only the case $0 < q < \infty$. Let $f \in B_p$, and let $(b_n)_{n \geq 1}$ be the rearranged sequence of $(b_I)_{I \in \mathcal{D}_+}$ as introduced earlier. We first claim that for every $k = 1, 2, \ldots$ we have

$$b_{n_{k-1}} \leq C n_k^{-1/p} t_{[\frac{n_k-1}{2}]}(f)_p. \tag{6.8}$$

To prove (6.8), we fix a value of k, define $N := [\frac{n_k-1}{2}]$ and consider the best tree $\mathcal{T}$ of cardinality N. Let $T = \sum_{I \in \mathcal{T}} A_I(f)$ be the corresponding approximant. Consider also the tree

$$\Lambda_{b_{n_k}} = \cup_{I \in \mathcal{D}^*_{b_{n_k}}} \mathcal{T}_I^* \text{ with } \#\Lambda_{b_{n_k}} = n_k - 1.$$

Then, the set $\Lambda_{b_{n_k}} \setminus \mathcal{T}$ has at least N elements, and we can write

$$\Lambda_{b_{n_k}} \setminus \mathcal{T} = \cup_{I \in \mathcal{D}^*_{b_{n_k}}} [\mathcal{T}_I^* \setminus (\mathcal{T} \cap \mathcal{T}_I^*)]. \tag{6.9}$$

Going further, for each $I \in \mathcal{D}^*_{b_{n_k}}$, we can use Remark 3.1 to write

$$\mathcal{T}_I^* \setminus (\mathcal{T} \cap \mathcal{T}_I^*) = \cup_{J \in \Lambda'_I} \mathcal{T}_J,$$

where $\mathcal{T}_J = \Gamma_J(\mathcal{T}_I^*)$ (see (4.3)). Thus, $\mathcal{T}_J$ is a subtree with root $J \in \mathcal{T}_I^*$. It follows from Lemma 4.2 and Property 3 of §4 that for each $J \in \Lambda'_I$, we have

$$\sum_{K \in \mathcal{T}_J} a^p_{K,p} \geq b_I^p \#\mathcal{T}_J \geq b^p_{n_{k-1}} \#\mathcal{T}_J.$$

If we sum these inequalities over all $J \in \Lambda'_I$, and then over all $I \in \mathcal{D}^*_{b_{n_k}}$, we obtain

$$t_N(f)_p = \|f - T\|_p^p \geq \sum_{I \in \mathcal{D}^*_{b_{n_k}}} \sum_{J \in \Lambda'_I} \sum_{K \in \mathcal{T}_J} a^p_{K,p} \geq b^p_{n_{k-1}} \sum_{I \in \mathcal{D}^*_{b_{n_k}}} \sum_{J \in \Lambda'_I} \#\mathcal{T}_J \geq b^p_{n_{k-1}} N. \tag{6.10}$$

This concludes the proof of (6.8).

Using (6.8) it follows immediately that

$$\begin{aligned} \|f\|^q_{X^\alpha_{p,q}} &= \sum_{k=1}^\infty \sum_{n_{k-1} \leq \nu < n_k} \frac{1}{\nu} [\nu^{1/\tau} b_\nu]^q = \sum_{k=1}^\infty b^q_{n_{k-1}} \sum_{n_{k-1} \leq \nu < n_k} \nu^{(q/\tau)-1} \\ &\leq C \sum_{k=1}^\infty n_k^{-q/p} t_{[\frac{n_k-1}{2}]}(f)_p^q \sum_{n_{k-1} \leq \nu < n_k} \nu^{(q/\tau)-1} \\ &\leq C \sum_{k=1}^\infty t_{[\frac{n_k-1}{2}]}(f)_p^q \sum_{n_{k-1} \leq \nu < n_k} \nu^{(q/\tau - q/p)-1} \\ &= C \sum_{k=1}^\infty t_{[\frac{n_k-1}{2}]}(f)_p^q \sum_{n_{k-1} \leq \nu < n_k} \nu^{\alpha q - 1}. \end{aligned} \tag{6.11}$$

