TECHNICAL RESEARCH REPORT

Multi-Dimensional Wavelet Frames

by T. Kugarajah and Q. Zhang

T.R. 95-31r1

INSTITUTE FOR SYSTEMS RESEARCH

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Multi-dimensional Wavelet Frames

T. Kugarajah and Q. Zhang

Abstract— **Pati & Krishnaprasad [1] first studied the connection between neural networks and wavelet transforms. Zhang & Benveniste [2] gave a different treatment of this connection. However, the problem of constructing multi-dimensional wavelet frames for use in neural networks has not been satisfactorily studied. In this paper, one-dimensional wavelet frame is generalized to the multi-dimensional case by using single-scaling and multi-scaling parameters. The construction of multi-dimensional wavelet frames is also discussed. These results provide more insight on the use of wavelets in neural networks.**

Keywords— **multi-dimensional wavelet, frame, neural network.**

I. Introduction

Wavelet theory is emerging as an important tool in many applications in signal processing and numerical analysis (see e.g. [3], [4]). However, studies on wavelets have often concentrated on one or two dimensional wavelets. The reason is that the implementation of wavelet transform of higher dimensions is of prohibitive cost [5]. Developing *neural networks* within the frame work provided by *wavelet theory* [1], [6], [2], [7], [8], represents one among many possible situations where use of higher dimensional wavelets is required. General multi-dimensional wavelets need to be studied for such purposes. In particular, multi-dimensional wavelet *frames* need more attention. In the Appendix of [2] a construction of multi-dimensional wavelet frame using tensor product wavelets was proposed, by claiming that the tensor product of one dimensional wavelet frames is also a frame. Unfortunately, an error has occurred in the proof of this proposition[1]. Among other things, this paper corrects this error and leads to a better understanding of wavelets as used in neural networks. The wavelet frame theorem given by Daubechies is generalized for this purpose, in both *single-scaling* and *multi-scaling* forms. The construction of multi-dimensional wavelet frames is also discussed.

In the single-scaling case, a single dilation parameter is used in all the dimensions of each wavelet, whereas in multi-scaling, an independent dilation parameter is used in each dimension. Apart from considerations of the complexities of these two scaling methods, the former results in families of wavelets whose scales are equal in all the dimensions, whereas the latter leads to mixed-scale wavelets. Therefore, the choice of scaling method should be in principle dictated by physical considerations.

T.K. is with the Institute for Systems Research and Department of Electrical Engineering, University of Maryland, College Park, MD20742 USA. This research was supported in part by the National Science Foundation's Engineering Research Centers Program: NSFD CDR 8803012.

Q.Z. is with IRISA/INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France.

[1] It was the first author of this paper who first pointed out this error.

When only the complexity for implementation is concerned, let us first consider the case of orthonormal wavelet bases in $L_2(\mathbf{R}^n)$. When single-scaling wavelets are used in separable form, it is known that $2^n - 1$ mother wavelets are required, where n is the dimension of the wavelets [10]. It would be difficult to implement such wavelet bases for $n > 3$. On the other hand, with multi-scaling wavelets, *one* mother wavelet suffices for generating an orthonormal basis of $L_2(\mathbf{R}^n)$. However, its implementation is in general not less complex, since all the wavelets of mixed-scales would be involved.

In the following we deal with wavelet *frames*, for which the situation is slightly different. In particular, we will make some extensions to Daubechies' theorem on sufficient conditions of wavelet frame (see [9] and the section 3.3.2. of [10]). As we will see in section II, it is possible to build single-scaling multi-dimensional wavelet frames by using a single mother wavelet. However, with single-scaling, we cannot generally ensure that the tensor product of one dimensional wavelet frames is a frame. In contrast, when multi-scaling is used, it is easy to construct multi-dimensional wavelet frames from the tensor product of one dimensional wavelet frames, as shown in sections III and IV-B. The drawback of multi-scaling frame is, as for orthonormal wavelet bases, its enormous complexity in implementation with respect to network size.

