Physics – Uspekhi **44** (5) 447 – 478 (2001)

REVIEWS OF TOPICAL PROBLEMS

PACS numbers: **02.60. – x**, **02.70. – c**, **05.40. – a**, **87.57. – s**

Wavelets and their uses

I M Dremin, O V Ivanov, V A Nechitaĭlo

DOI: 10.1070/PU2001v044n05ABEH000918

Contents

Abstract. This review paper is intended to give a useful guide for those who want to apply the discrete wavelet transform in practice. The notion of wavelets and their use in practical computing and various applications are briefly described, but rigorous proofs of mathematical statements are omitted, and the reader is just referred to the corresponding literature. The multiresolution analysis and fast wavelet transform have become a standard procedure for dealing with discrete wavelets. The proper choice of a wavelet and use of nonstandard matrix multiplication are often crucial for the achievement of a goal. Analysis of various functions with the help of wavelets allows one to reveal fractal structures, singularities etc. The wavelet transform of operator expressions helps solve some equations. In practical applications one often deals with the discretized functions, and the problem of stability of the wavelet transform and corresponding numerical algorithms becomes important. After discussing all these topics we turn to practical applications of the wavelet machinery. They are so numerous that we have to limit ourselves to a few examples only. The authors would be grateful for any comments which would move us closer to the goal proclaimed in the first phrase of the abstract.

I M Dremin, O V Ivanov, V A Nechitaĭlo Lebedev Physical Institute, Leninskiĭ pros. 53, 117924 Moscow, Russian Federation
Tel. (7-095) 132-29 29. Fax (7-095) 135-85 53
E-mail: dremin@lpi.ru

Received 13 December 2000
Uspekhi Fizicheskikh Nauk **171** (5) 465 – 501 (2001)
Translated by I M Dremin; edited by M S Aksent'eva

1. Introduction

Wavelets have become a necessary mathematical tool in many investigations. They are used in those cases when the result of the analysis of a particular signal [1] should contain not only a list of its typical frequencies (scales) but also knowledge of the definite local coordinates where these properties are important. Thus, analysis and processing of different classes of nonstationary (in time) or inhomogeneous (in space) signals is the main field of application of wavelet analysis. The most general principle of wavelet basis construction is to use dilations and translations. Any of the commonly used wavelets generates a complete orthonormal system of functions with a finite support constructed in such a way. That is why by changing the scale (dilations) they can distinguish the local characteristics of a signal at various scales, and by translations they cover the whole region in which it is studied. Due to the completeness of the system, they also

[1] The notion of a signal is used here for any ordered set of numerically recorded information about some processes, objects, functions etc. The signal can be a function of some coordinates — time, space or any other (in general, *n*-dimensional) scale. By the 'analysis' of a signal we imply not only its mathematical (in particular, wavelet) transform but also use the of such a transform to obtain some conclusions about specific features of the corresponding process or object.

allow for the inverse transformation to be done. In the analysis of nonstationary signals, the locality property of wavelets gives a substantial advantage over the Fourier transform which provides us only with knowledge of the global frequencies (scales) of the object under investigation because the system of the basic functions used (sine, cosine or imaginary exponential functions) is defined over an infinite interval[2]. However, as we shall see, the more general definitions and, correspondingly, a variety of forms of wavelets are used which admit a wider class of functions to be considered. According to Y Meyer [1], "the wavelet bases are universally applicable: 'everything that comes to hand', whether function or distribution, is the sum of a wavelet series and, contrary to what happens with Fourier series, the coefficients of the wavelet series translate the properties of the function or distribution simply, precisely and faithfully."

The literature devoted to wavelets is highly voluminous, and one can easily get a lot of references by sending the corresponding request to Internet web sites. Mathematical problems are treated in many monographs in detail (e.g., see [1 – 5]). Introductory courses on wavelets can be found in the books [6 – 9]. A nice review paper adapted for beginners and practical users with a demonstration of the wavelet transform of some signals was published in this journal about four years ago [10] and attracted much attention. However the continuous wavelet transform was mostly considered there whereas discrete wavelets were just briefly mentioned[3]. This choice was dictated by the fact that the continuous wavelets admit a somewhat more visual and picturesque presentation of the results of the analysis of a signal in terms of local maxima and skeleton graphs of wavelet coefficients with continuous variables.

At the same time, the main bulk of papers dealing with practical applications of wavelet analysis use discrete wavelets which will be our main concern here. This preference for discrete wavelets is related to the fact that widely used continuous wavelets are not, strictly speaking, orthonormal because the basis elements are both infinitely differentiable and exponentially decreasing at infinity which violates the orthonormalization property whereas there is no such problem for discrete wavelets. That is why discrete wavelets allow often for a more accurate transform and presentation of the analyzed signal, and, especially, for its inverse transform after the compression procedure. Moreover, they are better adapted to communication theory and practice. These comments do not imply that we insist on the use of only discrete wavelets for signal analysis. On the contrary, continuous wavelets may sometimes provide more transparent and analytical results in modeling the signal analysis than discrete wavelets.

The choice of a specific wavelet, be it discrete or continuous, depends on the analyzed signal and on the problem to be solved. Some functions are best analyzed using one method or another, and the advance of such an analysis depends on the relative simplicity of the decomposition achieved. The researcher's intuition and experience are decisive for the success. As an analogy, an example with number systems is often considered. It is a matter of tradition and convenience to choose systems with base 10, 2 or *e*. However, the Roman number system is completely excluded if one tries to use it for multiplication. At the same time, different problems can demand more or less effort both in their solution and in graphical presentation depending on the system chosen, and our intuition is important here.

Programs exploiting the wavelet transform are widely used now not only for scientific research but for commercial projects as well. Some of them have been even described in books (e.g., see [11]). At the same time, the direct transition from pure mathematics to computer programming and applications is non-trivial and often asks for an individual approach to the problem under investigation and for a specific choice of wavelets. Our main objective here is to describe in a suitable way the bridge that relates mathematical wavelet constructions to practical signal processing. Namely, the practical applications considered by A Grossman and I Morlet [12, 13] have lead to a quick development of wavelet methods by Y Meyer, I Daubechies et al.

The discrete wavelets look strange to those accustomed to analytical calculations because they cannot be represented by analytical expressions (except for the simplest one) or by solutions of some differential equations, and instead are given numerically as solutions of definite functional equations containing rescaling and translations. Moreover, in practical calculations their direct form is not even required, and only the numerical values of the coefficients of the functional equation are used. Thus the wavelet basis is defined by the iterative algorithm of the dilation and translation of a single function. This leads to a very important procedure called multiresolution analysis which gives rise to the multiscale local analysis of the signal and fast numerical algorithms. Each scale contains an independent non-overlapping set of information about the signal in the form of wavelet coefficients, which are determined from an iterative procedure called the *fast wavelet transform*. In combination, they provide its complete *analysis* and simplify the *diagnosis* of the underlying processes.

After such an analysis has been done, one can *compress* (if necessary) the resulting data by omitting some inessential part of the *encoded* information. This is done with the help of the so-called *quantization* procedure which commonly allocates different weights to various wavelet coefficients obtained. In particular, it helps erase some statistical fluctuations and, therefore, increase the role of the dynamical features of a signal. It can however falsify the diagnostic if the compression is done inappropriately. Usually, accurate compression gives rise to a substantial reduction of the required computer storage *memory* and *transmission* facilities, and, consequently, to a lower expenditure. The number of vanishing moments of wavelets is important at this stage. Unfortunately, the compression introduces unavoidable systematic errors. The mistakes one has made will consist of multiples of the deleted wavelet coefficients, and, therefore, the regularity properties of a signal play an essential role. Reconstruction after such compression schemes is then no longer perfect. These two objectives are clearly antagonistic. Nevertheless, when one tries to reconstruct the initial signal, the inverse transformation (*synthesis*) happens to be rather stable and reproduces its most important characteristics if proper methods are applied. The regularity properties of wavelets used also become crucial at the reconstruction stage. The

[2] A comparison of the wavelet transform with the so-called windowed (within a finite interval) Fourier transform will be briefly discussed below.

