The emergence of sparse analysis through the works of Yves Meyer

Albert Cohen

Since the turn of the century, the concept of *sparsity* has become prominent in many areas of applied mathematics, at the crossroad with other disciplines such as signal and image processing. Loosely speaking, sparse approximation refers to the possibility of representing certain complex objects by very few numbers up to a controlled loss of accuracy. This task is usually made possible once a proper mathematical representation of these objects has been provided, for example through a change of basis.

Such economical representations play an obvious role in data compression. They also found powerful applications in statistical estimation and inverse problems. The spectacular development of *compressed sensing* led to a deeper mathematical understanding of how sparsity could be exploited in many such applications. While Yves Meyer's work were not primarily focused on such areas, his contributions played a key role in the emergence and spreading of this concept and in the development of its mathematical foundations.

The present paper includes parts of a more detailed survey on the research works of Yves Meyer published in the Abel volume [6].

1 A golden decade

The process of *analyzing* and *representing* an arbitrary function f by means of more elementary functions has been at the heart of fundamental and applied advances in science and technology for several centuries. In more recent decades, implementation of this process on computers by fast algorithms has become of ubiquitous use in scientific computing. In the foundational example of the univariate Fourier expansions these elementary building blocks are the 1-periodic complex exponential functions defined by

$$e_n(t) = e^{i2\pi nt}, \quad n \in \mathbb{Z}, \tag{1.1}$$

and they form an orthonormal basis of $L^2(]0,1[)$.

The functions e_n are perfectly localized in frequency but have no localization in time, since their modulus is equal to 1 independently of t. This property constitutes a major defect when trying to efficiently detect the local frequency content of functions by means of Fourier analysis. It also makes Fourier representations numerically ineffective for functions that are not smooth everywhere. For example, the Fourier coefficients $c_n(f)$ of a 1-periodic piecewise smooth function f with a jump discontinuity at a single point $t_0 \in [0,1]$ decay like $|n|^{-1}$, which affects the convergence of the Fourier series on the whole of $\mathbb{R}$.

Wavelet bases are orthonormal bases of $L^2(\mathbb{R})$ with the general form

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \quad j \in \mathbb{Z}, k \in \mathbb{Z}, \tag{1.2}$$

where ψ is a function such that $\int_{\mathbb{R}} \psi = 0$. Due to their increased resolution as the scale level j tends to $+\infty$, they are better adapted than Fourier bases for capturing local phenomenons. The most basic example of the Haar system, which corresponds to $\psi = \chi_{[0,1/2[} - \chi_{[1/2,1[}$, was already known since 1911. In this example, the function ψ suffers from a lack of smoothness, also reflected by the slow decay at infinity of its Fourier transform.

The construction of modern wavelet theory took place during the decade of 1980-1990. It benefited greatly from ideas coming from various (and sometimes completely disjoint) sources: theoretical harmonic analysis, approximation theory, computer vision and image analysis, computer aided geometric design, digital signal processing. One of the fundamental contributions of Yves Meyer was to recognize and organize these separate developments into a unified and elegant theory.

After some attempts to disprove their existence, Yves Meyer turned the table in 1985 and gave a beautiful construction of orthonormal wavelet bases that belong to the Schwartz class

$$\mathcal{S}(\mathbb{R}) := \{f \in C^\infty(\mathbb{R}) \ : \ \sup_{x\in\mathbb{R}} |x|^k |f^{(l)}(x)| < \infty, \ k,l \geq 0\}, \tag{1.3}$$

and are therefore well localized both in time and frequency. Another orthonormal wavelet basis with smoothness and localization properties had been obtained earlier in the work of Jan-Olov Strömberg. By its elegant simplicity, Yves Meyer's construction was celebrated as a milestone.

A major turning point occured in 1986 when Stéphane Mallat introduced the general framework that was the key to the general construction of wavelets [20]. A *multiresolution approximation* is a dense nested sequence of approximation spaces

$$\{0\} \cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \cdots L^2(\mathbb{R}), \tag{1.4}$$

generated by a so-called *scaling function* φ in the sense that $(2^{j/2}\varphi(2^j \dot{-} k))_{k\in\mathbb{Z}}$ is a Riesz basis of V_j. A countable family $(e_k)_{k\in\mathcal{F}}$ in a Hilbert space V is called a Riesz basis if it is complete and there exists constants $0 < c \le C < \infty$ such that

$$c \sum_{k\in\mathcal{F}} |x_k|^2 \le \| \sum_{k\in\mathcal{F}} c_k e_k \|_V^2 \le C \sum_{k\in\mathcal{F}} |x_k|^2, \tag{1.5}$$

holds for any finitely supported coefficient sequence $(x_k)_{k\in\mathcal{F}}$, and therefore by density for any sequence in $\ell^2(\mathcal{F})$.