The following notations will be used :

- $x = (x_1, \ldots, x_n)^T \in \mathbf{R}^n$: the space domain variable,
- $\omega = (\omega_1, \ldots, \omega_n)^T \in \mathbf{R}^n$: the frequency domain variable,
- For integrations, the domains are $\mathbf{R}^n$; for summations, the ranges are $\mathbf{Z}$ or $\mathbf{Z}^n$ according to the dimension of the index. They are omitted in the formulae for the sake of brevity.
- $\overline{f}$ is the complex conjugate of f.
- For any square integrable functions f and g, their inner product $\langle f, g\rangle = \int \mathrm{d}x f(x)\overline{g(x)}$.
- For $f \in L_2(\mathbf{R}^n)$, $\widehat{f}$ is the n-dimensional Fourier transform of f.
- For $f \in L_2(\mathbf{R}^n)$, $\|f\| = \sqrt{\langle f, f\rangle}$.

This paper is organized as follows. In sections II and III Daubechies' frame theorem is generalized to multi-dimensional case, using single-scaling and multi-scaling respectively. In section IV, we discuss some constructions of wavelet frames. Finally, some concluding remarks are given in section V.

II. Single-scaling wavelet frame

In this section we show that single-scaling multi-dimensional wavelet frames can be built by using a single mother wavelet. For this purpose we generalize

Daubechies' theorem on sufficient conditions of wavelet frame [10] to the single-scaling multi-dimensional case.

Theorem 1: Let $\psi \in L_2(\mathbf{R}^n)$. Consider a family of dilated and translated functions of the form

$$\Psi(a,b) = \left\{ \begin{array}{ccc} \psi_{l,k}(x) & = & a^{-\frac{1}{2}nl}\psi(a^{-l}x - bk): \\ & & l \in \mathbf{Z}, k \in \mathbf{Z}^n \end{array} \right\} \quad (1)$$

where $x \in \mathbf{R}^n$, $a, b \in \mathbf{R}$ and $a > 1$. If the following three conditions (2), (3) and (4) are satisfied

$$m(\psi, a) \triangleq \operatorname*{ess\,inf}_{\|\omega\| \in [1,a]} \sum_{l \in Z} |\widehat{\psi}(a^l \omega)|^2 > 0 \quad (2)$$

$$M(\psi, a) \triangleq \operatorname*{ess\,sup}_{\|\omega\| \in [1,a]} \sum_{l \in Z} |\widehat{\psi}(a^l \omega)|^2 < \infty \quad (3)$$

$$\sup_{\eta \in \mathbf{R}^n} \left[(1 + \eta^T \eta)^{n(1+\epsilon)/2} \beta(\eta) \right] = C_\epsilon < \infty \quad (4)$$

for some $\epsilon > 0$, where

$$\beta(\eta) \triangleq \sup_{\|\omega\| \in [1,a]} \sum_{l \in Z} |\widehat{\psi}(a^l \omega)| \cdot |\widehat{\psi}(a^l \omega + \eta)| \quad (5)$$

then there exists $b_0 > 0$ such that $\forall b \in (0, b_0)$, the family $\Psi(a,b)$ in (1) constitutes a frame of $L^2(\mathbf{R}^n)$, in other words, there exist two constants $A > 0$ and $B < +\infty$, such that $\forall f \in L^2(\mathbf{R}^n)$, the following inequalities hold

$$A\|f\|^2 \le \sum_{l,k} |\langle \psi_{l,k}, f \rangle|^2 \le B\|f\|^2$$

where the sum ranges are $l \in \mathbf{Z}$ and $k \in \mathbf{Z}^n$, $\langle \cdot, \cdot \rangle$ denotes the inner product in $L_2(\mathbf{R}^n)$. □

Note that in the family $\Psi(a,b)$ defined by (1), the dilation index l is a scalar, and the scalar dilation parameter a^l is shared by all the dimensions of a wavelet. The proof of this theorem is given in Appendix I.

III. Multi-scaling wavelet frame

We introduce the dilation and translation matrices D_j and T as

$$D_j = diag\,(a^{j_1}, \cdots, a^{j_n}) \quad (6)$$

where

$$j = (j_1, \cdots, j_n)^T \in \mathbf{Z}^n \quad (7)$$

and

$$T = diag\,(b_1, \cdots, b_n)\ . \quad (8)$$

With D_j and T thus defined, separate dilation and translation parameters can be used in wavelet functions. The following theorem is an analog of Theorem 1 in the multi-scaling case.