[3] For short, in what follows we use the terms 'discrete' and 'continuous' wavelets implying 'the wavelets used for discrete and continuous transform', correspondingly. Even though the 'discrete' wavelets can be represented by rather smooth functions, this terminology should not lead to confusion because we shall also use the term 'continuous wavelets' only with the above mentioned meaning.

distortions of the reconstructed signal due to quantization can be kept small, although significant compression ratios are attained. Since the part of the signal which is not reconstructed is often called noise, in essence, what we are doing is denoising the signals. Namely at this stage the superiority of the discrete wavelets becomes especially clear.

Thus, the objectives of signal processing consist in accurate analysis with help of the transform, effective coding, fast transmission and, finally, careful reconstruction (at the transmission destination point) of the initial signal. Sometimes the first stage of signal analysis and diagnosis is enough for the problem to be solved and the anticipated goals to be achieved.

It has been proven that any function can be written as a superposition of wavelets, and there exists a numerically stable algorithm to compute the coefficients for such an expansion. Moreover, these coefficients completely characterize the function, and it is possible to reconstruct it in a numerically stable way by knowing these coefficients. Because of their unique properties, wavelets have been used in functional analysis in mathematics, in studies of (multi)fractal properties, singularities and local oscillations of functions, for solving some differential equations, for the investigation of inhomogeneous processes involving widely different scales of interacting perturbations, for pattern recognition, for image and sound compression, for digital geometry processing, for solving many problems of physics, biology, medicine, engineering etc (see the recently published books [11, 14–17]). This list is by no means exhaustive.

One should however stress that, even though this method is very powerful, the goals of wavelet analysis are rather modest. It helps us describe and reveal some features, in particular, symmetries, otherwise hidden in a signal, but it does not pretend to explain the underlying dynamics and physical origin although it may give some crucial hints to it. Wavelets present a new stage in optimization of this description providing, in many cases, the best known representation of a signal. With the help of wavelets, we merely see things a little more clearly. To understand the dynamics, standard approaches introduce models assumed to be driving the mechanisms generating the observations. To define the optimality of the algorithms of the wavelet transform, some (still debatable!) energy and entropy criteria have been developed. They are internal to the algorithm itself. However, the choice of the best algorithm is also tied to the objective goal of its practical use, i.e., to some external criteria. That is why in practical applications one should submit the performance of a 'theoretically optimal algorithm' to the judgements of experts and users to estimate its benefit over the previously developed ones.

Despite very active research and impressive results, the versatility of wavelet analysis implies that these studies are presumably not in their final form yet. We shall try to describe the situation in its *status nascendi*.

The main part of this paper (Sections 2–14) is devoted to the description of the general properties of wavelets and the use of the wavelet transform in computer calculations. Some applications to different fields are briefly described in Section 15.

2. Wavelets for beginners

Each signal can be characterized by its averaged (over some intervals) values (trend) and by its variations around this trend. Let us call these variations *fluctuations* independently of their nature, be they dynamic, stochastic, psychological, physiological or of any other origin. When processing a signal, one is interested in its fluctuations at various scales because from these one can learn about their origin. The goal of wavelet analysis is to provide tools for such processing.

Actually, physicists dealing with experimental histograms analyze their data at different scales when averaging over different size intervals. This is a particular example of a simplified wavelet analysis treated in this section. To be more definite, let us consider the situation when an experimentalist measures some function $f(x)$ within the interval $0 \leqslant x \leqslant 1$, and the best resolution obtained with the measuring device is limited to 1/16th of the whole interval. Thus the result consists of 16 numbers representing the mean values of $f(x)$ in each of these bins and can be plotted as a 16-bin histogram shown in the upper part of Fig. 1. It can be represented by the following formula

$$f(x) = \sum_{k=0}^{15} s_{4,k}\varphi_{4,k}(x)\,, \tag{2.1}$$

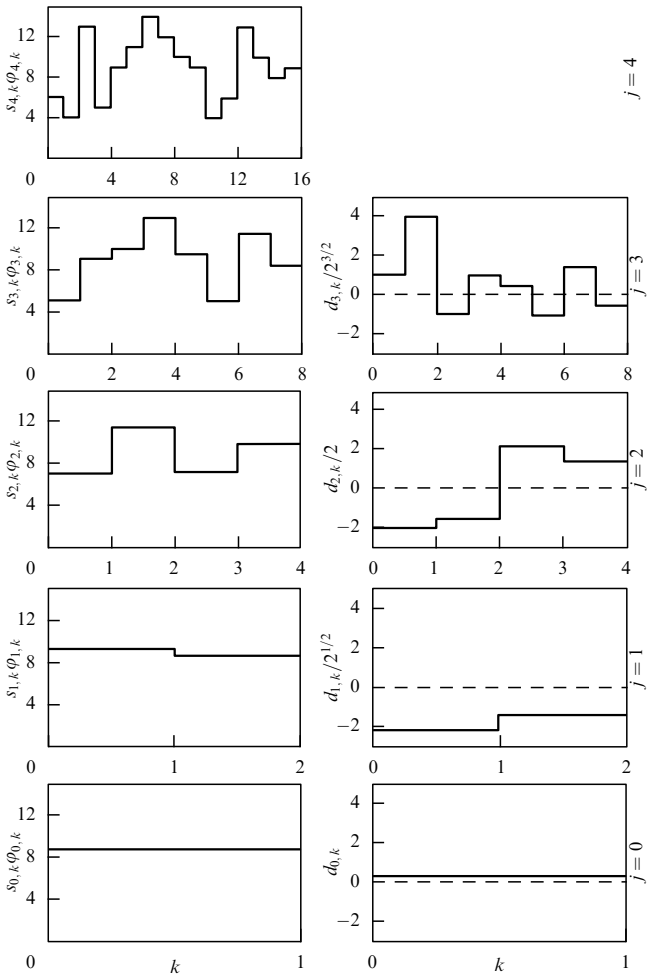

Figure 1. Histogram and its wavelet decomposition. The initial histogram is shown in the upper part of the figure. It corresponds to the level $j = 4$ with 16 bins [Eqn (2.1)]. The intervals are labeled on the vertical axis on their left-hand sides. The next level $j = 3$ is shown below. The mean values over two neighboring intervals of the previous level are shown on the left-hand side. They correspond to eight terms in the first sum in Eqn (2.4). On the right-hand side, the wavelet coefficients $d_{3,k}$ are shown. Other graphs for the levels $j = 2, 1, 0$ are obtained in a similar way.