On the other hand

$$\begin{aligned} \sum_{n=0}^\infty [(n+1)^\alpha t_n(f)_p]^q \frac{1}{n+1} &\geq C \sum_{k=1}^\infty \sum_{\frac{n_{k-1}}{2} \leq \nu < \frac{n_k}{2}} \frac{1}{\nu} [\nu^\alpha t_\nu(f)_p]^q \\ &\geq C \sum_{k=1}^\infty t_{[\frac{n_k-1}{2}]}(f)_p^q \sum_{n_{k-1} \leq \nu < n_k} \nu^{\alpha q - 1}. \end{aligned} \tag{6.12}$$

Combining (6.11) and (6.12) we obtain the desired result. □

Theorem 6.3 *For any $f \in B_p$ and any $\epsilon > 0$, the function $T_N(f)$, $N = N_\epsilon$, is the best tree approximation to f with $\leq N$ nodes, that is,*

$$\|f - T_N(f)\|_{B_p} = t_N(f)_p \tag{6.13}$$

Proof: Let $\mathcal{T}$ be any tree with cardinality N and let $T := \sum_{I \in \mathcal{T}} A_I(f)$. We shall show that T does not approximate f as well as T_N.

In the case where $\Lambda_\epsilon \cap \mathcal{T} = \emptyset$, that is, when $\mathcal{T}$ and Λ_ϵ do not have common roots, then $\mathcal{T}$ can be written as the disjoint union of subtrees $\mathcal{T}_J$, with roots $J \in \mathcal{D}^* \setminus \mathcal{D}^*_\epsilon$, and such that $\mathcal{T}_J \subset \mathcal{T}^*_J$. It follows from Property 5 of the last section that

$$\sum_{K \in \mathcal{T}_J} a_{K,p}(f)^p \leq \tilde{b}^p_J \#(\mathcal{T}_J) = b^p_J \#(\mathcal{T}_J) < \epsilon^p \#(\mathcal{T}_J), \tag{6.14}$$

where the last inequality follows from the fact that J is not in $\mathcal{D}^*_\epsilon$.

On the other hand, since Λ_ϵ is a union of supernodes, for every $J \in \mathcal{D}^*_\epsilon$

$$\sum_{K \in \mathcal{T}^*_J} a_{K,p}(f)^p = \tilde{b}^p_J \#(\mathcal{T}^*_J) = b^p_J \#(\mathcal{T}^*_J) > \epsilon^p \#(\mathcal{T}^*_J). \tag{6.15}$$

¿From (6.14) and (6.15) and taking into account the fact that $\mathcal{T}$ and Λ_ϵ have the same cardinality, we have that

$$\sum_{K \in \mathcal{T}} a^p_{K,p} < \sum_{K \in \Lambda_\epsilon} a^p_{K,p}. \tag{6.16}$$

This gives

$$\|f - T_N\|^p_p = \|f\|^p_p - \sum_{K \in \Lambda_\epsilon} a^p_{K,p} < \|f\|^p_p - \sum_{K \in \mathcal{T}} a^p_{K,p} = \|f - T\|^p_p, \tag{6.17}$$

as desired.

For the rest of the proof for any tree T we shall denote by $R(T)$ the set of its roots. Let us now assume that $\mathcal{T}$ and Λ_ϵ have common roots, in other words $R(\mathcal{T} \cap \Lambda_\epsilon) = R(\mathcal{T}) \cap R(\Lambda_\epsilon) \neq \emptyset$. If $\mathcal{T}_0 := \mathcal{T} \cap \Lambda_\epsilon$, then we can write the set $\mathcal{T}' := \mathcal{T} \setminus \mathcal{T}_0$ as a union of disjoint trees, $\mathcal{T}_J$ with $J \in R(\mathcal{T}) \setminus R(\mathcal{T}_0)$ the root of $\mathcal{T}_J$ satisfying $|J| = 1$, and (using Remark 3.1) of disjoint subtrees $\mathcal{T}'_J$, $J \in \Lambda'$, where each $J \in \Lambda'$ is a child of Λ_ϵ and hence it is the root of a supernode. Similarly to our analysis in the first part of the proof we have that for every $J \in R(\mathcal{T}) \setminus R(\mathcal{T}_0)$, the tree $\mathcal{T}_J$ can be written as the disjoint union of subtrees with roots in $\mathcal{D}^* \setminus \mathcal{D}^*_\epsilon$ and such that