In this framework, the generating wavelet ψ is then constructed so that the functions $(2^{j/2}\psi(2^j \dot{-} k))_{k\in\mathbb{Z}}$ constitute a Riesz basis for a complement W_j of V_j into V_{j+1}. This approach allowed in particular the construction of compactly supported orthonormal wavelets by Ingrid Daubechies [11], which was followed by that of biorthogonal wavelets [8].

The multiresolution analysis framework was immediately extended by Stéphane Mallat and Yves Meyer to multivariate functions, by tensorizing the spaces V_j in the different variables. This leads to multivariate wavelet bases of the form

$$\psi^{\varepsilon}_{\varepsilon,j,k} = 2^{dj/2}\psi_\varepsilon(2^j \cdot -k), \quad j \in \mathbb{Z},\ k \in \mathbb{Z}^d, \tag{1.6}$$

for $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d) \in \{0,1\}^d \setminus \{(0,\dots,0)\}$, where

$$\psi_\varepsilon(x_1,\dots,x_d) := \psi_{\varepsilon_1}(x_1)\cdots\psi_{\varepsilon_d}(x_x), \quad \psi_0 := \varphi,\ \psi_1 := \psi. \tag{1.7}$$

Adaptation of these bases to more general bounded domains of $\mathbb{R}^d$ as well as to various types of manifolds came in the following years, again based on the multiresolution concept.

All these developments are well documented in the classical textbooks [11, 26]. One of their major stimulus was the vision of powerful applications in areas as diverse as signal and image processing, statistics, or fast numerical simulation. This perspective was confirmed in the following decades. Yves Meyer played a key role in identifying the mathematical properties that are of critical use in such applications, in particular the ability of wavelets to characterize a large variety of function spaces. As we next discuss, these properties naturally led to the concept of sparse approximation. He was also one of the first to point out some intrinsic limitations of wavelets and promote alternative analysis strategies, some of which are discussed further.

2 Function spaces and unconditional bases

When expanding a function f into a given basis $(e_n)_{n\ge0}$, a desirable feature is that the resulting decomposition $f = \sum_{n\ge0} x_n e_n$ is numerically stable: operations such as perturbations, thresholding or truncation of the coefficients x_n should effect the norm of f in a well-controled manner. Such prescriptions can be encapsulated in the following classical property.

A sequence $(e_n)_{n\ge0}$ in a separable Banach space X is an *unconditional basis*, if the following properties hold:

(i) It is a *Schauder basis*: every $f \in X$ admits a unique expansion $\sum_{n\ge0} x_n e_n$ that converges towards f in X.

(ii) There exists a finite constant $C \ge 1$ such that for any finite set $F \subset \mathbb{N}$,

$$|x_n| \le |y_n|, \quad n \in F \implies \Big\| \sum_{n\in F} x_n e_n \Big\|_X \le C \Big\| \sum_{n\in F} y_n e_n \Big\|_X \tag{2.1}$$

The property (2.1) means that membership of f in X only depends on the moduli of its coordinates $|x_n|$. In other words, multiplier operators of the form

$$T : \sum_{n\geq 0} x_n e_n \to \sum_{n\geq 0} c_n x_n e_n, \tag{2.2}$$

should act boundedly in X if $(c_n)_{n\geq 0}$ is a bounded sequence. Orthonormal and Riesz bases are obvious examples of unconditional bases in Hilbert spaces.

While the trigonometric system is a Schauder bases in $L^p(]-\pi,\pi[)$ when $1 < p < \infty$, it does not constitute an unconditional basis when $p \neq 2$, and it is thus not possible to characterize the space L^p through a property of the moduli of the Fourier coefficients. The same situation is met for classical smoothness spaces, such as the Sobolev spaces $W^{m,p}_{\rm per}(]-\pi,\pi[)$ that consist of 2π-periodic functions having distributional derivatives up to order m in $L^p_{\rm loc}$: apart from the Hilbertian case $p = 2$, for which one has

$$f \in W^{m,2}_{\rm per}(]-\pi,\pi[) \iff \sum_{n\in\mathbb{Z}} (1+|n|^{2m})|c_n(f)|^2 < \infty, \tag{2.3}$$

no such characterization is available when $p \neq 2$.