where $s_{4,k} = f(k/16)/4$, and $\varphi_{4,k}$ is defined as a step-like block of the unit norm (i.e. of height 4 and width 1/16) different from zero only within the k-th bin. For an arbitrary j, one imposes the condition $\int |\varphi_{j,k}|^2 \, dx = 1$, where the integral is taken over the intervals of the lengths $\Delta x_j = 1/2^j$ and, therefore, $\varphi_{j,k}$ have the following form $\varphi_{j,k} = 2^{j/2}\varphi(2^j x - k)$ with φ denoting a step-like function of the unit height over such an interval. The label 4 is related to the total number of such intervals in our example. At the next coarser level the average over the two neighboring bins is taken as is depicted in the histogram just below the initial one in Fig. 1. Up to the normalization factor, we will denote it as $s_{3,k}$ and the difference between the two levels shown to the right of this histogram as $d_{3,k}$. To be more explicit, let us write down the normalized sums and differences for an arbitrary level j as

$$s_{j-1,k} = \frac{1}{\sqrt{2}}\left[s_{j,2k} + s_{j,2k+1}\right], \quad d_{j-1,k} = \frac{1}{\sqrt{2}}\left[s_{j,2k} - s_{j,2k+1}\right], \tag{2.2}$$

or for the backward transform (synthesis)

$$s_{j,2k} = \frac{1}{\sqrt{2}}\left(s_{j-1,k} + d_{j-1,k}\right), \quad s_{j,2k+1} = \frac{1}{\sqrt{2}}\left(s_{j-1,k} - d_{j-1,k}\right). \tag{2.3}$$

Since, for the dyadic partition considered, this difference has opposite signs in the neighboring bins of the previous fine level, we introduce the function ψ which is 1 and -1, correspondingly, in these bins and the normalized functions $\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$. This allows us to represent the same function $f(x)$ as

$$f(x) = \sum_{k=0}^{7} s_{3,k}\varphi_{3,k}(x) + \sum_{k=0}^{7} d_{3,k}\psi_{3,k}(x). \tag{2.4}$$

One proceeds further in the same manner to the sparser levels 2, 1 and 0 with averaging done over the interval lengths 1/4, 1/2 and 1, correspondingly. This is shown in the subsequent drawings in Fig. 1. The most sparse level with the mean value of f over the whole interval denoted as $s_{0,0}$ provides

$$f(x) = s_{0,0}\varphi_{0,0}(x) + d_{0,0}(x)\psi_{0,0}(x) + \sum_{k=0}^{1} d_{1,k}\psi_{1,k}(x) + \sum_{k=0}^{3} d_{2,k}\psi_{2,k}(x) + \sum_{k=0}^{7} d_{3,k}\psi_{3,k}(x). \tag{2.5}$$

The functions $\varphi_{0,0}(x)$ and $\psi_{0,0}(x)$ are shown in Fig. 2. The functions $\varphi_{j,k}(x)$ and $\psi_{j,k}(x)$ are normalized by the conservation of the norm, dilated and translated versions of them. In the next section we will give explicit formulae for them in a particular case of Haar scaling functions and wavelets. In practical signal processing, these functions (and more sophisticated versions of them) are often called low and high-pass filters, correspondingly, because they filter the large and small scale components of a signal. The subsequent terms in Eqn (2.5) show the fluctuations (differences $d_{j,k}$) at finer and finer levels with larger j. In all the cases (2.1)–(2.5) one needs exactly 16 coefficients to represent the function. In general, there are 2^j coefficients $s_{j,k}$ and $2^{j_n} - 2^j$ coefficients $d_{j,k}$, where j_n denotes the finest resolution level (in the above example, $j_n = 4$).

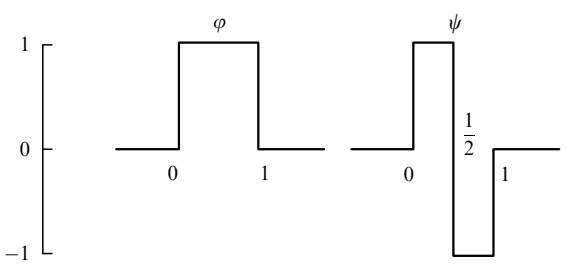

Figure 2. Haar scaling function $\varphi(x) \equiv \varphi_{0,0}(x)$ and 'mother' wavelet $\psi(x) \equiv \psi_{0,0}(x)$.

All the above representations of the function $f(x)$ [Eqns (2.1)–(2.5)] are mathematically equivalent. However, the latter one representing the wavelet analyzed function directly reveals the fluctuation structure of the signal at different scales j and various locations k present in a set of coefficients $d_{j,k}$ whereas the original form (2.1) hides the fluctuation patterns in the background of a general trend. The final form (2.5) contains the overall average of the signal depicted by $s_{0,0}$ and all its fluctuations with their scales and positions well labeled by 15 normalized coefficients $d_{j,k}$ while the initial histogram shows only the normalized average values $s_{j,k}$ in the 16 bins studied. Moreover, in practical applications the latter wavelet representation is preferred because for rather smooth functions, strongly varying only at some discrete values of their arguments, many of the high-resolution d-coefficients in relations similar to Eqn (2.5) are close to zero (compared to the 'informative' d-coefficients) and can be discarded. Bands of zeros (or close to zero values) indicate those regions where the function is fairly smooth.

At first sight, this simplified example looks somewhat trivial. However, for more complicated functions and more data points with some elaborate forms of wavelets it leads to a detailed analysis of the signal and to possible strong compression with subsequent good quality restoration. This example also provides an illustration of the very important feature of the whole approach with successive coarser and coarser approximations to f called the *multiresolution analysis* which is discussed in more detail below.

3. Basic notions and Haar wavelets

To analyze any signal, one should, first of all, choose the corresponding basis, i.e., the set of functions to be considered as 'functional coordinates'. In most cases we will deal with signals represented by the square integrable functions defined on the real axis (or by the square summable sequences of complex numbers). They form the infinite-dimensional Hilbert space $L^2(R)$ ($l^2(Z)$). The scalar product of these functions is defined as

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\,\bar{g}(x)\,dx, \tag{3.1}$$

where the bar stands for the complex conjugate. For finite sets it becomes a finite-dimensional Hilbert space. Hilbert spaces always have orthonormal bases, i.e., families of vectors (or functions) e_n such that

$$\langle e_n, e_m\rangle = \delta_{nm}, \tag{3.2}$$

$$||f||^2 = \int |f(x)|^2\,dx = \sum_n \left|\langle f, e_n\rangle\right|^2. \tag{3.3}$$

In the Hilbert space, there exist some more general families of linear independent basis vectors called a Riesz basis which generalize equation (3.3) to

$$\alpha||f||^2 \leqslant \sum_n \left|\langle f, e_n\rangle\right|^2 \leqslant \beta||f||^2 \tag{3.4}$$

with $\alpha > 0$, $\beta < \infty$. It is an unconditional basis where the order in which the basis vectors are considered does not matter. Any bound operator with a bound inverse maps an orthonormal basis into a Riesz basis.

Sometimes we will consider the spaces $L^p(R)$ $(1 \leqslant p < \infty;$ $p \neq 2)$, where the norm is defined as

$$||f||_{L^p} = \left[\int |f(x)|^p \,\mathrm{d}x\right]^{1/p}, \tag{3.5}$$

as well as some other Banach spaces [4] (see Sections 11, 12).

The Fourier transform with its trigonometric basis is well suited for the analysis of stationary signals. Then the norm $||f||$ is often called energy. For nonstationary signals, e.g., the location of that moment when the frequency characteristics has abruptly been changed is crucial. Therefore the basis should have a compact support. The wavelets are just such functions which span the whole space by translation of the dilated versions of a definite function. That is why every signal can be decomposed into a wavelet series (or integral). Each frequency component is studied with a resolution matched to its scale. The above procedure of normalization of functions $\varphi_{j,k}$ is directly connected with the requirement of conservation of the norm of a signal at its decompositions.

