http://math-sci.ui.ac.ir

Journal of Mathematics and Society
ISSN (print): 2345-6493, ISSN (on-line): 2345-6507
Vol. 9 No. 3 (2024), pp. 35-79.

http://www.ui.ac.ir

WAVELETS, APPROXIMATION AND COMPRESSION**

MARTIN VETTERLI
TRANSLATORS: SARKOUT ABDI, ARAM AZIZI*, MAHMOUD SHAFIEE AND JAMSHID SAEIDIAN

ABSTRACT. Over the last decade or so, wavelets have had a growing impact on signal processing theory and practice, both because of their unifying role and their successes in applications. Filter banks, which lie at the heart of wavelet-based algorithms, have become standard signal processing operators, used routinely in applications ranging from compression to modems. The contributions of wavelets have often been in the subtle interplay between discrete-time and continuous-time signal processing. The purpose of this article is to look at recent wavelet advances from a signal processing perspective. In particular, approximation results are reviewed, and the implication on compression algorithms is discussed. New constructions and open problems are also addressed. Finding a good basis to solve a problem dates at least back to Fourier and his investigation of the heat equation. The series proposed by Fourier has several distinguishing features: The series is able to represent any finite energy function on an interval. The basis functions are eigenfunctions of linear shift invariant systems or, in other words, Fourier series diagonalize linear, shift invariant operators.

1. **Introduction**

Finding a good basis to solve a problem dates at least back to Fourier and his investigation of the heat equation [8]. The series proposed by Fourier has several distinguishing features:

- The series is able to represent any finite energy function on an interval.

Keywords: signal processing, filter banks, Fourier transform, wavelet transform, compress algoritm.
Article Type: Translation Paper.
Communicated by Alireza Amini Harandi.
*Corresponding author. Received: 31-05-2023, Accepted: 12-03-2024, Published Online: 19-10-2024.
The above abstract has been extracted by the translator from the original article (M. Vetterli, Wavelets, approximation and compression, *IEEE signal processing Magazine*, **18 no. 5 (2001) 59–73.)
http://dx.doi.org/10.22108/msci.2024.137864.1581 .

- The basis functions are eigenfunctions of linear shift invariant systems or, in other words, Fourier series diagonalize linear, shift invariant operators.

Assume we have a space of functions S and we wish to represent an element $f \in S$. The space Scan be, for example, integrable functions on the interval $[0, 1]$ with finite square integral

$$\int_0^1 |f(t)|^2 \mathrm{d}t < \infty, \tag{1.1}$$

which we denote $L_2(0, 1)$. The first question is then to find a basis for S, that is a set of basis functions $\{\varphi_i\}_{i \in I}$ in S such that any element $f \in S$ can be written as a linear combination

$$f = \sum_{i \in I} a_i \varphi_i. \tag{1.2}$$

The example closest to the heart of signal processing people is certainly the expansion of bandlimited functions in terms of the sinc function. Given the representation of f in an orthonormal basis as in formola, its orthogonal projection into a fixed subspace of dimension N spanned by$\{\varphi_n\}_{n=0,\ldots,N-1}$ is denoted $\hat{f}_N(t)$

$$\hat{f}_N = \sum_{n=0}^{N-1} \langle \varphi_n, f \rangle \varphi_n. \tag{1.3}$$

Given objects of interest and spaces in which they are embedded, we wish to know how fast an N-term approximation converges

$$\|f(t) - f_N(t)\|_2 \sim fct(N), \tag{1.4}$$

where $f_N(t)$ stands for an approximation of $f(t)$ which involves N elements, to be chosen appropriately. This immediately raises a number of questions. Different bases can give very different rates of approximation. Then, there are various ways to choose the N terms used in the approximation. A fixed subset (e.g., the first N terms) leads to a linear, subspace approximation as in (1.3). Adaptive schemes, to be discussed later, are nonlinear. Will different choices of the subset lead to different rates of approximation? Such questions are at the heart of approximation theory and are relevant when choosing a basis and an approximation method for a given signal processing problem. For example, denoising in wavelet bases has led to interesting results for piecewise smooth signals precisely because of the superior approximation properties of wavelets for such signals. We are now ready to address the last problem we shall consider, namely the compression problem. This involves not only approximation quality, but also description complexity. There is a cost associated with describingf_N , and this cost depends on the approximation method. Typically, the coefficient values and their locations need to be described, which involves quantization of the coefficients and indexing their locations.