Theorem 2: Let $\psi \in L_2(\mathbf{R}^n)$. For $a \in \mathbf{R}$, $a > 1$, $b = (b_1, \cdots, b_n) \in \mathbf{R}^n$, and $b_i > 0$, $i = 1, \cdots, n$, consider the family of translated and dilated functions of the form

$$\Psi(a,b) = \{\psi_{j,k}(x) = \det D_j^{\frac{1}{2}} \psi(D_j x - Tk) : j, k \in \mathbf{Z}^n\}.$$

If

$$m(\psi, a) \triangleq \operatorname*{ess\,inf}_{|\omega_i| \in [1,a], i=1,\cdots,n} \sum_{j \in Z^n} |\widehat{\psi}(D_{-j} w)|^2 > 0,$$

$$M(\psi, a) \triangleq \operatorname*{ess\,sup}_{|\omega_i| \in [1,a], i=1,\cdots,n} \sum_{j \in Z^n} |\widehat{\psi}(D_{-j} w)|^2 < \infty$$

and

$$\sup_{\eta \in \mathbf{R}^n} [(1 + \eta^T \eta)^{n(1+\epsilon)/2} \beta(\eta)] = C_\epsilon < \infty$$

for some $\epsilon > 0$, where

$$\beta(\eta) \triangleq \sup_{|\omega_i| \in [1,a], i=1,\cdots,n} \sum_{j \in Z^n} |\widehat{\psi}(D_{-j} w)| . |\widehat{\psi}(D_{-j} w + \eta)|,$$

then there exists[2] $b_0 > 0$ such that $\forall b \in (0, b_0)$, the family defined above constitutes a frame for $L_2(\mathbf{R}^n)$; i.e. , $\exists$ two constants $A > 0$ and $B < \infty$, such that $\forall f \in L_2(\mathbf{R}^n)$, the following inequalities hold

$$A\|f\|^2 \le \sum_{j,k} |\langle \psi_{j,k}, f \rangle|^2 \le B\|f\|^2$$

□

The proof of this theorem is given in Appendix II.

IV. Construction of wavelet frames

We are interested in a methodology that allows us to construct the multi-dimensional wavelet function leading to frames; i.e., the problem is to find a wavelet function that satisfies, together with its dilation and translation parameters, the sufficiency conditions outlined in the above theorems. In this section we first consider the single-scaling case and the tensor product construction of multi-scaling wavelet frames; then we discuss some non tensor product constructions.

Daubechies [10] shows that in 1-D, a single sufficient condition on the decay of ψ as given by

$$|\widehat{\psi}(\omega)| \le C|\omega|^\alpha (1 + |\omega|^2)^{-\frac{\gamma}{2}}$$

where $\omega \in \mathbf{R}$ with constants $C > 0$, $\alpha > 0$ and $\gamma > \alpha + 1$, guarantees the second and third conditions of the theorems as applied to 1-D.

Since in practice this decay condition is rather mild, for construction in the following we can assume that it is satisfied by the 1-D wavelet chosen.

A. Single-scaling wavelet frame

Let ϕ be a symmetric function ($\phi(-t) = \phi(t)$, $t \in \mathbf{R}$) satisfying condition (2) as a 1-D wavelet function, with a suitably chosen value of a. Assume also that ϕ satisfies the decay condition

$$|\widehat{\phi}(\omega)| \le C|\omega|^\alpha (1 + |\omega|^2)^{-\frac{\gamma}{2}}, \quad \omega \in \mathbf{R} \quad (9)$$

[2] abusing notation, we consider element-wise bounds when we refer to vector bounds in this paper.

with constants $C > 0$, $\alpha > 0$ and $\gamma > \alpha + n$. When $n > 1$, this is a stronger condition than that given by Daubechies, but is nevertheless satisfied by many wavelet functions. In particular, any wavelet function that includes exponential decay will satisfy condition (9). Now take

$$\widehat{\psi}(\omega) = \widehat{\phi}(\|\omega\|), \quad \omega \in \mathbf{R}^n$$

where $\|\cdot\|$ denotes the Euclidean norm. Then, the corresponding radial wavelet function $\psi(x)$, $x \in \mathbf{R}^n$, will satisfy Theorem 1, with a suitably chosen value of a. We give the proof in Appendix III.

As an example, consider the 1-D Mexican Hat: $\phi(x) = (1-x^2)e^{-\frac{x^2}{2}}$, and $\widehat{\phi}(\omega) = \omega^2 e^{-\frac{\omega^2}{2}}$, which satisfies the decay condition in (9). Then in n-D, let $\widehat{\psi}(\omega) = \|\omega\|^2 e^{-\frac{\|\omega\|^2}{2}}$, which leads to the n-D wavelet $\psi(x) = \left(n - \|x\|^2\right) e^{-\frac{\|x\|^2}{2}}$. Applying single-scaling dilations and translations to this mother wavelet gives single scaling wavelet frames.

