NONLINEAR APPROXIMATION WITH REDUNDANT DICTIONARIES

L. Borup, M. Nielsen

Aalborg University, Dept. of Math.
Fredrik Bajers Vej 7G
DK-9220 Aalborg East, Denmark
Email:{*lasse,mnielsen*}*@math.aau.dk*

R. Gribonval

IRISA
Campus de Beaulieu
F-35760 Rennes Cedex, France
Email: *remi.gribonval@irisa.fr*

ABSTRACT

In this paper we study nonlinear approximation and data representation with redundant function dictionaries. In particular, approximation with redundant wavelet bi-frame systems is studied in detail. Several results for orthonormal wavelets are generalized to the redundant case. In general, for a wavelet bi-frame system the approximation properties are limited by the number of vanishing moments of the system. In some cases this can be overcome by oversampling, but at a price of replacing the canonical expansion by another linear expansion. Moreover, for special non-oversampled wavelet bi-frames we can obtain good approximation properties not restricted by the number of vanishing moments, but again without using the canonical expansion.

1. INTRODUCTION

Advances in communication, sensing, and computational power have led to an explosion of data. The size and varied formats for these datasets challenge existing techniques for transmission, storage, querying, display, and numerical manipulation. It is thus vital to develop efficient and robust data representations that lend themselves to scientific analysis and computation. An important theoretical tool along this line is the theory of constructive approximation.

Say we have an expansion of a target function f in an ordered basis, and a positive integer m. By m-term *nonlinear approximation* we mean the process of approximating f by truncating the expansion using only the m "most important" terms. Important terms could be those minimizing a given cost function.

It is interesting to observe how the approximant changes as m increases. In constructive approximation the target functions are divided into a family of *approximation spaces* classified by how fast the approximant converges to f as m increases. The study of these spaces is important for practical purposes since they offer some information on how

much we can expect to be able to compress a given function without substantial loss of information.

For nice systems such as trigonometric- and wavelet bases, there already exist characterizations of the corresponding approximation spaces [1]. However, if these systems are replaced by more general dictionaries, the picture becomes less clear. Examples of structured redundant dictionaries are wavelet bi-frame systems and Gabor frames. The first system has the same basic structure as a wavelet system, while the latter consists of local trigonometric functions.

Many open problems remain regarding function representation using redundant dictionaries. But there is no doubt that sparse representations using such systems will have numerous applications, in fact, there is a growing interest within signal and image processing communities in this approach.

In this note we study the approximation properties of wavelet bi-frame systems in details. The results extends the known results for wavelets to the redundant case.

2. APPROXIMATION SPACES

In this section we introduce the theoretical tools to study nonlinear approximation with redundant function dictionaries. A dictionary $\mathcal{D} = \{g_k\}_{k \in \mathbb{N}}$ in L_p is a countable collection of normalized elements from L_p. For $\mathcal{D}$ we consider the collection of all possible m-term expansions with elements from $\mathcal{D}$:

$$\Sigma_m(\mathcal{D}) := \Big\{ \sum_{i \in \Lambda} c_i g_i \ \Big|\ c_i \in \mathbb{C}, \operatorname{card} \Lambda \le m \Big\}.$$

The error of the best m-term approximation to an element $f \in L_p$ is then

$$\sigma_m(f, \mathcal{D})_p := \inf_{f_m \in \Sigma_m(\mathcal{D})} \|f - f_m\|_{L_p}.$$

The approximation space $\mathcal{A}_q^\alpha(L_p, \mathcal{D})$ essentially consists of all signals that can be approximated at the rate $O(m^{-\alpha})$. The parameter q is for fine-tuning only.

Definition 2.1 (Approximation spaces) *The approximation space $\mathcal{A}_q^\alpha(L_p, \mathcal{D})$ is defined by*

$$|f|_{\mathcal{A}_q^\alpha(L_p,\mathcal{D})}^q := \sum_{m \in \mathbb{N}} \left(m^\alpha \sigma_m(f, \mathcal{D})_p\right)^q \frac{1}{m} < \infty,$$

and (quasi)normed by $\|f\|_{\mathcal{A}_q^\alpha(L_p,\mathcal{D})} = \|f\|_p + |f|_{\mathcal{A}_q^\alpha(L_p,\mathcal{D})}$, for $0 < q, \alpha < \infty$. When $q = \infty$, the ℓ_q norm is replaced by the sup-norm.