 http://dx.doi.org/10.22108/msci.2024.137864.1581

2. Orthogonal Filter Banks

When thinking of filtering, one usually thinks about frequency selectivity. For example, an ideal discrete-time lowpass filter with cut-off frequency $\omega_c < \pi$ takes any input signal and projects it onto the subspace of signals bandlimited to $[-\omega_c, \omega_c]$. Orthogonal discrete-time filter banks perform a similar projection which we now review. Assume a discrete-time filter with finite impulse response $g_0[n] = \{g_0[1], g_0[1], g_0[L-1]\}$, L even, and the property

$$\langle g_0[n], g_0[n-2k]\rangle = \delta_k, \tag{2.1}$$

that is, the impulse response is orthogonal to its even shifts, and $\|g_0\|_2 = 1$. Denote by $G_0(z)$ the z-transform of the impulse response $g_0[n]$

$$G_0(z) = \sum_{n=0}^{L-1} g_0[n] z^{-n}, \tag{2.2}$$

with an associated region of convergence covering the z-plane except the origin. Assume further that $g_0[n]$ is a lowpass filter, that is, its discrete-time Fourier transform has most of its energy in the region $[-\pi/2, \pi/2]$. Then define a high-pass filter $g_1[n]$ with z-transform $G_1(z)$ as follows:

$$G_1(z) = z^{-L+1} G_0(-z^{-1}). \tag{2.3}$$

3. Discrete-Time Polynomials and Filter Banks

Signal processing specialists intuitively think of problems in terms of sinusoidal bases. Approximation theory specialists think often in terms of other series, like the Taylor series, and thus, of polynomials as basic building blocks. We now look at how polynomials are processed by filter banks. A discrete-time polynomial signal of degree M is composed of a linear combination of monomial signals

$$\rho^{(m)}[n] = n^m, \qquad 0 \le m \le M. \tag{3.1}$$

We shall now see that such monomial (and therefore polynomial) signals are eigensignals of certain multirate operators. We need to consider lowpass filters $G_0(z)$ that have a certain number$N > 0$ of zeroes at $z = 1$, or $\omega = \pi$ on the unit circle. That is, the filter factors as

$$G_0(z) = (1 + z^{-1})^N R_0(z). \tag{3.2}$$

Clearly, because of (2.3), the highpass $G_1(z)$ has N zeros at $z = 1$ or $(\omega = 0)$, while $H_0(z)$ and $H_1(z)$ have N zeros at z= 1 and 1, respectively because of formoula.

4. Continuous-Time Polynomials and Wavelets

As is well known, a strong link exists between iterated filter banks and wavelets. For example, filter banks can be used to generate wavelet bases[4], and filter banks can be used to calculate wavelet series [9]. It comes thus as no surprise that the properties seen in discrete time regarding polynomial representation carry over to continuous time. While these properties are directly related to moment properties of wavelets and thus hold in general, we review them in the context of wavelets generated from orthogonal finite impulse response (FIR) filter banks. Assume again that the lowpass filter has N zeros at $\omega = \pi$, and thus, the highpass has N zeros at $\omega = 0$. From the two scale relation of scaling function and wavelet, we get that the Fourier transform of the wavelet can be factored as

$$\Psi(w) = \frac{1}{\sqrt{2}} G_1(l^{jw/2}) \cdot \phi\left(\frac{w}{2}\right). \tag{4.1}$$

where $\phi(w)$ is the Fourier transform of the scaling function.

5. Discontinuities in Filter Bank and Wavelet Representations

What happens if a signal is discontinuous at some point t_0 ? we know that Fourier series do not like discontinuities, since they destroy uniform convergence. Wavelets have two desirable properties as far as discontinuities are concerned. First, they focus locally on the discontinuity as we go to finer and finer scales. That is because of the scaling relation of wavelets, where the function set $\Psi_{m,n}(t)$ is defined as

$$\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n) \qquad m, n \in \mathbb{Z} \tag{5.1}$$

where $m \to -\infty$ corresponds to fine details. Thus, as m grows negative, the wavelet becomes "sharper." If the discontinuity is isolated, and the surroundings are smooth, all wavelet inner products except the ones at the discontinuity will be zero, and around the discontinuity,$L1$inner products are different from zero when the wavelet has support length L. Second, the magnitude evolution across scales of the nonzero wavelet inner products characterizes the discontinuity. This is a well-known characteristic of the continuous wavelet transform [5, 10] and holds as well for the orthonormal wavelet series.