B. *Multi-scaling wavelet frame*

B.1 Tensor product frames

Let $\psi(x)$ be a tensor product of 1-dimensional wavelet functions, i.e.,

$$\psi(x) = \psi_1(x_1) \cdots \psi_n(x_n).$$

Then,

$$\widehat{\psi}(w) = \widehat{\psi}_1(\omega_1) \cdots \widehat{\psi}_n(\omega_n).$$

$\psi_i(x_i), i = 1, \cdots, n$, must satisfy the admissibility condition:

$$\int \frac{|\widehat{\psi}_i(\omega_i)|^2 \mathrm{d}\omega_i}{|w_i|} < \infty.$$

Under mild conditions of decay, this is satisfied if we choose $\psi_i(x_i)$ such that

$$\int \psi_i(x_i) \mathrm{d}x_i = 0.$$

If these 1-dimensional functions can constitute frames, they must satisfy the first two conditions outlined in Theorem 2, as applied to the one dimensional case. These conditions are known to be *necessary* conditions as well in 1-D [10] .

The assumption made on mild decay conditions ensures that the second and third conditions are satisfied, and hence all conditions of Theorem 2 when reduced to 1-D are satisfied.

In the multidimensional case, by using the inequalities in 1-D above, and the fact that the *infimum* and *supremum* can now be taken over the sum in each dimension,we have

$$\begin{aligned} m(\psi, a) &= \operatorname*{ess\,inf}_{|\omega_i| \in [1,a], i=1,\cdots,n} \Big\{ \sum_{j_1} |\widehat{\psi}_1(a^{-j_1}\omega_1)|^2 \cdots \\ &\qquad \sum_{j_n} |\widehat{\psi}_n(a^{-j_n}\omega_n)|^2 \Big\} \\ &> 0 \end{aligned}$$

$$\begin{aligned} M(\psi, a) &= \operatorname*{ess\,sup}_{|\omega_i| \in [1,a], i=1,\cdots,n} \Big\{ \sum_{j_1} |\widehat{\psi}_1(a^{-j_1}\omega_1)|^2 \cdots \\ &\qquad \sum_{j_n} |\widehat{\psi}_n(a^{-j_n}\omega_n)|^2 \Big\} \\ &< \infty \end{aligned}$$

For the third condition, we have the following inequality,

$$\begin{aligned} &\sum_{|k| \neq 0} \left[\beta(2\pi T^{-1}k)\beta(-2\pi T^{-1}k)\right]^{1/2} \\ \leq \quad & \textstyle\sum_{k_1} (1 + (2\pi b_1^{-1} k_1)^2)^{-\frac{(1+\epsilon)}{2}} \cdots \\ & \qquad \textstyle\sum_{k_n} (1 + (2\pi b_n^{-1} k_n)^2)^{-\frac{(1+\epsilon)}{2}} \\ < \quad & \textstyle\sum_{k_1} |2\pi b_1^{-1} k_1|^{-(1+\epsilon)} \cdots \sum_{k_n} |2\pi b_n^{-1} k_n|^{-(1+\epsilon)}. \end{aligned}$$

Since each sum over k_i converges, $i = 1, \ldots, n$, we have that the sum involving β converges. Moreover, as $b_i \to 0$, $i = 1, \cdots, n$, this sum tends to 0. Hence all conditions of Theorem 2 are satisfied.

Therefore the tensor product construction leads to valid frames of wavelets.

B.2 Necessary conditions

To make the results complete, we are interested in obtaining *necessary* conditions as in the 1-D case. In particular, it would be appropriate to check whether the admissibility condition for discrete wavelet frames has the same structure as *continuous wavelets* in the multidimensional case. In the tensor product set up, this follows trivially since the 1-D admissibility conditions lead to

$$\int_{\mathrm{R}^n} \psi(x) \mathrm{d}x = 0.$$

From recent extensions on the bounds for 1-D case [11], [10], the following holds for any frame $\psi_{i_{j_i}, k_i}$ ($i = 1, \cdots, n$ is the dimension index, j_i, k_i are dilation and translation indexes respectively):

$$A_i \leq \frac{2\pi}{b_i} \sum_{j_i} |\widehat{\psi}_i(a^{-j_i} w)|^2 \leq B_i.$$

Considering multiplication of the above inequalities over $i = 1, \cdots, n$, we have

$$\begin{aligned} A &= A_1 \cdots A_n \leq (2\pi)^n \det T^{-1} \sum_j |\widehat{\psi}(D_{-j} w)|^2 \\ &\leq B_1 \cdots B_n = B. \end{aligned}$$

In [11], this bound is obtained for the case of *Riesz bases.* However, the proof relies only on the frame condition, and therefore the above inequality is general in that it holds for arbitrary frames (not necessarily of the tensor product type). Another problem is to construct such arbitrary frames.