In what follows, we use the notation $V \hookrightarrow W$ to indicate V is continuously embedded in W for two (quasi)normed spaces V and W, i.e., $V \subset W$ and there is a constant $C < \infty$ such that $\| \cdot \|_W \leq C \| \cdot \|_V$.

It is known that the main tool in the characterization of $\mathcal{A}_q^\alpha(L_p, \mathcal{D})$ comes from the link between approximation theory and interpolation theory (see *e.g.* [2, Theorem 9.1, Chapter 7]). Let Y_p be a (quasi)Banach space continuously embedded in L_p with semi-(quasi)norm $|\cdot|_{Y_p}$. Given $\alpha > 0$, the Jackson inequality

$$\sigma_m(f, \mathcal{D})_p \leq C m^{-\alpha} |f|_{Y_p}, \qquad \forall f \in Y_p, \ \forall m \in \mathbb{N} \quad (1)$$

and the Bernstein inequality

$$|S|_{Y_p} \leq C' m^\alpha \|S\|_p, \qquad \forall S \in \Sigma_m(\mathcal{D}) \quad (2)$$

(with some constants C and C' independent of f, S and m) imply, respectively, the continuous embedding

$$(L_p, Y_p)_{\beta/\alpha, q} \hookrightarrow \mathcal{A}_q^\beta(L_p, \mathcal{D})$$

and the converse embedding

$$(L_p, Y_p)_{\beta/\alpha, q} \hookleftarrow \mathcal{A}_q^\beta(L_p, \mathcal{D})$$

for all $0 < \beta < \alpha$ and $q \in (0, \infty]$. Here $(X, Y)_{\theta, q}$ denotes the real interpolation space between the Banach spaces X and Y. We refer the reader to [3] for the definition of the real method of interpolation.

3. WAVELET BI-FRAME SYSTEMS

A dictionary of particular interest is a wavelet bi-frame. The redundant representation offered by wavelet bi-frames has already been put to good use for signal denoising, and is currently explored for image compression. Motivated by these and other applications, we would like to investigate the approximation properties of wavelet bi-frames. In particular, we are interested in wavelet bi-frames based on a multiresolution analysis, since this guarantees the existence of fast implementation algorithms.

Given a finite collection of functions $\Psi = \{\psi^\ell\}_{\ell \in E}$, $E := \{1, 2, \ldots, L\}$, in L_2, we use the notation $X(\Psi)$ to denote the corresponding "wavelet" system,

$$X(\Psi) := \{\psi_{j,k}^\ell \mid j, k \in \mathbb{Z}, \ell \in E\},$$

where $\psi_{j,k} := 2^{j/2} \psi(2^j \cdot - k)$. A **wavelet bi-frame for** L_2 consists of two sequences of wavelets $\Psi = \{\psi^\ell\}_{\ell \in E}$ and $\tilde{\Psi} = \{\tilde{\psi}^\ell\}_{\ell \in E}$ for which the systems $X(\Psi)$ and $X(\tilde{\Psi})$ are Bessel systems, and satisfy the reconstruction formula

$$f = \sum_{\ell \in E} \sum_{j,k \in \mathbb{Z}} \langle f, \tilde{\psi}_{j,k}^\ell \rangle \psi_{j,k}^\ell, \qquad \forall f \in L_2. \quad (3)$$

This definition implies that both $X(\Psi)$ and $X(\tilde{\Psi})$ are frames for L_2 and in fact the roles of Ψ and $\tilde{\Psi}$ are interchangeable in (3). The special case with $\Psi = \tilde{\Psi}$ corresponds to a so-called tight wavelet frame.