Yves Meyer showed that, in contrast to the trigonometric system, wavelet bases are unconditional bases for most classical function spaces that are known to possess one. The case of L^p spaces for $1 < p < \infty$ is treated by the following observation: if the general wavelet ψ has C^1 smoothness, the multiplier operator (2.2) by a bounded sequence belongs to a classical class of integral operators introduced by Calderon and Zygmund, which are proved to act boundedly in $L^p(\mathbb{R}^d)$. Conversely, Yves Meyer showed that Calderon-Zygmund operators are “almost diagonalized” by wavelet bases in the sense that the resulting matrices have fast off-diagonal decay. This property plays a key role in the numerical treatment of partial differential and integral equations by wavelet methods, as discussed for instance in [1, 7].

The characterization of more general function spaces by the size properties of wavelet coefficients is particularly simple for an important class of smoothness spaces introduced by Oleg Besov. There exist several equivalent definitions of Besov spaces. The original one uses the m-th order L^p-modulus of smoothness

$$\omega_m(f,t)_p := \sup_{|h|\leq t} \|\Delta_h^m f\|_{L^p}, \tag{2.4}$$

where Δ_h^m is the m-th power of the finite difference operator $\Delta_h : f \mapsto f(\cdot + h) - f$. For $s > 0$, any integer $m > s$, and $0 < p, q < \infty$, a function $f \in L^p(\mathbb{R}^d)$ belongs to the space $B^{s,p}_q(\mathbb{R}^d)$ if and only if the function $g : t \to t^{-s}\omega_m(f,t)_p$ belongs to $L^q([0,\infty[, \frac{dt}{t})$. One may use

$$\|f\|_{B^{s,p}_q} := \|f\|_{L^p} + |f|_{B^{s,p}_q}, \quad \text{with} \quad |f|_{B^{s,p}_q} := \|g\|_{L^q([0,\infty[,\frac{dt}{t})}, \tag{2.5}$$

as a norm for such spaces, also sometimes denoted by $B^s_q(L^p(\mathbb{R}^d))$. Roughly speaking, functions in $B^{s,p}_q(\mathbb{R}^d)$ have up to s (integer or not) derivatives L^p. The third index q may be viewed as a fine tuning parameter, which appears naturally when viewing Besov spaces as real interpolation spaces between Sobolev space [2]: for example, with $0 < s < m$,

$$B^s_q(L^p) = [L^p, W^{m,p}]_{\theta,q}, \quad s = \theta m. \tag{2.6}$$

Particular instances are the Hölder spaces $B^{s,\infty}_\infty = C^s$ and Sobolev spaces $B^{s,p}_p = W^{s,p}$, when s is not an integer or when $p = 2$ for all values of s.

Let (ψ_λ) denote a multivariate wavelet basis of the type (1.6), where for simplicity λ denotes the three indices (e, j, k) in (1.6). Denoting by $|\lambda| := j = j(\lambda)$ the scale level of $\lambda = (e, j, k)$, we consider the expansion

$$f = \sum_{|\lambda|\geq 0} d_\lambda \psi_\lambda, \tag{2.7}$$

where the coarsest scale level $|\lambda| = 0$ also includes the translated scaling functions that span V_0.

The characterization of $B^{s,p}_q(\mathbb{R}^d)$ established by Yves Meyer for such expansions requires some minimal prescriptions: one assumes that for an integer $r > s$ the univariate generating wavelet ψ and scaling functions φ that defines (1.6) have derivatives up to order r that decay sufficiently fast at infinity, for instance faster than any polynomial rate, and that $\int_{-\infty}^{+\infty} t^k \psi(t)dt = 0$ for all $k = 0, 1, \dots, r-1$. Then, one has the norm equivalence

$$\|f\|_{B^{s,p}_q} \sim \|\varepsilon\|_{\ell^q}, \tag{2.8}$$

where the sequence $\varepsilon = (\varepsilon_j)_{j\geq 0}$ is defined by

$$\varepsilon_j := 2^{sj} 2^{(\frac{d}{2}-\frac{d}{p})j} \|(d_\lambda)_{|\lambda|=j}\|_{\ell^p}. \tag{2.9}$$

A closely related characterization of Besov spaces uses the Littlewood-Paley decomposition

$$f = S_0 f + \sum_{j\geq 0} \Delta_j f, \quad \Delta_j f := S_{j+1} f - S_j f, \tag{2.10}$$

where $\widehat{S_j f}(\omega) := \Theta(2^{-j}\omega)\hat{f}(\omega)$ with Θ a smooth compactly support function with value 1 for $|\omega| \leq 1$. It has the same form as above, with now $\varepsilon_j := \|\Delta_j f\|_{L^p}$. In the wavelet characterization the dyadic blocks are further discretized into the local components $d_\lambda \psi_\lambda$. Similar results have been obtained for Besov spaces defined on general bounded Lipschitz domains $\Omega \subset \mathbb{R}^d$ with wavelet bases adapted to such domains.