B.3 Non-separable frames

The observation that all conditions of the theorem on sufficient conditions hinge on the boundedness and decay of terms involving $|\widehat{\psi}(\cdot)|$ suggests the possibility of multiplying the tensor product wavelet in the frequency domain by a function of the form

$$p(w) = \sum_{l \in Z^n} c_l \, e^{-il^T w}, \; c_l \in \mathbf{R},$$

which can be the Fourier series of a periodic function. Let the new wavelet be

$$\widehat{\psi}_p(w) = p(w)\widehat{\psi}(w)$$

where $\psi(\cdot)$ corresponds to the wavelet constructed as a tensor product. If

$$0 < \sum_l |c_l| < \infty,$$

then the fact that

$$0 < |\widehat{\psi}_p(w)| \leq (\sum_l |c_l|)|\widehat{\psi}(w)|$$

implies that all conditions of Theorem 2 are satisfied.

Therefore, one can construct non-tensor product wavelet frames from the tensor product frames. In the case of *Riesz bases,* similar results are obtained in [11].

The 1-D wavelet functions could be any of the following: the Laplacian of the Gaussian, a combination of a few sigmoids (e.g.[6]). The choice of the wavelet used in networks for *learning* is dictated by considerations of smoothness, implementability in analog hardware, separability, etc.

V. Concluding remarks

In this paper we studied two different generalizations of one-dimensional wavelet frames and some construction procedures for frames. Theorem 1 suggests a radial construction since the condition $\|\omega\| \in [1, a]$ considers a hypersphere in the frequency (ω) space, whereas Theorem 2 trivially suggests a tensor product construction since we consider $|\omega_i|, i = 1, \cdots, n$. It should be noted that considering separate values of b in Theorem 1 or considering separate values of a in each dimension in Theorem 2 does not alter the results.

It is difficult to implement very high dimensional wavelet bases or frames by following conventional mathematical methods. However, using general multi-dimensional wavelets in neural networks is practically feasible, since only incomplete sets of wavelet frames need to be implemented for the purpose of learning (function approximation) [8]. Although the multi-scaling wavelet frames are more difficult to implement than the single-scaling ones, there are situations where one deals with vastly disparate input variables, and having a shared scaling variable does not give much flexibility. In recent years, researchers have investigated methods that combine good heuristics with rigorous mathematics (e.g., MARS[12]) to combat the exponential increase in complexity. One such method [13] exploits separability of the 'basis' function with the well-known LMS algorithm under certain assumptions. Other methods are being investigated by the authors. Moreover parallel implementation may benefit from separability. These techniques should be developed further so that the flexibility afforded by multi-scaling (and separability) can effectively offset the increase in network complexity. Rigorous results considering these trade-offs, i.e., when exactly one construction leads to better results than the other, need to be obtained.

Acknowledgements

T.K wishes to thank Prof. P.S. Krishnaprasad and Prof. W.P. Dayawansa for useful discussions.

Appendix

I. Proof of Theorem 1

As our proof closely follows Daubechies' scheme for 1-D (see [9] and the section 3.3.2. of [10]), we just present mathematical results that show the validity of the conditions given to many dimensions.

First, we need the following generalization of the Poisson formula:

$$\sum_{k \in Z^n} e^{iCk^T x} = \left(\frac{2\pi}{C}\right)^n \prod_{j=1}^{n} \sum_{k_j \in Z} \delta\left(x_j - \frac{2\pi}{C} k_j\right)$$

where i is the imaginary unity and C is any real non zero constant. It can be verified by simple computations.