The choice of the analyzing wavelet is, however, not unique. One should choose it in accordance with the problem to be solved. The simplicity of operations (computing, in particular) and of representation (minimum parameters used) also plays a crucial role. A bad choice of a particular wavelet shape may even prevent one from getting any result as in the above example with Roman numbers. There are several methods of estimating how well the chosen function is applicable to the solution of a particular problem (see Section 6).

Let us try to construct functions satisfying the above criteria. An educated guess would be to relate the function $\varphi(x)$ to its dilated and translated version. The simplest linear relation with $2M$ coefficients is

$$\varphi(x) = \sqrt{2}\sum_{k=0}^{2M-1} h_k\,\varphi(2x-k) \tag{3.6}$$

with the dyadic dilation 2 and integer translation k. At first sight, the chosen normalization of the coefficients h_k with the 'extracted' factor $\sqrt{2}$ looks somewhat arbitrary. Actually, it is defined *a posteriori* by the traditional form of fast algorithms for their calculation [see Eqns (5.2) and (5.3) below] and normalization of functions $\varphi_{j,k}(x), \psi_{j,k}(x)$. It is used in all the books cited above. However, sometimes (see [2], Chapter 7) it is replaced by $c_k = \sqrt{2}h_k$.

For discrete values of the dilation and translation parameters one gets discrete wavelets. The value of the dilation factor determines the size of cells in the lattice chosen. The integer M defines the number of coefficients and the length of the wavelet support. They are interrelated because from the definition of h_k for orthonormal bases

$$h_k = \sqrt{2}\int \varphi(x)\bar{\varphi}(2x-k)\,\mathrm{d}x \tag{3.7}$$

it follows that only finitely many h_k are nonzero if φ has a finite support. The normalization condition is chosen as

$$\int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x = 1\,. \tag{3.8}$$

The function $\varphi(x)$ obtained from the solution of this equation is called a *scaling function* [5]. If the scaling function is known, one can form a 'mother wavelet' (or a basic wavelet) $\psi(x)$ according to

$$\psi(x) = \sqrt{2}\sum_{k=0}^{2M-1} g_k\,\varphi(2x-k)\,, \tag{3.9}$$

where

$$g_k = (-1)^k h_{2M-k-1}\,. \tag{3.10}$$

The simplest example would be for $M = 1$ with two nonzero coefficients h_k equal to $1/\sqrt{2}$, i.e., the equation leading to the Haar scaling function $\varphi_{\mathrm{H}}(x)$:

$$\varphi_{\mathrm{H}}(x) = \varphi_{\mathrm{H}}(2x) + \varphi_{\mathrm{H}}(2x-1)\,. \tag{3.11}$$

One easily gets the solution of this functional equation

$$\varphi_{\mathrm{H}}(x) = \theta(x)\,\theta(1-x)\,, \tag{3.12}$$

where $\theta(x)$ is the Heaviside step-function equal to 1 for positive arguments and 0 for negative ones. The additional boundary condition is $\varphi_{\mathrm{H}}(0) = 1$, $\varphi_{\mathrm{H}}(1) = 0$. This condition is important for the simplicity of the whole procedure of computing the wavelet coefficients when two neighboring intervals are considered.

The 'mother wavelet' is

$$\psi_{\mathrm{H}}(x) = \theta(x)\,\theta(1-2x) - \theta(2x-1)\,\theta(1-x) \tag{3.13}$$

with boundary values defined as $\psi_{\mathrm{H}}(0) = 1$, $\psi_{\mathrm{H}}(1/2) = -1$, $\psi_{\mathrm{H}}(1) = 0$. This is the *Haar wavelet* [18] known since 1910 and used in the functional analysis . Namely this example was considered in the previous section for the histogram decomposition. Both the scaling function $\varphi_{\mathrm{H}}(x)$ and the 'mother wavelet' $\psi_{\mathrm{H}}(x)$ are shown in Fig. 2. It is the first one of a family of compactly supported orthonormal wavelets ${}_M\psi\colon \psi_{\mathrm{H}} = {}_1\psi$. It possesses the locality property since its support $2M - 1 = 1$ is compact.

The dilated and translated versions of the scaling function φ and the 'mother wavelet' ψ

$$\varphi_{j,k} = 2^{j/2}\varphi(2^j x - k)\,, \tag{3.14}$$

$$\psi_{j,k} = 2^{j/2}\psi(2^j x - k) \tag{3.15}$$

[4] These are linear spaces with a norm which generally is not derived from a scalar product and complete with respect to this norm [$L^p(R)$ $(1 \leqslant p < \infty;$ $p \neq 2)$ is a particular example of it].

[5] It is often also called a 'father wavelet ' but we will not use this term.

form the orthonormal basis as can be (easily for Haar wavelets) checked [6]. The choice of 2^j with the integer valued j as a scaling factor leads to the unique and selfconsistent procedure of computing the wavelet coefficients. In principle, there exists an algorithm of derivation of the compact supported wavelets with an arbitrary rational number in place of 2. However, only for this factor, it has been shown that there exists an explicit algorithm with the regularity of the wavelet increasing linearly with its support. For example, for the factor 3 the regularity index only grows logarithmically. The factor 2 is probably distinguished here as in music where octaves play a crucial role. If the dilation factor is 2, then the Fourier transform of the 'mother' wavelet is essentially localized between π and 2π. However, for some practical applications, a sharper frequency localization is necessary, and it may be useful to have wavelet bases with a narrower bandwidth. The fractional dilation wavelet bases provide one of the solutions of this problem but there also exist other possibilities.

The Haar wavelet oscillates so that

$$\int_{-\infty}^{\infty} \psi(x)\,\mathrm{d}x = 0\,. \tag{3.16}$$

This condition is common for all the wavelets. It is called the *oscillation* or *cancellation condition*. From it, the origin of the name wavelet becomes clear. One can describe a 'wavelet' as a function that oscillates within some interval like a wave but is then localized by damping outside this interval. This is a necessary condition for wavelets to form an unconditional (stable) basis. We conclude that for special choices of coefficients h_k one gets the specific forms of 'mother' wavelets, which give rise to orthonormal bases.

One may decompose any function f of $L^2(R)$ at any resolution level j_n into a series

$$f = \sum_k s_{j_n,k}\,\varphi_{j_n,k} + \sum_{j \geqslant j_n,k} d_{j,k}\,\psi_{j,k}\,. \tag{3.17}$$

At the finest resolution level $j_n = j_{\max}$ only s-coefficients are left, and one gets the scaling-function representation

$$f(x) = \sum_k s_{j_{\max},k}\,\varphi_{j_{\max},k}\,. \tag{3.18}$$

In the case of the Haar wavelets it corresponds to the initial experimental histogram with the finest resolution. Since we will be interested in its analysis at varying resolutions, this form is used as an initial input only. The final representation of the same data (3.17) shows all the fluctuations in the signal. The wavelet coefficients $s_{j,k}$ and $d_{j,k}$ can be calculated as

$$s_{j,k} = \int f(x)\varphi_{j,k}(x)\,\mathrm{d}x\,, \tag{3.19}$$

$$d_{j,k} = \int f(x)\psi_{j,k}(x)\,\mathrm{d}x\,. \tag{3.20}$$

However, in practice their values are determined from the fast wavelet transform described below.

In reference to the particular case of the Haar wavelet, considered above, these coefficients are often referred to as sums (s) and differences (d), thus related to mean values and fluctuations.