The most common method for constructing wavelet bi-frames relies on so-called extension principles. The resulting bi-frames are based on a multiresolution analysis (MRA), and the generators are called **bi-framelets**. When the wavelet bi-frame is based on an MRA there exist $\phi, \tilde{\phi} \in L_2$ and 2π-periodic functions $\tau_\ell, \tilde{\tau}_\ell, \ell = 0, 1, \ldots, L$, such that

$$\hat{\phi}(2\xi) = \tau_0(\xi)\hat{\phi}(\xi), \qquad \qquad \hat{\tilde{\phi}}(2\xi) = \tilde{\tau}_0(\xi)\hat{\tilde{\phi}}(\xi),$$

$$\hat{\psi}^\ell(2\xi) = \tau_\ell(\xi)\hat{\phi}(\xi), \quad \text{and} \quad \hat{\tilde{\psi}}^\ell(2\xi) = \tilde{\tau}_\ell(\xi)\hat{\tilde{\phi}}(\xi).$$

The construction of multiresolution-based wavelet frames have been studied extensively, see *e.g.* [4, 5, 6].

4. JACKSON AND BERNSTEIN INEQUALITIES FOR WAVELET BI-FRAMES

In this section we consider Jackson and Bernstein inequalities for wavelet bi-frames. Denote by $W^s(L_\infty)$ the Sobolev space consisting of functions with all s distributional derivatives in L_∞.

Proposition 4.1 ([7]) *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a bi-framelet system and assume that $X(\Psi)$ is based on a compactly supported refinable function $\phi \in W^s(L_\infty)$, $s \geq 0$. Then the Bernstein inequality*

$$|S|_{B_\tau^{d\alpha}(L_\tau)} \leq C m^\alpha \|S\|_{L_p}, \quad \forall S \in \Sigma_m(X(\Psi)), \quad (4)$$

holds true for each $m \geq 1$, $0 < \alpha < s$, $0 < p \leq \infty$, with $1/\tau := \alpha + 1/p$ and $C = C(\alpha, p)$.

A function f is said to have $N \in \mathbb{N}$ vanishing moments if $\int_{\mathbb{R}} x^k f(x) dx = 0$ for $k = 0, 1, \ldots, N - 1$.

The following Jackson inequality is valid.

Proposition 4.2 ([7]) *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a bi-frame. Suppose each $\tilde{\psi} \in \tilde{\Psi}$ has $N \in \mathbb{N}$ vanishing moments and $|\tilde{\psi}(x)| \leq C(1+|x|)^{-2(N+1)}$, and suppose there exist $\beta, \varepsilon > 0$ such that for all $\psi \in \Psi \cup \tilde{\Psi}$, $\psi \in C^\beta$ and $|\psi(x)| \leq C(1 + |x|)^{-1-\varepsilon}$. Then, we have the Jackson inequality*

$$\sigma_m(f, X(\Psi))_p \leq C m^{-\alpha} \|f\|_{B_\tau^{d\alpha}(L_\tau)}$$

for $p \in (1, \infty)$, $\alpha < N$, and $1/\tau = \alpha + 1/p$.

Remark 4.3 *It was shown in [7] that the rate of approximation given by Proposition 4.2 can be obtained simply by thresholding the canonical frame expansion (for sufficiently "nice" bi-frame systems).*

Finally, combining Proposition 4.1 and Proposition 4.2 above, we obtain a full characterization of the approximation spaces using wavelet bi-frames.

Theorem 4.4 *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a wavelet bi-frame system that satisfies the hypothesis of Propositions 4.1 and 4.2. Let $p \in (1, \infty)$ and $\tau := (\alpha + 1/p)^{-1}$ where we assume $0 < \alpha < \min\{s, N\}$. Then, for each $0 < \beta < \alpha$, $q \in (0, \infty]$, we have the characterization*

$$\mathcal{A}_q^\beta(L_p, X(\Psi)) = (L_p, B_\tau^\alpha(L_\tau))_{\beta/\alpha, q}. \quad (5)$$