By applying this generalized Poisson formula and the Parseval's theorem, straight forward computations give

$$\sum_{l,k} |\langle \psi_{l,k}, f \rangle|^2 = \left(\frac{2\pi}{b}\right)^n \sum_l \int d\omega \, |\widehat{\psi}(a^l \omega)|^2 \cdot |\widehat{f}(\omega)|^2 + \Lambda$$

where

$$\Lambda = \left(\frac{2\pi}{b}\right)^n \sum_l \sum_{k \neq 0} \int d\omega \, \widehat{\psi}(a^l \omega) \overline{\widehat{\psi}\left(a^l\omega - \frac{2\pi}{b}k\right) \widehat{f}(\omega)} \, \widehat{f}\left(\omega - \frac{2\pi}{a^l b}k\right).$$

where $k \neq 0$ means *at least one component* of k is not zero.

By applying the Cauchy-Schwarz inequality, we get

$$|\Lambda| \leq \left(\frac{2\pi}{b}\right)^n \sum_{k \neq 0} \left[\beta\left(\frac{2\pi}{b}k\right) \beta\left(-\frac{2\pi}{b}k\right)\right]^{\frac{1}{2}} \|f\|^2$$

where $\beta(\cdot)$ is as defined in (5). This inequality together with the three conditions of the theorem gives

$$\left(\frac{2\pi}{b}\right)^n \left\{m(\psi,a) - \sum_{k\neq 0}\left[\beta\left(\frac{2\pi}{b}k\right)\beta\left(-\frac{2\pi}{b}k\right)\right]^{\frac{1}{2}}\right\}\|f\|^2$$
$$\leq \sum_{l,k}|\langle\psi_{l,k},f\rangle|^2 \leq \left(\frac{2\pi}{b}\right)^n\left\{M(\psi,a) + \right.$$
$$\left.\sum_{k\neq 0}\left[\beta\left(\frac{2\pi}{b}k\right)\beta\left(-\frac{2\pi}{b}k\right)\right]^{\frac{1}{2}}\right\}\|f\|^2.$$

The only thing left is to verify that condition (4) ensures the convergence of the multi-indexed series

$$\sum_{k\neq 0}\left[\beta\left(\frac{2\pi}{b}k\right)\beta\left(-\frac{2\pi}{b}k\right)\right]^{\frac{1}{2}}$$

and implies that the sum tends to zero when $b \to 0$, so that the coefficients of $\|f\|^2$ in the above inequalities are strictly positive for small enough b.

By (4) we have

$$\beta(\eta) \leq C_\epsilon\left(1+\eta^T\eta\right)^{-\frac{n(1+\epsilon)}{2}}.$$

This leads to

$$\sum_{k\neq 0}\left[\beta\left(\frac{2\pi}{b}k\right)\beta\left(-\frac{2\pi}{b}k\right)\right]^{\frac{1}{2}} \leq$$
$$C_\epsilon\left(\frac{b}{2\pi}\right)^{n(1+\epsilon)}\sum_{k\neq 0}\left[\left(\left(\frac{b}{2\pi}\right)^2\right.\right.$$
$$\left.\left.+|k_1|^2+\cdots+|k_n|^2\right)^n\right]^{-\frac{1+\epsilon}{2}}.$$

Considering the inequality

$$\begin{aligned}(C+|k_1|^2+\cdots+|k_n|^2)^n &\geq (C+|k_1|^2)\\ &\quad\cdots(C+|k_n|^2)\end{aligned}$$

where C is any positive constant, we can see that the series in k converges. Moreover, as $b \to 0$, this sum tends to zero. The proof of the theorem is thus established. □

II. Proof of Theorem 2

As in the proof of theorem 1, we use the n-dimensional Poisson formula, and the Parseval's Theorem in relation to Fourier Transforms. The steps involved are similar to the proof of Theorem 1 as shown below.

We arrive at

$$\begin{aligned}\sum_{j,k}|\langle\psi_{j,k},f\rangle|^2 &= (2\pi\ \det T^{-1})^n\sum_j\\ &\quad\int d\omega|\widehat{\psi}(D_{-j}\omega)|^2\ |\widehat{f}(\omega)|^2+\Lambda\end{aligned}$$

where

$$\begin{aligned}\Lambda &= (2\pi\ \det T^{-1})^n\sum_j\sum_{|k|\neq 0}\int d\omega\widehat{\psi}(D_{-j}\omega)\cdots\\ &\quad\overline{\widehat{\psi}(D_{-j}w-2\pi T^{-1}k)\widehat{f}(\omega)}\widehat{f}(\omega-2\pi D_{-j}T^{-1}k)\end{aligned}$$