6. Linear Approximation

Assume a space$V$ and an orthonormal basis $\{g_n\}_{n\in N}$ for V. Thus, a function $f \in V$ can be written as a linear combination

$$f = \sum_{n\in N} \langle g_n, f\rangle g_n. \tag{6.1}$$

http://dx.doi.org/10.22108/msci.2024.137864.1581

The best (in the squared error sense) linear approximation of f in the subspace$V$, denoted as $\hat{f}$, is given by the orthogonal projection of f onto a fixed subspace of V (see (1.3)). Assume it is a subspaceW of dimension M, and spanned by the first M vectors of the basis, or

$$W = span\{g_0, g_1, g_2, \dots, g_{M-1}\}$$

. Then $\hat{f}$ is given by

$$\hat{f}_M = \sum_{n=0}^{M-1} \langle g_n, f\rangle g_n. \tag{6.2}$$

and the squared error of the approximation is

$$\hat{\varepsilon} = \|f - \hat{f}\|_2^2 = \sum_{n=M}^{\infty} |\langle g_n, f\rangle|^2.$$

Because the subspace$W$ is fixed, independent of f , the approximation is linear.

7. Nonlinear Approximation in Orthonormal Bases

Consider the same set up as above, but with a different approximation rule. Instead of (6.1) , where the first$M$ coefficients in the orthonormal expansion are used, we keep the M largest coefficients instead. That is, we define an index set I_M of the M largest inner products, or

$$|\langle g_m, f\rangle| \geq |\langle g_n, f\rangle|, \qquad m \in I_m,\ n \notin I_m. \tag{7.1}$$

Then, we define, the best nonlinear approximation as:

$$\hat{f}_M = \sum_{n\in I_M} \langle g_n, f\rangle g_n \tag{7.2}$$

which leads to an approximation error

$$\tilde{\varepsilon}_M = \|f - \tilde{f}\|_2^2 = \sum_{n\in I_M} |\langle g_n, f\rangle|^2$$

Clearly

$$\tilde{\varepsilon}_M \leq \hat{\varepsilon}_M. \tag{7.3}$$

(We could call this adaptive linear approximation or adaptive subspace approximation. However, the commonly used term is nonlinear approximation. More general nonlinear schemes could also be considered, but are beyond the scope of this article.)

8. Nonlinear Approximation and Compression

So far, we have considered keeping M elements from a basis, either a fixed set (the first M typically) or an adaptive set (corresponding to the largest projection). In compression, we have first to describe the coefficient set, which has zero cost when this set is fixed (linear approximation) but has a nontrivial cost when it is adaptive (nonlinear approximation). In that case, there are $\binom{N}{M}$ possible subsets, and the rate to describe the subset is equal to the entropy of the distribution of the subsets, or at most

$$\log_2 \binom{N}{M} \tag{8.1}$$

9. Compression of Piecewise Polynomial Signals

Let us return to one-dimensional piecewise smooth signals. Wavelets are well suited to approximate such signals when nonlinear approximation is allowed. To study compression behavior, consider the simpler case of piecewise polynomials, with discontinuities. To make matters easy, let us look again at the signal we used earlier to study nonlinear approximation, but this time include quantization and bit allocation. A simple analysis of the approximate rate distortion behavior of a step function goes as follows. Coefficients decay as $2^{m/2}$, so the number of scales J involved, if a quantizer of size Δ is used, is of the order of $\log_2(1/\Delta)$. The number of bits per coefficient is also of the order of $\log_2(1/\Delta)$, so the rate R is of the order

$$R \sim \log_2^2 (1/\Delta) \sim J^2 \tag{9.1}$$

The distortion or squared error is proportional to Δ^2 (for each coefficient), times the number of scales or, using $\Delta = 2^{-J}$

$$D \sim J \cdot 2^{-J}. \tag{9.2}$$

Using (9.1), we get

$$D_W(R) = C_1 \cdot \sqrt{R} \cdot 2^{C_2\sqrt{R}}. \tag{9.3}$$

for the distortion-rate behavior of a wavelet scheme. Note that we ignored the cost of indexing the location. This cost turns out to be quite small (order J), because the coefficients are all gathered around the discontinuity.