Remark 4.5 *It is known (see, e.g., [1]) that the right hand side of (5) equals the Besov space $B_q^\beta(L_q)$ when $1/q = \beta + 1/p$, $0 < \beta < \alpha$, and $q \in (0, \infty]$.*

5. BEYOND THE CANONICAL FRAME DECOMPOSITION

In the proof of Proposition 4.2 in [7] it is essential to require that the dual system $X(\tilde{\Psi})$ has a sufficient number of vanishing moments. The remainder of this paper is devoted to study ways of avoiding this requirement since many bi-framelet systems only have few vanishing moments. The key is to oversample the system. Given a finite collection of functions $\Psi = \{\psi^\ell\}_{\ell \in E}$ in L_2 and $R \geq 1$ we let $X_R(\Psi)$ denote the oversampled system,

$$X_R(\Psi) := \left\{ 2^{j/2} \psi^\ell(2^j \cdot - k/R) : j, k \in \mathbb{Z}, \ell \in E \right\}.$$

It turns out that if $X(\Psi)$ is a frame in L_2 so is the oversampled system $X_R(\Psi)$. Moreover, if $X(\Psi), X(\tilde{\Psi})$ is a "nice" bi-framelet system, the oversampled system $X_{2^N}(\Psi)$, $N \in \mathbb{N}$, again, gives rise to an approximation space no larger than a Besov space, i.e. the Bernstein inequality proven in Proposition 4.1 holds true for the oversampled system.

Again, in order to get a complete characterization of the approximation spaces based on $X_R(\Psi)$ in terms of Besov spaces, we need to prove a matching Jackson estimate. The following result was given in [7].

Proposition 5.1 *Let $X(\Psi)$, $X(\tilde{\Psi})$ be a wavelet bi-frame system, $X(\eta)$ a bi-orthogonal wavelet basis and $r > 0$ such that the Besov space $B_\tau^r(L_\tau)$, $0 < \tau < \infty$, can be characterized by*

$$\left\{ f \in L_\tau : \sum_{j,k \in \mathbb{Z}} (2^{j(r+1/2-1/\tau)} |\langle f, \eta_{j,k} \rangle|)^\tau < \infty \right\}. \quad (6)$$

Assume there exists a sequence $\{d_k^\ell\}_{\ell \in E, k \in \mathbb{Z}} \in \ell_{1/(r+1)}$, such that

$$\eta(x) = \sum_{\ell \in E} \sum_{k \in \mathbb{Z}} d_k^\ell \psi^\ell(x - k/R).$$

Then, for $1 < p < \infty$, and $0 < \alpha = 1/\tau - 1/p < r$, we have the Jackson inequality

$$\sigma_m(f, X_R(\Psi))_p \leq Cm^{-\alpha} \|f\|_{B_\tau^\alpha(L_\tau)}.$$

Now, by combining Proposition 4.1 (for the oversampled system) and Proposition 5.1 we obtain a complete characterization of the approximation spaces $\mathcal{A}_q^\beta(L_p, X_2(\Psi))$, not limited by the number of vanishing moments of the dual system $\tilde{\Psi}$.

In our strategy to get a Jackson inequality for the (oversampled) framelet system $X_R(\Psi)$, the crucial issue is to identify some "nice" wavelet(s) that can be expanded sparsely in terms of the oversampled bi-frame system. For spline-based *tight* framelets, it was shown in [8] how to get a finite expansion of a nice semi-orthogonal wavelet in the twice oversampled ($R = 2$) framelet system. In [7] a construction is given which is valid for more general bi-framelets and only requires $R = 2$.

Proposition 5.2 *Let $X(\Psi)$, $X(\tilde{\Psi})$ be an MRA-based wavelet bi-frame system and let ϕ and η be respectively the scaling function and the associated standard orthonormal wavelet. Suppose that each filter τ_ℓ, $\ell = 0, 1, \ldots, L$, is a trigonometric polynomial,*

$$\sum_{\ell \in E} |\tau_\ell(\xi)|^2 > 0, \quad \text{for } \xi \neq 0,$$

and ϕ is an r-regular scaling function (not necessarily orthonormal). Then there exists $\{d_k^\ell\} \in \bigcap_{\tau > 0} \ell_\tau$ such that

$$\eta(x) := \sum_{\ell \in E} \sum_{k \in \mathbb{Z}} d_k^\ell \psi^\ell(x - k/2).$$

5.1. Wavelets build with no oversampling

We conclude by improving the result from Proposition 5.2 by showing that in certain cases no oversampling is needed to build a wavelet out of the framelet system.