The third condition of the theorem implies the decay of β as

$$\beta(\eta) \leq (1+\eta^T\eta)^{-\frac{n(1+\epsilon)}{2}}\cdot C_\epsilon.$$

Hence

$$\sum_{|k|\neq 0}\left[\beta(2\pi T^{-1}k)\beta(-2\pi T^{-1}k)\right]^{\frac{1}{2}} <$$
$$C_\epsilon\left(\frac{\det T}{2\pi}\right)^{n(1+\epsilon)}\sum_{|k|\neq 0}\left[\left(\left(\frac{\det T}{2\pi}\right)^2+k^Tk\right)^n\right]^{-\frac{1+\epsilon}{2}}.$$

The multi-indexed series converges as in theorem 1. Moreover, it is easily seen that when $b_i, i=1,\cdots,n \to 0$, the sum $\to 0$; and the limit on b is given by
$b_c =$

$$\inf\left\{b\middle| m(\psi,a) \leq \sum_{|k|\neq 0}\left[\beta(2\pi T^{-1}k)\beta(-2\pi T^{-1}k)\right]^{\frac{1}{2}}\right\}.$$

Again, the bounds and inequalities are considered element-wise.

This completes the proof of the theorem. □

III. Proof of single-scaling construction

We note that the first two conditions of Theorem 1 are trivially satisfied by the construction $\widehat{\psi}(\omega) = \widehat{\phi}(\|\omega\|)$. For the third condition, we examine

$$\beta(\eta) \triangleq \sup_{\|\omega\|\in[1,a]}\sum_{l\in Z}|\widehat{\psi}(a^l\omega)|\cdot|\widehat{\psi}(a^l\omega+\eta)|. \quad (10)$$

Using the above construction for ψ and condition (9) we derive the following. For comparisons with the one dimensional case, the reader can refer to [10, page 102–103].

$$\begin{aligned}\beta(\eta) &= \sup_{\|\omega\|\in[1,a]}\sum_{l\in Z}|\widehat{\phi}(\|a^l\omega\|)|\cdot|\widehat{\phi}(\|a^l\omega+\eta\|)|\\ &\leq C^2\sup_{\|\omega\|\in[1,a]}\left\{a^\alpha\sum_{l=-\infty}^{-1}a^{l\alpha}(1+\right.\\ &\qquad\qquad\|a^l\omega+\eta\|^2)^{-(\gamma-\alpha)/2}\\ &\quad+\sum_{l=0}^{\infty}[(1+\|a^l\omega\|^2)\\ &\quad\left.\cdot(1+\|a^l\omega+\eta\|^2)]^{-(\gamma-\alpha)/2}\right\}.\end{aligned}$$

For the first term notice that for $-\infty \leq l \leq -1$ and $\|\omega\| \in [1,a]$, $\|a^l\omega\| \in [0,1]$. Hence if $\|\eta\| \geq 2$, then $\|\eta + a^l\omega\| \geq \frac{1}{2}\|\eta\|$, implying $\left(1+\|a^l\omega+\eta\|^2\right)^{-1} \leq 4\left(1+\|\eta\|^2\right)^{-1}$. If $\|\eta\| \leq 2$, we have $5\left(1+\|a^l\omega+\eta\|^2\right) \geq 5 \geq 1+\|\eta\|^2$. Therefore the first term can be bounded by

$$C_1\left(1+\|\eta\|^2\right)^{-(\gamma-\alpha)/2}\sum_{l=-\infty}^{-1}a^{l\alpha}$$

with some constant $C_1 > 0$.

For the second term, let us consider the fact that for all $x, y \in \mathbf{R}^n$, the following inequality always holds (to check this, just develop the left side):

$$(1+\|x+y\|^2)(1+\|x-y\|^2) \geq 1+\|y\|^2.$$

Let $x = a^l\omega + \frac{\eta}{2}$ and $y = \frac{\eta}{2}$. Then, for all $a^l\omega$ and η, the following holds:

$$\left[(1+\|a^l\omega\|^2)(1+\|a^l\omega+\eta\|^2)\right]^{-1} \leq 4(1+\|\eta\|^2)^{-1}.$$