The norm equivalence (2.8) shows that membership of f in Besov spaces is characterized by simple weighted summability properties of its wavelet coefficients. In the particular case $q = p$, this equivalence takes the very simple form

$$\|f\|_{B^{s,p}_p} \sim \|(2^{(s+\frac{d}{2}-\frac{d}{p})|\lambda|} d_\lambda)\|_{\ell^p}. \tag{2.11}$$

As an immediate consequence, classical results such as the critical Sobolev embedding $B^{s,p}_p \subset L^2$ for $s = \frac{d}{p} - \frac{d}{2}$ take the trivial form of the embedding $\ell^p \subset \ell^2$ for $p < 2$. While this embedding is not compact, an interesting approximation property holds: when retaining only the n largest coefficients in the wavelet decomposition of f, the resulting approximation f_n satisfies

$$\|f - f_n\|_{L^2} \leq C n^{-r} \|f\|_{B^{s,p}_p}, \quad r := \frac{s}{d}. \tag{2.12}$$

This follows immediately from the fact that, for $p < 2$, the decreasing rearrangement of $(d_k)_{k\geq 1}$ of a sequence $(d_\lambda) \in \ell^p$ satisfies the tail bound

$$\Big(\sum_{k\geq n} d_k^2\Big)^{1/2} \leq n^{\frac{1}{2}-\frac{1}{p}} \|(d_\lambda)\|_{\ell^p}. \tag{2.13}$$

This last estimate shows that ℓ^p summability governs the compressibility of a sequence, in the sense of how fast it can be approximated by *n-sparse* sequences. The theory of best n-term wavelet approximation, generalizing the above remarks, has been developed by Ronald DeVore and his collaborators, in close relation with other nonlinear approximation procedures such as free knot splines or rational approximation, see [13] for a detailed survey.

A particularly useful feature of nonlinear wavelet approximation is that piecewise smooth signals, such as images, can be efficiently captured since the large coefficients are only those of the few wavelets whose supports contain the singularities. This is an instance of *sparse approximation* which aims at accurately capturing functions by a small number of well chosen coefficients in a basis or dictionary expansion. Sparse approximation in unconditional bases was identified in [14] by David Donoho as a key ingredient for powerful applications in data compression and statistical estimation, in particular through thresholding algorithms that were developped jointly with Iain Johnstone, Gérard Kerkyacharian and Dominique Picard [16]. A detailed exposition of these applications can be found in [21]. Pushed into the forefront by the work of Yves Meyer, David Donoho and Ronald DeVore, sparse approximation became within a few years a prominent concept in signal processing and scientific computing.

3 Taking off from the wavelet world

The estimate (2.13) shows that the rate n^{-r} of best n-term approximation of a function, using an orthonormal or Riesz basis (ψ_λ), is implied by the ℓ^p summability of is coefficient sequence (d_λ) with

$\frac{1}{p} = r + \frac{1}{2}$. A more refined analysis shows that this rate is exactly equivalent to the slightly weaker property that (d_λ) belongs to $w\ell^p$, which means that its decreasing rearrangement has the decay property

$$d_k \leq Ck^{-1/p}. \tag{3.1}$$

The spaces ℓ^p and $w\ell^p$ are thus natural ways of quantifying sparsity of a function when decomposed in an arbitrary orthonormal or Riesz basis (ψ_λ) of a Hilbert space.

In the case of wavelet bases, these summability properties are equivalent to Besov smoothness. From an applicative point of view, a more natural question is: given a class of functions $\mathcal{K}$ in a Hilbert space, which basis should be used in order to obtain the sparsest possible representations of the element of this class ? In view of the previous observations, this basis should be picked so that the coefficient sequence of any element of $\mathcal{K}$ belongs to $w\ell^p$, for the smallest possible value of p.