For physicists familiar with experimental histograms, it generalizes the particular example discussed in the previous section. The first sum in (3.17) with the scaling functions $\varphi_{j,k}$ shows the average [7] values of f within the dyadic intervals $[k2^{-j}, (k+1)2^{-j})$, and the second term contains all the fluctuations of the function f in this interval. They come from ever smaller intervals which correspond to larger values of the scale parameter j. One would say that it 'brings into focus' the finer details of a signal. This touching in of details is regularly spaced, taking account of dimension — the details of dimension 2^{-j} are placed at the points $k2^{-j}$. At the lowest (most sparse) level j_0 the former sum consists of a single term with the overall weighted average $\langle f\rangle = s_{j_0,k_0}$, where k_0 is the center of the histogram. The second sum in (3.17) shows fluctuations at all the scales. At the next, more refined level $j_1 > j_0$, there are two terms in the first sum which show the average values of f within half-intervals with their centers positioned at k_1, k_2. The number of terms in the second sum becomes less by one term which was previously responsible for the fluctuations at the half-interval scale. The total number of terms in the expansion stays unchanged. Here we just mention that according to (3.17) the number of terms in each sum depends on a definite resolution level. Changing the level index by 1, we move some terms to another sum, and each of these representations are 'true' representations of the histogram at different resolution levels.

Formally a similar procedure may be done the other way round by going to the sparser resolutions $j < j_0$. Even if we 'fill out' the whole support of f, we can still keep going with our averaging trick. Then the average value of f diminishes, and one can neglect the first sum in (3.17) in the L^2 sense because its L^2-norm (3.3) tends to zero. In the histogram example, it decreases as $|\langle f\rangle| \propto N^{-1}$ and $|\langle f\rangle|^2 \propto N^{-2}$ whereas the integration region is proportional to N, i.e.,

$$||\langle f\rangle||^2 \propto N^{-1} \to 0 \quad \text{for} \quad N \to \infty\,.$$

That is why often only the second term in (3.17) is considered, and the result is often called the wavelet expansion. It also works if f is in the space $L^p(R)$ for $1 < p < \infty$ but it can not be done if it belongs to $L^1(R)$ or $L^\infty(R)$. For example, if $f = 1$ identically, all wavelet coefficients $d_{j,k}$ are zero, and only the first sum matters. For the histogram interpretation, the neglect of this sum would imply that one is not interested in its average value but only in its shape determined by fluctuations at different scales. Any function can be approximated to a precision $2^{j/2}$ (i.e., to an arbitrary high precision at $j \to -\infty$) by a finite linear combination of Haar wavelets.

The Haar wavelets are also suitable for the studies of functions belonging to L^p spaces, i.e., possessing higher moments.

Though the Haar wavelets provide a good tutorial example of an orthonormal basis, they suffer from several deficiencies. One of them is the bad analytic behavior with the abrupt change at the interval bounds, i.e., its bad regularity properties. By this we mean that all finite rank moments of the Haar wavelet are different from zero — only its zeroth moment, i.e., the integral (3.16) of the function itself is zero. It shows that this wavelet is not orthogonal to any polynomial

[6] We return to the general case and therefore omit the index H because the same formula will be used for other wavelets.

[7] The averaging is done with the weight functions $\varphi_{j,k}(x)$.

apart from a trivial constant. The Haar wavelet does not have good time-frequency localization. Its Fourier transform decays like $|\omega|^{-1}$ for $\omega \to \infty$.

It would be desirable to build up wavelets with better regularity. The advantage of them compared to the Haar system shows up in the smaller number of wavelet coefficients which are sufficient to account for and in their applicability to a wider set of functional spaces besides L^2. The former feature is related to the fact that the wavelet coefficients are significantly different from zero only near singularities of f (strong fluctuations!). Therefore wavelet series of standard functions with isolated singularities are 'sparse' series in contrast to the Fourier series which are usually dense ones for rather regular functions. The latter feature allows us to get access to local and global regularities of the functions under investigation. The way to this program was opened by multiresolution analysis .

4. Multiresolution analysis and Daubechies wavelets

Relation (3.17) shows that a general function f can be approximated by a sequence of very simple functions $\varphi_{j,k}$, $\psi_{j,k}$. The above example has demonstrated that the Haar functions are local and cover the whole space $L^2(R)$ using the translation k. They are orthogonal for different resolution scales j. The transition from j to $j+1$ is equivalent to the replacement of x by $2x$, i.e., to the rescaling which allows for the analysis to be done at various resolutions.

However the Haar wavelets are oversimplified and not regular enough. The goal is to find a general class of those functions which would satisfy the requirements of locality, regularity and oscillatory behavior. Note that in some particular cases the orthonormality property sometimes can be relaxed. They should be simple enough in the sense that they are sufficiently explicit and regular to be completely determined by their samples on the lattice defined by the factors 2^j.

The general approach which respects these properties is known as the *multiresolution approximation*. A rigorous mathematical definition is given in Section 17.1. Here we just describe its main ingredients.

Multiresolution analysis consists of a sequence of successive approximation spaces V_j which are scaled and invariant under integer translation versions of the central functional space V_0. To explain the meaning of these spaces in a simple example, we show in Fig. 3 what the projections of some function on the Haar spaces V_0, V_1 might look like. One easily recognizes the histogram representation of this function. The comparison of histograms at the two levels shows that the first sum in Eqn (3.17) provides the 'blurred image' or 'smoothed means' of $f(x)$ in each interval, while the second sum of this equation adds finer and finer details of smaller sizes. Thus the general distributions are decomposed into a series of correctly localized fluctuations having a characteristic form defined by the wavelet chosen.

The functions $\varphi_{j,k}$ form an orthonormal basis in V_j. The orthogonal complement of V_j in V_{j+1} is called W_j. The subspaces W_j form a mutually orthogonal set. The sequence of $\psi_{j,k}$ constitutes an orthonormal basis for W_j at any definite j. The whole collection of $\psi_{j,k}$ and $\varphi_{j,k}$ for all j is an orthonormal basis for $L^2(R)$. This ensures us that we have constructed a multiresolution analysis approach, and the functions $\psi_{j,k}$ and $\varphi_{j,k}$ constitute the small and large scale

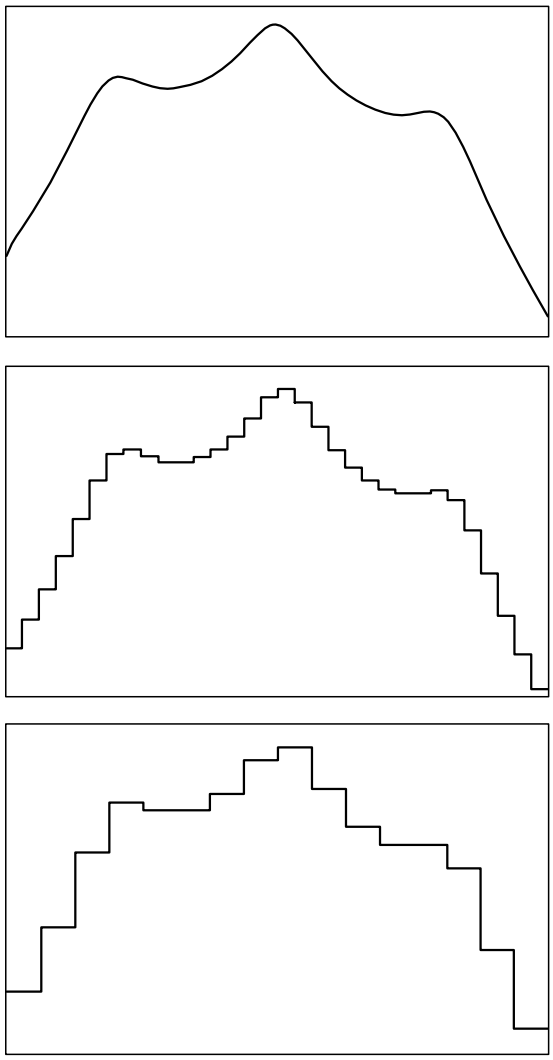

Figure 3. Analyzed function (a) and its Haar projections onto two subsequent spaces V_0 and V_1.

filters, correspondingly. The whole procedure of multiresolution analysis is demonstrated in the graphs of Fig. 4.