The following lemma is a restatement of the so-called oblique extension principle, see [5].

Lemma 5.3 *Let $X(\Psi)$ be a tight framelet system with associated trigonometric polynomials τ_ℓ, $\ell = 0, 1, \ldots, L$. Define $m_\ell(e^{-i\xi}) := \tau_\ell(\xi)$. Then there exists a rational function $T(z)$, with $T(1) = 1$, such that*

$$m_0(z)\overline{m_0(z^{-1})}T(z^2) + \sum_{\ell \in E} m_\ell(z)\overline{m_\ell(z^{-1})} = T(z), \quad (7)$$

$$m_0(z)\overline{m_0(-z^{-1})}T(z^2) + \sum_{\ell \in E} m_\ell(z)\overline{m_\ell(-z^{-1})} = 0. \quad (8)$$

From Lemma 5.3 we can deduce the following result. Let $G(z) = \gcd\{m_\ell(z) \colon \ell \in E\}$.

Theorem 5.4 *Suppose $m_0(z) = \left(\frac{1+z}{2}\right)^r R(z)$, for some $r \in \mathbb{N}$, where $|R(1)| = 1$ and $R(-1) \neq 0$. Then $G(z) = (1-z)^n \tilde{G}(z)$ for some $1 \leq n \leq r$, where $\tilde{G}(1) \neq 0$.*

Proof. By (7) we have

$$\sum_{\ell \in E} |m_\ell(1)|^2 = 1 - |m_0(1)|^2 = 1 - |R(1)|^2 = 0.$$

Thus $(1-z)|G(z)$, i.e., $n \geq 1$.

Now, suppose the lemma holds for some $n \geq 1$. Then (8) yields $(1-z)^n(1+z^{-1})^n = (-1)^n(1+z)^n(1-z^{-1})^n$ is a factor of $m_0(z)m_0(-z^{-1})T(z^2)$. Since $T(1) = 1$, $R(1) \neq 0$ and $R(-1) \neq 0$, this implies $n \leq r$. $\qquad\square$

Let ϕ be a scaling function generated by the refinement filter $\tau_0(\xi)$, and let $P(\xi) := \sum_{k \in \mathbb{Z}} |\hat{\phi}(\xi - k)|^2$. Define

$$\tilde{\tau}(\xi) := \frac{e^{-i\xi}\overline{\tau_0(\xi+\pi)}\sqrt{P(\xi+\pi)}}{\sqrt{P(2\xi)P(\xi)}},$$

and $M(e^{-i\xi}) = \tilde{\tau}(\xi)$. Notice that if $m_0(z) = \left(\frac{1+z}{2}\right)^r R(z)$ for some $r \in \mathbb{N}$, then $(1-z)^r$ is a factor of $M(z)$.

Recall that any orthonormal wavelet η associated with the scaling function ϕ is given by

$$\hat{\eta}(2\xi) = v(2\xi)\tilde{\tau}(\xi)\hat{\phi}(\xi) \qquad (9)$$

for some 2π-periodic function v with $|v| = 1$ a.e..