One class of particular interest for modeling real images is the space $\mathrm{BV}(Q)$ of bounded variation functions on the unit cube $Q := [0,1]^2$ that consists of functions $f \in L^1(Q)$ such that ∇f is a finite measure. In particular, if $\Omega \subset Q$ is a set of finite perimeter, the characteristic function χ_Ω belongs to $\mathrm{BV}(Q)$. More generally, piecewise smooth images with edges discontinuities across curves of finite length have bounded variation. While the space $\mathrm{BV}(Q)$ admits no unconditional basis, it was shown in [9] that it can be "almost" characterized by its decomposition in a bivariate wavelet basis (ψ_λ) in the following sense: if $f = \sum d_\lambda \psi_\lambda$, one has

$$(d_\lambda) \in \ell^1 \implies f \in \mathrm{BV}(Q) \implies (d_\lambda) \in w\ell^1. \tag{3.2}$$

In view of the previous remarks shows that for general images of bounded variation, the rate of best n-term approximation in wavelet bases is $n^{-1/2}$. It can be shown that this rate is also the best that can be achieved by any basis. In particular, no polynomial rate can be achieved when using the Fourier basis. In this sense wavelets appear as the optimal tool for piecewise smooth images with edges of finite length.

The situation becomes quite different if one considers images with edges enjoying some geometric smoothness in addition to finite length. The simplest model consists of piecewise constant images with straight edges. For such images, Yves Meyer noticed in [27] that the decreasingly rearranged Fourier coefficients decay at rate

$$c_k \leq Ck^{-1}\log(k), \tag{3.3}$$

therefore comparable to wavelet coefficients up to the logarithmic factor. When going to a higher dimensional cube $Q = [0,1]^d$, this rate persist for Fourier coefficients while wavelet representation become less effective.

A more elaborate model consists of the functions which are piecewise C^m with edges discontinuities having C^n geometric smoothness. For such classes $\mathcal{K}(n,m)$, both wavelet and Fourier decompositions can be outperformed by more sophisticated representations into function that combine local support with directional selectivity. One representative example are the *curvelets*, introduced by Emmanuel Candes and David Donoho in [3], which have the form

$$\psi_\lambda = 2^{3j/2}\psi(D^j R_j^l \cdot - k), \quad k \in \mathbb{Z}, \quad l = 0, \dots, 2^j - 1, \tag{3.4}$$

where D is the anisotropic dilation matrix $\left(\begin{smallmatrix} 4 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ and R_j the rotation of angle $2^{-j-1}\pi$. The anisotropic scaling and angular selectivity allow to better capture the geometry of edges, leading to improved sparsity: for example, it is known that

$$f \in \mathcal{K}(2,2) \implies (d_\lambda) \in \ell^p, \quad p > \frac{2}{3}, \tag{3.5}$$

where d_λ are the coefficients of f in the curvelet expansion. The value $\frac{2}{3}$ is optimal for this class. Other representation methods have since then proposed and studied for better capturing geometry: contourlets, shearlets, bandlets, anisotropic finite elements.

Returning to univariate signals, one object of long term interest to Yves Meyer are signals whose "instantaneous frequency" evolves with time in some controlled manner. Such signals are called *chirps* and take the general form

$$f(t) = Re\left(a(t)e^{i\varphi(t)}\right), \tag{3.6}$$

where $\left|\frac{a'(t)}{a(t)}\right| << |\varphi'(t)|$ and $|\varphi''(t)| << |\varphi'(t)|^2$.

Typical examples of chirp are ultrasounds emitted by bats and recordings of voice signals, but the most famous one is the gravitational wave signal first detected in 2016 which has for a large part the behaviour

$$f(t) \sim |t - t_0|^{-1/4} \cos(|t - t_0|^{5/8} + \varphi_0). \tag{3.7}$$

Wavelets are not the right tool for sparse representation of chirps. Time-frequency analysis such as the short-time Fourier transform provides more natural tools, once proper orthonormal bases have been provided.

The first example of such a basis was originally suggested by Kenneth Wilson and formalized by Ingrid Daubechies, Stéphane Jaffard and Jean-Lin Journé [12]: an orthonormal basis of $L^2(\mathbb{R})$ is constructed by taking for all $n \in \mathbb{Z}$ the functions $\varphi_{0,n}(t) = \varphi(t - n)$ and

$$\varphi_{l,n}(t) = \begin{cases} \sqrt{2}\varphi\left(t - \frac{n}{2}\right)\cos(2\pi lt) & l \geq 0,\ l + n \in 2\mathbb{Z}, \\ \\ \sqrt{2}\varphi\left(t - \frac{n}{2}\right)\sin(2\pi lt) & l > 0,\ l + n \in 2\mathbb{Z} + 1, \end{cases} \tag{3.8}$$

The generating function φ should satisfy certain symmetry properties. One possible choice is the scaling function associated with the orthonormal wavelet basis of Yves Meyer, which is defined by $\hat{\varphi} = \sqrt{\kappa}$ where κ is the symmetric and smooth cut-off function. A variant of this system, where the family is made redundant by additional dilations, was proposed in the papers of Sergei Klimenko and his collaborators for the sparse representation of gravitational waves and used for their detection. Other examples of time-frequency bases are given in the textbook [18].