Then, $\forall\delta \in (0,1)$, for the second term we have

$$\begin{aligned}
&\sum_{l=0}^{\infty}[(1+\|a^l\omega\|^2)(1+\|a^l\omega+\eta\|^2)]^{-(\gamma-\alpha)/2} \\
&= \sum_{l=0}^{\infty}\Big\{[(1+\|a^l\omega\|^2)(1+ \\
&\qquad \|a^l\omega+\eta\|^2)]^{-(1-\delta)(\gamma-\alpha)/2} \\
&\qquad \cdot [(1+\|a^l\omega\|^2)(1+\|a^l\omega+\eta\|^2)]^{-\delta(\gamma-\alpha)/2}\Big\} \\
&\leq C_2(1+\|\eta\|^2)^{-(1-\delta)(\gamma-\alpha)/2} \\
&\qquad \cdot \sum_{l=0}^{\infty}[(1+\|a^l\omega\|^2)(1+\|a^l\omega+\eta\|^2)]^{-\delta(\gamma-\alpha)/2} \\
&\leq C_2(1+\|\eta\|^2)^{-(1-\delta)(\gamma-\alpha)/2} \\
&\qquad \cdot \sum_{l=0}^{\infty}(1+a^l\|\omega\|^2)^{-\delta(\gamma-\alpha)/2}
\end{aligned}$$

where $C_2 > 0$ is a constant. Notice that $\|\omega\| \in [1, a]$, the second term

$$\begin{aligned}
&\sum_{l=0}^{\infty}[(1+\|a^l\omega\|^2)(1+\|a^l\omega+\eta\|^2)]^{-(\gamma-\alpha)/2} \leq \\
&\qquad C_2(1+\|\eta\|^2)^{-(1-\delta)(\gamma-\alpha)/2}\sum_{l=0}^{\infty}\left(a^l\right)^{-\delta(\gamma-\alpha)/2}.
\end{aligned}$$

Since $\gamma > \alpha + n$, there exist $\delta > 0$ and $\epsilon > 0$, both small enough so that $(1-\delta)(\gamma-\alpha) \geq n(1+\epsilon)$. Therefore the second term can be bounded by

$$C_2\left(1+\|\eta\|^2\right)^{-n(1+\epsilon)/2}\sum_{l=0}^{\infty}\left(a^{-\delta(\gamma-\alpha)/2}\right)^l.$$

Thus the third condition of Theorem 1 is satisfied by $\beta(\eta)$.

References

[1] Y.C. Pati and P.S. Krishnaprasad, "Discrete affine wavelet transforms for analysis and synthesis feedforward neural networks", in *Advances in Neural Information Processing Systems III*, R. Lippman, J. Moody, and D. Touretzky, Eds., San Mateo, CA, 1990, pp. 743–749, Morgan Kaufmann, Publishers.

[2] Q. Zhang and A. Benveniste, "Wavelet networks", *IEEE Trans. on Neural Networks*, vol. 3, no. 6, pp. 889–898, Nov. 1992.

[3] C.K. Chui, *Wavelets : a tutorial in theory and applications*, Academic Press, Inc. Boston, San Diego, 1992.

[4] M.B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, Eds., *Wavelets and their applications*, Jones and Bartlett books in mathematics, Boston, 1992.

[5] A. Benveniste, A. Juditsky, B. Delyon, Q. Zhang, and P.Y. Glorennec, "Wavelets in identification", in *SYSID'94, 10th IFAC Symposium on System Identification*, Copenhagen, 1994.

[6] Y.C. Pati and P.S. Krishnaprasad, "Analysis and synthesis of feedforward neural networks using discrete affine wavelet transforms", *IEEE Trans. on Neural Networks*, vol. 4, no. 1, pp. 73–85, Jan. 1993.

[7] Q. Zhang, "Regressor selection and wavelet network construction", Technical Report 709, Inria, Apr. 1993.

[8] Q. Zhang, "Using wavelet network in nonparametric estimation", Tech. Rep. 833, IRISA, 1994.

[9] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis", *IEEE Trans. on Information Theory*, vol. 36, no. 5, pp. 961–1005, Sept. 1990.

[10] I. Daubechies, *Ten lectures on wavelets*, CBMS-NSF regional series in applied mathematics. SIAM, 1992.

[11] C.K. Chui and X.L. Shi, "Multi-frequency wavelet decomposition", in *Recent Advances in Wavelet Analysis*, L. Schumaker and G. Webb, Eds., pp. 155 – 189. Academic Press, 1994.

[12] J.H. Friedman, "Multivariate adaptive regression splines", *The Annals of Statistics*, vol. 19, no. 1, pp. 1–141, 1991.

[13] T.D. Sanger, "A tree-structured adaptive network for function approximation in high-dimensional spaces", *IEEE Trans. on Neural Networks*, vol. 2, no. 2, pp. 285–293, March 1991.