Let us try to expand such a wavelet in terms of integer shifts of the framelets ψ^ℓ. A **neccessary and sufficient condition** for such an expansion to be possible is the existence of polynomials $p^\ell(z)$ such that $\hat{\eta}(\xi) = \sum_{\ell \in E} p^\ell(e^{-i\xi})\hat{\psi}^\ell(\xi)$. Using (9) this is equivalent to the existence of polynomials $p^\ell(z)$ such that

$$\sum_{\ell \in E} p^\ell(e^{-i2\xi})\tau^\ell(\xi) = v(2\xi)\tilde{\tau}(\xi). \qquad (10)$$

According to Theorem 5.4 we have $m_\ell(z) = (1-z)^n \cdot (a^\ell(z^2) + zb^\ell(z^2))$ for some $1 \leq n \leq r$. Likewise we can write $M(z) = (1-z)^n(A(z^2) + zB(z^2))$. Thus, with $q(e^{-i\xi}) := v(\xi)$, (10) can be rewritten

$$q(z)\begin{bmatrix} A(z) \\ B(z) \end{bmatrix} = \begin{bmatrix} a^1(z) & \cdots & a^L(z) \\ b^1(z) & \cdots & b^L(z) \end{bmatrix}\begin{bmatrix} p^1(z) \\ \vdots \\ p^L(z) \end{bmatrix}. \qquad (11)$$

Now, using Proposition 4.1 and Proposition 5.1 we obtain.

Theorem 5.5 *Let $X(\Psi)$ be a tight framelet system based on a compactly supported refinable function $\phi \in W^s(L_\infty)$, $s \geq 0$. Suppose the wavelet η given by (9) can be used to characterize the Besov space $B_\tau^r(L_\tau)$ as given in (6), and*

suppose there exist polynomials $\{p^\ell(z)\}_{\ell \in E}$ such that (11) is satisfied. Let $p \in (1, \infty)$, $\alpha < \min\{s, r\}$, and $\tau := (\alpha + 1/p)^{-1}$. Then, for each $0 < \beta < \alpha$ and $q \in (0, \infty]$, we have the characterization

$$\mathcal{A}_q^\beta(L_p, X(\Psi)) = (L_p, B_\tau^\alpha(L_\tau))_{\beta/\alpha, q}.$$

6. CONCLUSION

We have studied nonlinear approximation with redundant wavelet bi-frame systems. Several results for orthonormal wavelets have been extended to the redundant case. In particular, the approximation spaces for such systems have been completely characterized. For "nice" wavelet bi-frames, the optimal approximation rate can be obtained simply by thresholding the canonical frame decomposition of a given signal.

For oversampled wavelet bi-frame systems, the approximation results are obtained without requireing vanishing moments. It is also shown that an oversampling ratio of two is sufficient. For special wavelet bi-frames, oversampling is not needed at all to obtain the approximation results.

7. REFERENCES

[1] Ronald A. DeVore, "Nonlinear approximation," in *Acta numerica, 1998*, pp. 51–150. Cambridge Univ. Press, Cambridge, 1998.

[2] Ronald A. DeVore and George G. Lorentz, *Constructive approximation*, Springer-Verlag, Berlin, 1993.

[3] Colin Bennett and Robert Sharpley, *Interpolation of operators*, Academic Press Inc., Boston, MA, 1988.

[4] Charles K. Chui, Wenjie He, and Joachim Stöckler, "Tight frames with maximum vanishing moments and minimum support," in *Approximation theory, X (St. Louis, MO, 2001)*, Innov. Appl. Math., pp. 187–206. Vanderbilt Univ. Press, Nashville, TN, 2002.

[5] Ingrid Daubechies, Bin Han, Amos Ron, and Zuowei Shen, "Framelets: MRA-based constructions of wavelet frames," *Appl. Comput. Harmon. Anal.*, vol. 14, no. 1, pp. 1–46, 2003.

[6] Amos Ron and Zuowei Shen, "Affine systems in $L_2(\mathbf{R}^d)$: the analysis of the analysis operator," *J. Funct. Anal.*, vol. 148, no. 2, pp. 408–447, 1997.

[7] Lasse Borup, Rémi Gribonval, and Morten Nielsen, "Bi-framelet systems with few vanishing moments characterize Besov spaces," *Appl. Comput. Harmon. Anal.*, vol. 17, no. 1, pp. 3–28, 2004.

[8] Rémi Gribonval and Morten Nielsen, "On approximation with spline generated framelets," *Constr. Approx.*, vol. 20, no. 2, pp. 207–232, 2004.