In recent years, sparse approximation has also been intensively exploited for the treatment of high-dimensional approximation. Problems which involve functions of a very large number of variables are challenged by the so-called "curse of dimensionality": the complexity of standard discretization methods blows up exponentially as the number of variables grows. Such problems arise naturally in learning theory, partial differential equations or numerical models depending on parametric or stochastic variables. Wavelet representation are not well suited for extracting sparsity in such high dimensional applications. This motivated the development of better adapted tools, such as sparse grids, sparse polynomials, and sparse tensor formats.

4 Compressed sensing and quasi-crystals

The most usual approach for obtaining a sparse approximation of a discrete signal represented by a vector $x \in \mathbb{R}^N$ is to choose an appropriate basis, compute the coefficients of x in this basis, and then retain only the n largest of these, with $n << N$.

This approximation process is *adaptive* since the indices of the retained coefficients vary from one signal to another. The view expressed by Emmanuel Candès, Justin Romberg, and Terence Tao [5, 4] and David Donoho [14] is that since only a few of these coefficients are in the end, it should be possible to only compute a few *non-adaptive* linear measurements in the first place and still retain the information needed in order to build a compressed representation. These ideas have led since the turn of the century to the very active area *compressed sensing.*

If m is the number of linear measurements, the observed data has the form

$$y = \Phi x, \tag{4.1}$$

where Φ is an $m \times N$ matrix. Any n-sparse vector x is uniquely characterized by its measurement if and only if no $2n$-sparse vector lies in the kernel of Φ. In other words, any submatrix of Φ_T obtained by retaining a set $T \subset \{1, \dots, N\}$ of column with $\#(T) = 2n$ should be injective. It is easily seen that a generic $m \times N$ matrix satisfies this property provided that $m \geq 2n$, and therefore $m = 2n$ linear measurement are in principle sufficient to be able to reconstruct n-sparse vectors. However, the reconstruction from $2n$ measurements will generally be computationally untractable when N is large and unstable due to the fact that $\Phi_T^*\Phi_T$, even if invertible, can be very ill-conditionned.

Stability and computational feasibility can be recovered at the expense of a stronger condition introduced by Emmanuel Candes, Justin Romberg and Terence Tao: the matrix Φ satisfies the *restricted isometry property* (RIP) of order k with parameter $0 < \delta < 1$, if and only if

$$(1 - \delta)\|z\|_2^2 \leq \|\Phi_T z\|_2^2 \leq (1 + \delta)\|z\|_2^2, \quad z \in \mathbb{R}^k, \tag{4.2}$$

for all set $T \subset \{1, \dots, N\}$ such that $\#(T) = k$. Under such a property with $k = 2n$ and $\delta < \frac{1}{3}$, it was shown that n-sparse vector can be stably reconstructed from its linear measurements by a convex optimization algorithm which consists in searching for the solution of (4.1) with minimal ℓ^1-norm. The value $\frac{1}{3}$ is not optimal and has been further improved [17].

Measurement matrices Φ of size $m \times N$ that satisfy RIP of order to k are known to exist in the regime $m \sim k \log(N/k)$. Therefore, with $k = 2n$ the measurement budget m is linear in n up to logarithmic factors. However the constructions of such matrices rely upon probabilistic arguments: they are realizations of random matrices for which it is proved that RIP of order $2n$ holds with high probability under this type of regime. Two notable examples are the matrix which entries consisting of independent centered Gaussian variables of variance $1/m$ and the matrix obtained by picking at random m rows from the $N \times N$ discrete Fourier transform matrix. The currently available deterministic constructions of matrices satisfying RIP of order k require the non-optimal regime $m \sim k^2$ up to logarithmic factors.

In recent years, Yves Meyer studied the problem of sampling continuous *bandlimited signals* with unknown Fourier support, which may be viewed as an analog counterpart to the above compressed sensing problem. For any set $E \subset \mathbb{R}^d$, we denote by $\mathcal{F}_E$ the Paley-Wiener space of functions $f \in L^2(\mathbb{R}^d)$ such that their Fourier transform

$$\hat{f}(\omega) = \int_{\mathbb{R}^d} f(x) \exp(-i2\pi\omega \cdot x) dx, \tag{4.3}$$

is supported on E. Sampling theory for such functions has been motivated since the 1960's by the development of discrete telecommunications. It is well known, since the foundational work of Claude Shannon and Harry Nyquist, that regular grids, which are full rank lattices

$$L = B\mathbb{Z}^d, \tag{4.4}$$

where B is a $d \times d$ invertible matrix, are particularly suitable for the sampling of certain band-limited functions. This may be seen as a direct consequence of the Poisson summation formula which says that for any sufficiently nice function f,

$$|L| \sum_{\lambda \in L} f(\lambda) e^{i2\pi\langle\lambda,\omega\rangle} = \sum_{\lambda^* \in L^*} \hat{f}(\omega + \lambda^*). \tag{4.5}$$