In accordance with the above declared goal, one can define the notion of wavelets (see Section 17.1) so that the functions $2^{j/2}\psi(2^j x - k)$ are the wavelets (generated by the 'mother' ψ), possessing the regularity, the localization and the oscillation properties.

At first sight, from our example with Haar wavelets, it looked as if one is allowed to choose the coefficients h_k freely. This impression is, however, completely wrong. The general properties of scaling functions and wavelets define these coefficients in a unique way in the framework of the multiresolution analysis approach.

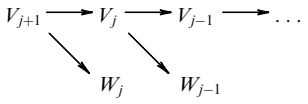

Figure 4. Graphical representation of multiresolution analysis with decomposition of V_{j+1} space into its subspace V_j and the orthogonal complement W_j iterated to the lower levels.

Let us show how the program of the multiresolution analysis works in practice when applied to the problem of finding out the coefficients of any filter h_k and g_k. They can be directly obtained from the definition and properties of the discrete wavelets. These coefficients are defined by relations (3.6) and (3.9)

$$\varphi(x)=\sqrt{2}\sum_k h_k\varphi(2x-k)\,,\quad \psi(x)=\sqrt{2}\sum_k g_k\varphi(2x-k)\,, \tag{4.1}$$

where $\sum_k |h_k|^2<\infty$. The orthogonality of the scaling functions defined by the relation

$$\int \varphi(x)\,\varphi(x-m)\,\mathrm{d}x=\delta_{0m} \tag{4.2}$$

leads to the following equation for the coefficients:

$$\sum_k h_k\,h_{k+2m}=\delta_{0m}\,. \tag{4.3}$$

The orthogonality of wavelets to the scaling functions

$$\int \psi(x)\varphi(x-m)\,\mathrm{d}x=0 \tag{4.4}$$

gives the equation

$$\sum_k h_k\,g_{k+2m}=0\,, \tag{4.5}$$

having a solution of the form

$$g_k=(-1)^k h_{2M-1-k}\,. \tag{4.6}$$

Thus the coefficients g_k for wavelets are directly defined by the scaling function coefficients h_k.

Another condition of the orthogonality of wavelets to all polynomials up to the power $(M-1)$, defining its regularity and oscillatory behavior

$$\int x^n\psi(x)\,\mathrm{d}x=0\,,\qquad n=0,\ldots,M-1\,, \tag{4.7}$$

provides the relation

$$\sum_k k^n g_k=0\,, \tag{4.8}$$

giving rise to

$$\sum_k (-1)^k\,k^n\,h_k=0\,. \tag{4.9}$$

when the formula (4.6) is taken into account.

The normalization condition

$$\int \varphi(x)\,\mathrm{d}x=1 \tag{4.10}$$

can be rewritten as another equation for h_k:

$$\sum_k h_k=\sqrt{2}\,. \tag{4.11}$$

Let us write down equations (4.3), (4.9), (4.11) for $M=2$ explicitly:

$$\begin{aligned}
&h_0h_2+h_1h_3=0\,,\\
&h_0-h_1+h_2-h_3=0\,,\\
&-h_1+2h_2-3h_3=0\,,\\
&h_0+h_1+h_2+h_3=\sqrt{2}.
\end{aligned}$$

The solution of this system is

$$\begin{aligned}
&h_3=\frac{1}{4\sqrt{2}}\,(1\pm\sqrt{3})\,,\quad h_2=\frac{1}{2\sqrt{2}}+h_3\,,\\
&h_1=\frac{1}{\sqrt{2}}-h_3\,,\quad h_0=\frac{1}{2\sqrt{2}}-h_3\,,
\end{aligned} \tag{4.12}$$

that, in the case of the minus sign for h_3, corresponds to the well known filter

$$\begin{aligned}
&h_0=\frac{1}{4\sqrt{2}}(1+\sqrt{3})\,,\quad h_1=\frac{1}{4\sqrt{2}}(3+\sqrt{3})\,,\\
&h_2=\frac{1}{4\sqrt{2}}(3-\sqrt{3})\,,\quad h_3=\frac{1}{4\sqrt{2}}(1-\sqrt{3})\,.
\end{aligned} \tag{4.13}$$

These coefficients define the simplest D^4 (or ${}_2\psi$) wavelet from the famous family of orthonormal Daubechies wavelets with finite support. It is shown in the upper part of Fig. 5 by the dotted line with the corresponding scaling function shown by the solid line. Some other higher rank wavelets are also shown there. It is clear from this figure (especially, for D^4) that wavelets are smoother at some points than at others. The choice of the plus sign in the expression for h_3 would not change the general shapes of the scaling function and wavelet D^4. It results in their mirror symmetrical forms obtained by a simple reversal of the signs on the horizontal and vertical axes, correspondingly. However, for higher rank wavelets different choices of signs would correspond to different forms of the wavelet. After the signs are chosen, it is clear that compactly supported wavelets are unique, for a given multiresolution analysis up to a shift in the argument (translation) which is inherently there. The dilation factor must be rational within the framework of the multiresolution analysis. Let us note that ${}_2\varphi$ is Hölder continuous with the global exponent $\alpha=0.55$ [see Eqn (11.1) below] and has different local Hölder exponents on some fractal sets. Typically, wavelets are more regular at some points than at others.

For the filters of higher order in M, i.e., for higher rank Daubechies wavelets, the coefficients can be obtained in an analogous manner. It is however necessary to solve the equation of the M-th power in this case. Therefore, the numerical values of the coefficients can be found only approximately, but with any predefined accuracy. The wavelet support is equal to $2M-1$. It is wider than for the Haar wavelets. However the regularity properties are better. The higher order wavelets are smoother compared to D^4 as seen in Fig. 5. The Daubechies wavelet with M vanishing moments has μM continuous derivatives where $\mu\approx 0.2$ as was estimated numerically [8]. This means that about 70–80%

[8] This asymptotic estimate was obtained with the help of the Fourier transform [2]. More precise methods at finite values of M allow us to get the relation between the regularity and the number of vanishing moments of any function. The linear interpolation in the region of the practically used values $6<M<12$ leads to the estimate $\mu\approx 0.275$.