$$\sum_{K \in \mathcal{T}_J} a_{K,p}(f)^p \leq \tilde{b}^p_J \#(\mathcal{T}_J) = b^p_J \#(\mathcal{T}_J) < \epsilon^p \#(\mathcal{T}_J). \tag{6.18}$$

If $J \in \Lambda'$ from Property 5 of the last section we obtain

$$\sum_{K \in \mathcal{T}'_J} a_{K,p}(f)^p \leq \tilde{b}^p_J \#(\mathcal{T}'_J) = b^p_J \#(\mathcal{T}'_J) < \epsilon^p \#(\mathcal{T}'_J). \tag{6.19}$$

where the last inequality follows from the fact that J is not in $\mathcal{D}^*_\epsilon$.

On the other hand, consider the set $\mathcal{T}'' := \Lambda_\epsilon \setminus \mathcal{T}_0$. This set can be written as a union of subtrees $\mathcal{T}_J''$, $J \in \Lambda''$ where $\mathcal{T}_J'' = \Gamma_J(\mathcal{T}_I^*)$ for some supernode $\mathcal{T}_I^*$, $I \in \mathcal{D}_\epsilon^*$. It follows from Lemma 4.2 that

$$\sum_{K \in \mathcal{T}_J''} a_{K,p}(f)^p \geq \epsilon^p \#(\mathcal{T}_J''). \tag{6.20}$$

Now, $\mathcal{T} \setminus \mathcal{T}_0$ and $\Lambda_\epsilon \setminus \mathcal{T}_0$ have the same cardinality and therefore, if these sets are not empty, we will have

$$\sum_{K \in \mathcal{T}} a_{K,p}^p < \sum_{K \in \Lambda_\epsilon} a_{K,p}^p. \tag{6.21}$$

Again, this gives

$$\|f - T_N\|_p^p < \|f - T\|_p^p, \tag{6.22}$$

which completes the proof of the Theorem. □

References

[BJ] R. Baraniuk and D. Jones, A Signal-Dependent Time-Frequency Representation: Fast Algorithm for Optimal Kernel Design, *IEEE Trans. Signal Processing*, **45**(1), pp. 134–146, 1994.

[Ba] R. Baraniuk, Optimal Tree Approximation using Wavelets, in *Proc. SPIE Tech. Conf. on Wavelet Applications in Signal Processing VII*, **3813**, Denver, July 1999.

[BL] J. Bergh and J. Löfström, *Interpolation Spaces*, Springer, New York, 1976.

[CDDD] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, Tree approximation and optimal encoding, IMI Preprint Series, University of South Carolina, #9 of 1999.

[CDF] A. Cohen, I. Daubechies and J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, *Comm. Pure and Appl. Math*, **45**(1992), pp. 485–560.

[CDGO] A. Cohen, I. Daubechies, O. Guleryuz, and M. Orchard, On the importance of combining wavelet based non-linear approximation in coding strategies, preprint, Princeton University, 1998.

[CDH] A. Cohen, R. DeVore, and R. Hochmuth, Restricted approximation, Constructive Approximation **16**(2000), 85–113.

[Da1] I. Daubechies, Orthonormal bases of compactly supported wavelets *Comm. Pure and Appl. Math* **41**(1988), 909–996.

[Da2] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[DJP] R. DeVore, B. Jawerth, and V. Popov , Compression of wavelet decompositions, *Amer. J. Math.* **114**(1992), 737–785.

[DL] R. DeVore and G. G. Lorentz, *Constructive Approximation*, Springer, New York, 1993.