Here $|L| := |\det(B)|$ is the measure of the foundamental volume of L, and $L^* = (B^t)^{-1}\mathbb{Z}^d$ its dual lattice. This formula shows that, if $E \subset \mathbb{R}^d$ is a compact set with translates $(E + \lambda^*)_{\lambda^* \in L^*}$ having intersections of null measure, functions with Fourier transform supported in E are then stably determined by their sampling on L. Such sets E should in particular have measure smaller than the density of L, that is,

$$|E| \leq \mathrm{dens}(L) := |L^*| = |L|^{-1}. \tag{4.6}$$

One elementary example, for which equality holds in the above, is the fundamental volume of the lattice L^*, that is,

$$E_{L^*} = (B^*)^{-1}([0,1]^d), \tag{4.7}$$

or any of its translates.

A theory of stable sampling on more general discrete sets was developed in the 1960's by Henry Landau and Arne Beurling. The possibility of reconstructing any $f \in \mathcal{F}_E$ from its samples over a discrete set $\Lambda \subset \mathbb{R}^d$, is described by the property of *stable sampling*: there exists a constant C such that

$$\|f\|_{L^2}^2 \leq C \sum_{\lambda \in \Lambda} |f(\lambda)|^2, \quad f \in \mathcal{F}_E. \tag{4.8}$$

Henry Landau proved that a necessary condition for such a property to hold is that

$$\underline{\mathrm{dens}}(\Lambda) \geq |E|, \tag{4.9}$$

where

$$\underline{\mathrm{dens}}(\Lambda) = \liminf_{R \to \infty} \inf_{x \in \mathbb{R}^d} \frac{\#(\Lambda \cap B(x,R))}{|B(x,R)|}, \tag{4.10}$$

is the lower density of Λ, which is the usual density $\mathrm{dens}(\Lambda)$ when the standard limit exists.

A continuous signal is *s-sparse in the Fourier domain* if it belongs to $\mathcal{F}_E$ for some set of Lebesgue measure $|E| \leq s$. Stable reconstruction of any s-sparse signal from its sampling on a discrete set Λ requires that this set has the property of stable sampling for *all* sets of measure $|E| \leq r := 2s$. Such sets Λ are called *universal sampling sets*. Obviously, they should have density larger than r, however this condition cannot be sufficient. The case of a regular lattice L is instructive: on the one hand, the set E_{L^*} has measure $|E_{L^*}| = |L|^{-1} = \text{dens}(L)$ and satisfies the stable sampling property in view of (4.5). On the other hand, other sets E with the same or even smaller measure could have their translates by Λ^* overlapping with non-zero measure, which is a principle obstruction to these properties. This phenomenon is well known in electrical engineering as *aliasing*. This shows that universal sampling sets cannot be regular lattices.

Alexander Olevskii and Alexander Ulanovskii [28, 29] gave the first construction of a set Λ of uniform density that has the stable sampling property for any set E such that

$$|E| < \text{dens}(\Lambda). \tag{4.11}$$

Yves Meyer had the intuition that the mathematical models of *quasicrystals* that emerged from his early work on harmonic analysis and number theory [24, 25] could provide with a natural alternative solution to this problem.

One such model is obtained by the following *cut and project* scheme which was implicit in earlier work on algebraic number theory: the set of interest is obtained by projecting a "slice" cut from a higher dimensional lattice in general position. More precisely, if L is a full rank lattice of $\mathbb{R}^{d+m}$ for some $d, m > 0$, we denote by $p_1(x) \in \mathbb{R}^d$ and $p_2(x) \in \mathbb{R}^m$ the components of $x \in \mathbb{R}^{d+m}$ such that $x = (p_1(x), p_2(x))$ and assume that p_1 is a bijection between L and $p_1(L)$ with dense image. A similar property is assumed for p_2. Let $K \subset \mathbb{R}^m$ be a Riemann integrable compact set of positive measure. The associated *model set* $\Lambda = \Lambda(L, K) \in \mathbb{R}^d$ is defined by