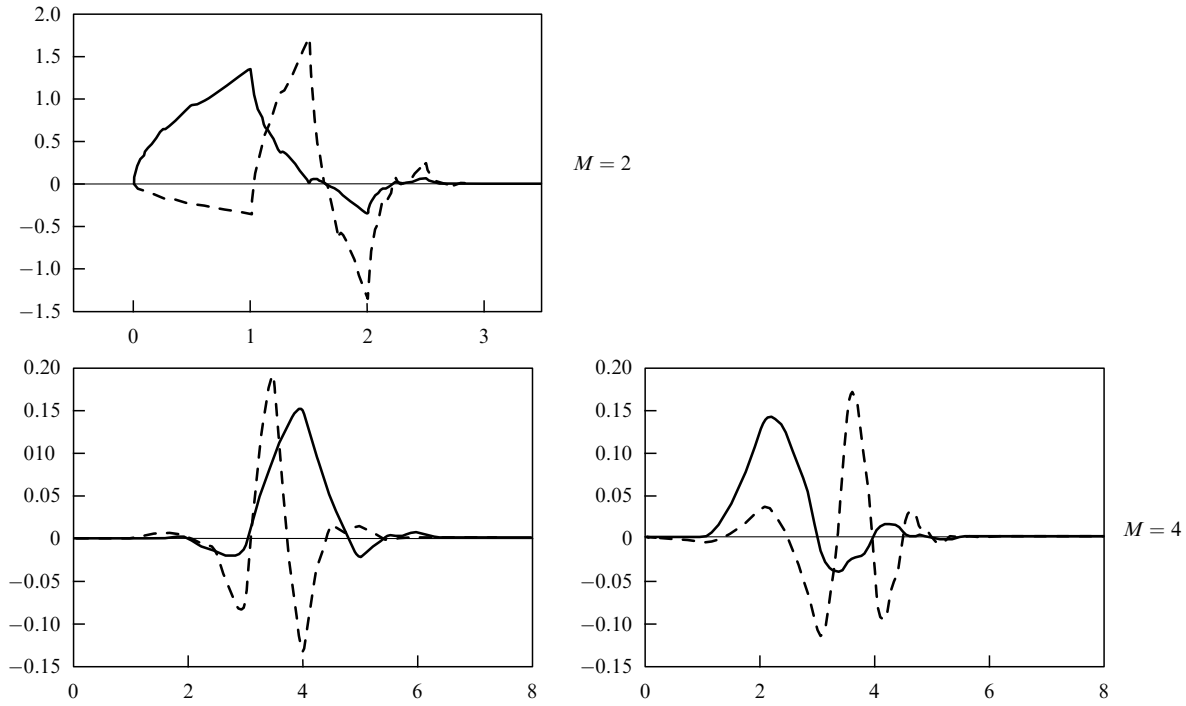

Figure 5. Daubechies scaling functions (solid lines) and wavelets (dotted lines) for $M = 2, 4$.

of zero moments are 'wasted'. As the regularity M increases, so does the support in general. For sufficiently regular functions, Daubechies wavelet coefficients are much smaller (2^{Mj} times) than the Haar wavelet coefficients, i.e., the signal can be compressed much better with Daubechies wavelets. Since they are more regular, the synthesis is also more efficient.

One can ask the question whether the regularity or the number of vanishing moments is more important. The answer depends on the application, and is not always clear. It seems that the number of vanishing moments is more important for stronger compression which increases for a larger number of vanishing moments, while the regularity can become crucial in inverse synthesis to smooth the errors due to the compression (omission of small coefficients).

In principle, by solving the functional equation (3.6) one can find the form of the scaling function and, from (3.9), the form of the corresponding 'mother wavelet'. There is no closed-form analytic formula for the compactly supported $\varphi(x)$, $\psi(x)$ (except for the Haar case). Nevertheless one can compute their plots, if they are continuous, with arbitrarily high precision using a fast cascade algorithm with the wavelet decomposition of $\varphi(x)$ which is a special case of a refinement scheme (for more details, see [2]). Instead of a refinement cascade one can compute $\varphi(2^{-j}k)$ directly from Eqn (3.9) starting from appropriate $\varphi(n)$. However, in practical calculations the above coefficients h_k are used only without referring to the shapes of the wavelets.

Except for the Haar basis, all real orthonormal wavelet bases with compact support are asymmetric, i.e., they have neither a symmetry nor an antisymmetry axis (see Fig. 5). The deviation of a wavelet from symmetry is judged by how much the phase of the expression $m_0(\omega) = \sum_k h_k \exp(-ik\omega)$ deviates from a linear function. The 'least asymmetric' wavelets are constructed by minimizing this phase. Better symmetry for a wavelet necessarily implies better symmetry for the coefficients h_k but the converse statement is not always true.

5. Fast wavelet transform and coiflets

After calculation of the coefficients h_k and g_k, i.e., the choice of a definite wavelet transform of a signal $f(x)$, one is able to perform its wavelet analysis because the wavelet orthonormal basis ($\psi_{j,k}$, $\varphi_{j,k}$) has been defined. Any function $f \in L^2(R)$ is completely characterized by the coefficients of its decomposition in this basis and may be decomposed according to formula (3.17). Let us make the sum limits in this formula more precise. The function $f(x)$ may be considered at any n-th resolution level j_n. Then the separation of its average values and fluctuations at this level looks like

$$f(x) = \sum_{k=-\infty}^{\infty} s_{j_n,k}\varphi_{j_n,k}(x) + \sum_{j=j_n}^{\infty}\sum_{k=-\infty}^{\infty} d_{j,k}\psi_{j,k}(x)\,. \qquad (5.1)$$

Over the infinite interval, the first sum may be omitted as explained above, and one gets the pure wavelet expansion. As we stressed already, the coefficients $s_{j,k}$ and $d_{j,k}$ carry information about the content of the signal at various scales. They can be calculated directly using the formulas (3.19), (3.20). However this algorithm is inconvenient for numerical computations because it requires many (N^2) operations where N denotes the number of the sampled values of the function. We will describe a faster algorithm. It is clear from Fig. 6, and the fast algorithm formulas are presented below.

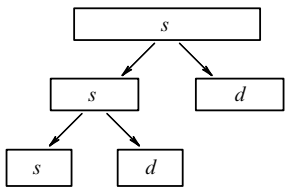

Figure 6. Fast wavelet transform algorithm.

In real situations with digitized signals, we have to deal with finite sets of points. Thus, there always exists the finest level of resolution where each interval contains only a single number. Correspondingly, the sums over k will get finite limits. It is convenient to reverse the level indexation assuming that the label of this fine scale is $j = 0$. It is then easy to compute the wavelet coefficients for more sparse resolutions $j \geqslant 1$.

Multiresolution analysis naturally leads to an hierarchical and fast scheme for the computation of the wavelet coefficients of a given function. The functional equations (3.6), (3.9) and the formulas for the wavelet coefficients (3.19), (3.20) give rise, in the case of Haar wavelets, to the relations (2.2), or for the backward transform (synthesis) to (2.3).

In general, one can get the iterative formulas of the fast wavelet transform

$$s_{j+1,k} = \sum_m h_m \, s_{j,2k+m} \,, \tag{5.2}$$

$$d_{j+1,k} = \sum_m g_m \, s_{j,2k+m} \tag{5.3}$$

where

$$s_{0,k} = \int f(x)\varphi(x-k)\,\mathrm{d}x \,. \tag{5.4}$$

These equations yield fast algorithms (the so-called pyramid algorithms) for computing the wavelet coefficients, asking now just for $O(N)$ operations to be done. Starting from $s_{0,k}$, one computes all other coefficients provided the coefficients h_m, g_m are known. The explicit shape of the wavelet is not used in this case any more. The simple form of these equations is the only justification for introducing the factor $\sqrt{2}$ into the functional equation (3.6). In principle, the coefficients h_m, g_m could be renormalized. However, in practice Eqns (5.2) and (5.3) are used much more often than others, and this normalization is kept intact. After choosing a particular wavelet for analysis, i.e., choosing h_m, g_m, one uses only Eqns (5.2) and (5.3) for computing the wavelet coefficients, and additional factors in these equations would somewhat complicate the numerical processing.