10. True Two-Dimensional Bases

As the wavelet example shows, separable bases are not suited for "true" two-dimensional objects. What is needed are transforms and bases that include some form of "geometry" and that are truly two dimensional. (The notion of geometry is not easy to formalize in our context, but the intuition is that the dimensions are not independent, and certain shapes are more likely than others.) Besides

http://dx.doi.org/10.22108/msci.2024.137864.1581

the two-dimensional Fourier and wavelet transforms, which are both separable, the Radon transform plays a key role. This transform, studied early in the 20th century [12], was rediscovered several times in fields ranging from astronomy to medical imaging (see [6] for an excellent overview). The Radon transform maps a function $f(x, y)$ into $RA_f(\theta, t)$ by taking line integrals at angle θ and location t

$$RA_f(\theta, t) = \iint f(x, y)\delta(x\cos\theta + y\sin\theta - t)\mathrm{d}x\mathrm{d}y \tag{10.1}$$

A key insight to construct directional bases from the Radon transform was provided by Candès and Donoho [2, 3] , with the ridgelet transform. The idea is to map a one-dimensional singularity, like a straight edge, into a point singularity using the Radon transform. Then, the wavelet transform can be used to handle the point singularity.

11. Directional Filter Banks

To get directional analysis, one can alternatively use directional filter banks[1]. In such a case, the basis functions are given by the filter impulse responses and their translates with respect to the subsampling grid. Such filter banks can be designed directly, or through iteration of elementary filter banks. They lead to bases if critically subsampled, or frames if oversampled.

12. Two-Dimensional Bases and Compression

As we had seen in one dimension, a good N-term approximation is not yet a guarantee for good compression. While a powerful N-term approximation is desirable, it must be followed by appropriate compression. Thus, the topic of compression of two-dimensional piecewise smooth functions is still quite open. Several promising approaches are currently under investigation, including compression in ridgelet and curvelet domain, compression along curves using “bandelets”[11] and generalization of footprints in two dimensions or edgeprints [7].

13. Conclusions

The interplay of representation, approximation, and compression of signals was reviewed. For piecewise smooth signals, we showed the power of wavelet-based methods, in particular for the one-dimensional case. For two-dimensional signals, where wavelets do not provide the answer for piecewise smooth signals with curve singularities, new approaches and open problems were indicated. Such approaches rely on new bases with potentially high impact on image processing, for such problems as denoising, compression and classification.

http://dx.doi.org/10.22108/msci.2024.137864.1581

References

[1] R. H. Bamberger and M. J. T. Smith, A filter bank for the directional decomposition of images: Theory and design, *IEEE Trans. Signal Processing*, **40** (1992) 882–893.

[2] E. Candès, Ridgelets: Theory and applications, Ph.D. dissertation, Dept. Statistics, Stanford University, Stanford, CA, 1998.

[3] E. J. Candès and D. L. Donoho, Curvelets—A surprisingly effective nonadaptive representation for objects with edges, *in Curve and Surface Fitting*, A. Cohen, C. Rabutand, and L. L. Schumaker, Eds. Saint-Malo: Vanderbilt University Press, 1999.

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

[5] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia, PA: SIAM, 1992.

[6] S. R. Deans, *The Radon Transform and Some of its Applications*, New York: Wiley, 1983.

[7] P. L. Dragotti and M. Vetterli, Footprints and edgeprints for image denoising and compression, *in Proc. IEEE Int. Conf. Image Processing, ICIP 2001*, Patras, Greece, 2001.

[8] J. Fourier, *Théorie Analytique de la Chaleur*, Paris, France: Gauthier-Villars, 1888.

[9] S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Recognition Machine Intell.*, **11** (1989) 674–693.

[10] S. Mallat and W. L. Hwang, Singularity detection and processing with wavelets, *IEEE Trans. Inform. Theory (Special Issue on Wavelet Transforms and Multiresolution Signal Analysis)*, **38** (1992) 617–643.

[11] E. Le Pennec and S. Mallat, Image compression with geometric wavelets, *in Proc. IEEE Int. Conf. Image Processing, ICIP 2000*, Vancouver, Canada, Sept. (2000) 661–664.

[12] J. Radon, Ueber die bestimmung von funktionen durch ihre integralwerte längst gewisser mannigfaltigkeiten, *Berichte Sächsische Akademie der Wissenschaften, Leipzig*, (1917) 262–267.

Sarkout Abdi

Department of Mathematics, Rasht Branch, Islamic Azad University, Rasht, Iran

Aram Azizi

Department of Mathematics, Payame Noor University, P.O. Box 19395-3697 Tehran, Iran

Email:a.azizi@pnu.ac.ir

Mahmoud Shafiee

Department of Mathematics, Rasht Branch, Islamic Azad University, Rasht, Iran

Jamshid Saeidian

Faculty of Mathematical Sciences and Computer, Kharazmi University, 50 Taleghani avenue, Tehran 1561836314, Iran

http://dx.doi.org/10.22108/msci.2024.137864.1581