$$\Lambda := \{p_1(x) \ : \ x \in L, \ p_2(x) \in K\}, \tag{4.12}$$

The density of a model set $\Lambda = \Lambda(L, K)$ is uniform and given by

$$\text{dens}(\Lambda) = \frac{|K|}{|L|}. \tag{4.13}$$

Bassarab Matei and Yves Meyer showed in [22, 23] that the stable sampling property holds for any E under the condition (4.11) for model sets $\Lambda := \Lambda(L, K) \subset \mathbb{R}^d$ such that K is a univariate interval. Such sets are called *simple quasicrystals*. They are universal sampling sets and may therefore be used for the reconstruction of s-sparse signals with $2s < \text{dens}(\Lambda)$. A remarkable fact is that, in contrast to compressed sensing matrices, their construction does not rely on any probabilistic argument.

References

[1] G. Beylkin, R. Coifman and V. Rokhlin, *Fast wavelet transforms and numerical algorithms*, Comm. Pure and Appl. Math. 44, 141-183, 1991.

[2] J. Bergh and J. Löfström, *Interpolation spaces*, Grundlehren 223, Springer, 1976.

[3] E. Candes and D.L. Donoho, *Curvelets and curvilinear integrals*, J. Approx. Theory 113, 5990, 2000.

[4] E. Candes, J. Romberg and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. 59, 1207-1223, 2005.

[5] E. Candes and T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inf. Theory 53, 5406-2425, 2006.

[6] A. Cohen, *A journey through the mathematics of Yves Meyer*, The Abel Volume, 2018.

[7] A. Cohen, W. Dahmen and R. DeVore, *Adaptive wavelet methods fo elliptic operator equations - convergence rates*, Math. Comp. 70, 27-75, 2000.

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

[9] A. Cohen, W. Dahmen, I. Daubechies and R. DeVore, *Harmonic analysis of the space* BV, Revista Mathematica Iberoamericana 19, 235-263, 2003.

[10] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure and Appl. Math. 41, 909-996, 1988.

[11] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, 1992.

[12] I. Daubechies, S. Jaffard and J.-L. Journé, *A simple wilson orthonormal basis with exponential decay*, SIAM J. Math. Anal. 22, 554-573, 1991.

[13] R. DeVore, *Nonlinear approximation*, Acta Numerica, 51-150, 1998.

[14] D.L. Donoho, *Unconditional bases are optimal for data compression and statistical estimation*, Appl. Comp. Harm. Anal. 1, 100-115, 1993.

[15] D. Donoho, *Compressed Sensing*, IEEE Transactions on Information Theory 52, 1289-1306, 2006.

[16] D. Donoho, I. Johnstone, G. Kerkyacharian and D. Picard, *Wavelet shrikage: Asymptotia? (with discussion)*, J. Roy. Stat. Soc. Ser. 57, 301-369, 1994.

[17] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Birkhaüser, 2013.

[18] S. Jaffard, Y. Meyer and R. Ryan, *Wavelets, tools for science and technology*, SIAM, Philadelphia, 2001.

[19] H.J. Landau, *Necessary density conditions for sampling and interpolation of certain entire functions*, Acta Math. 117, 37-52, 1967.

[20] S. Mallat, *Multiresolution approximation and wavelet orthonormal bases of* $L^2(\mathbb{R})$, Trans. Amer. Math. Soc. 315, 69-88, 1989.

[21] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1998.

[22] B. Matei and Y. Meyer, *A variant of compressed sensing*, Rev. Mat. Iberoam. 25, 669-692, 2009.

[23] B. Matei and Y. Meyer, *Simple quasicrystals are sets of stable sampling*, Journal of Complex Variables and Elliptic Equation 55, 947-964, 2010.

[24] Y. Meyer, *Nombres de Pisot, nombres de Salem et analyse harmonique*, Lecture Notes in Mathematics 117, Springer, 1970.

[25] Y. Meyer, *Algebraic numbers and harmonic analysis*, North-Holland, New York, 1972

[26] Y. Meyer, *Wavelets and Operators*, vol. 1, Cambridge university press, 1997.

[27] Y. Meyer, *Oscillating patterns in image processing and in some nonlinear evolution equations*, Jacqueline Lewis Memorial Lectures, AMS, 2001.

[28] A. Olevskii and A. Ulanovskii, *A universal sampling of band-limited signals*. C.R. Math. Acad. Sci. Paris 342, 927-931, 2006.

[29] A. Olevskii and A. Ulanovskii, *Universal sampling and interpolation of band-limited signals*, Geom. Funct. Anal. 18, 1029-1052, 2008.