The remaining problem lies in the initial data. If an explicit expression for $f(x)$ is available, the coefficients $s_{0,k}$ may be evaluated directly according to (5.4). But this is not so in the situation when only discrete values are available. To get good accuracy, one has to choose very small bins (a dense lattice) which is often not accessible with finite steps of sampling. In such a case, the usually adopted simplest solution consists in directly using the values $f(k)$ of the available sample in place of the coefficients $s_{0,k}$ and starting the fast wavelet transform using formulas (5.2), (5.3). This is a safe operation since the pyramid algorithm yields perfect reconstruction, and the coefficient $s_{0,k}$ essentially represents a local average of the signal provided by the scaling function.

In general, one can choose

$$s_{0,k} = \sum_m c_m f(k-m) \,. \tag{5.5}$$

The above supposition $s_{0,k} = f(k)$ corresponds to $c_m = \delta_{0m}$. This supposition may be almost rigorous for some specific choices of scaling functions named coiflets after R. Coifman whose ideas inspired I Daubechies to build up these wavelets). It is possible to construct multiresolution analysis with the scaling function having vanishing moments, i.e., such that

$$\int x^m \varphi(x)\,\mathrm{d}x = 0 \,, \qquad 0 < m < M \,. \tag{5.6}$$

To construct such wavelets (coiflets), one has to add to the equations for determining the coefficients h_k a new condition

$$\sum_k h_k \, k^m = 0 \,, \qquad 0 < m < M \,, \tag{5.7}$$

which follows from the requirement (5.6).

Coiflets are more symmetrical than Daubechies wavelets as is seen from Fig. 7 if compared to Fig. 5. The latter do not have the property (5.6). The price for this extra generalization is that coiflets are longer than Daubechies wavelets. If in the latter case the length of the support is $2M - 1$, for coiflets it is equal to $3M - 1$. The error in the estimation of $s_{j,k}$ decreases with the number of vanishing moments as $O(2^{-jM})$. At the same time, the variation of the smoothness for coiflets of a given order is larger than that for Daubechies wavelets of the same order.

There are other proposals to improve the first step of the iterative procedure of the fast wavelet transform promoted by using T-polynomials by W Sweldens [19] or the so-called 'lazy' or interpolating wavelets by S Goedecker and O Ivanov [20]. The latter is most convenient for simultaneous analysis at different resolution levels, in particular for non-equidistant lattices.

Inverse fast wavelet transform allows for the reconstruction of the function starting from its wavelet coefficients.

6. Choice of wavelets

Above, we demonstrated three examples of discrete orthonormal compactly supported wavelets. The regularity property, the number of vanishing moments and the number of wavelet coefficients exceeding some threshold value were considered as possible criteria for the choice of a particular wavelet not to mention computing facilities. Sometimes the so-called information cost functional used by statisticians is introduced, and one tries to minimize it and thus select the optimal basis. In particular, the entropy criterion for the probability distribution of the wavelet coefficients is also considered [4, 11]. The entropy of f relative to the wavelet basis measures the number of significant terms in the decomposition (3.17). It is defined by $\exp(-\sum_{j,k} |d_{j,k}|^2 \log |d_{j,k}|^2)$. If we have a collection of orthonormal bases, we will choose for the analysis of f the particular basis that yields the minimum entropy.

The number of possible wavelets at our disposal is much larger than the above examples show. We will not discuss all of them, just mentioning some and referring the reader to the cited books.

• First, let us mention splines which lead to wavelets with non-compact support but with exponential decay at infinity and with some (limited) number of continuous derivatives. Special orthogonalization tricks should be used here. The splines are intrinsically associated with interpolation schemes for finding more precise initial values of $s_{0,k}$ relating them to some linear combinations of the sampled values of $f(x)$.

• To insure both the full symmetry and exact reconstruction, one has to use so-called biorthogonal wavelets. This means that two dual wavelet bases $\psi_{j,k}$ and $\tilde{\psi}_{j,k}$, associated with two different multiresolution ladders, are used. They can

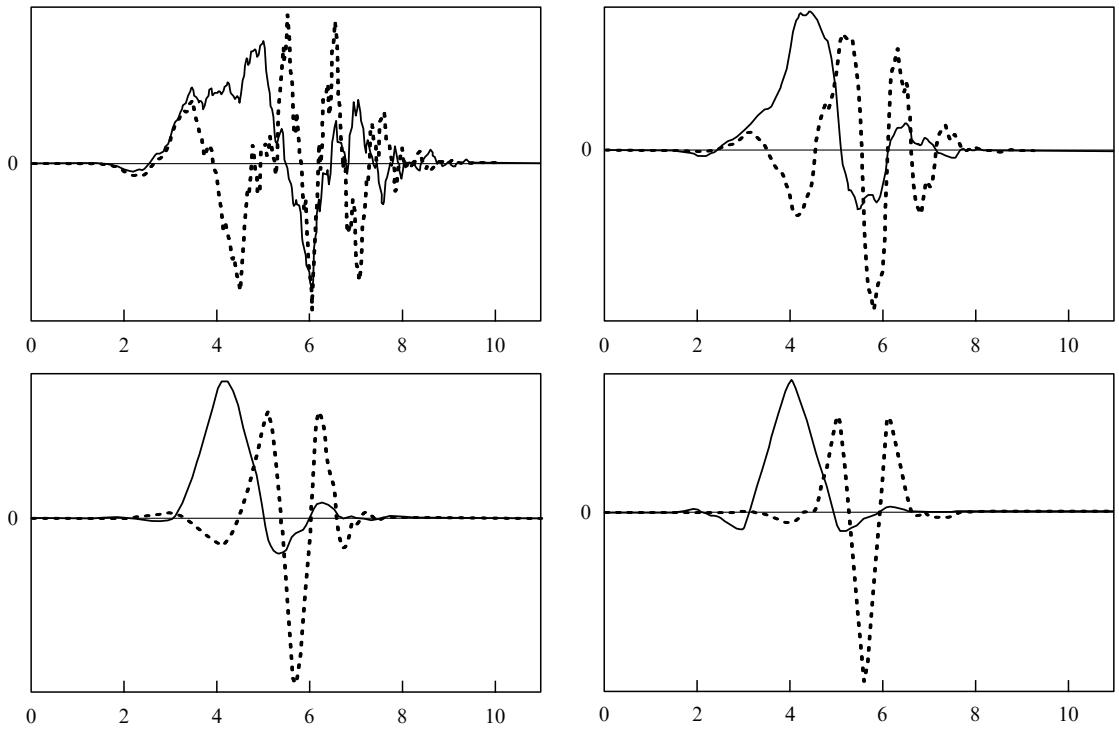

Figure 7. Coiflets (dotted lines) and their scaling functions (solid lines) for $M = 4$.

have very different regularity properties. A function f may be represented in two forms, absolutely equivalent until the compression is done:

$$f = \sum_{j,k} \langle f, \psi_{j,k} \rangle \widetilde{\psi}_{j,k} \,, \tag{6.1}$$

$$f = \sum_{j,k} \langle f, \widetilde{\psi}_{j,k} \rangle \psi_{j,k} \,, \tag{6.2}$$

where ψ and its dual wavelet satisfy the biorthogonality requirement $\langle \psi_{j,k} | \widetilde{\psi}_{j',k'} \rangle = \delta_{j,k;\,j',k'}$. In contrast to Daubechies wavelets, where regularity is tightly connected with the number of vanishing moments, biorthogonal wavelets have much more freedom. If one of them has a regularity of the order r, then its dual partner wavelet has automatically at least r vanishing moments. If $\widetilde{\psi}_{j,k}$ is much more regular than $\psi_{j,k}$, then $\psi_{j,k}$ has many more vanishing moments than $\tilde{\psi}^{j,k}$. This allows us to choose, e.g., very regular $\widetilde{\psi}_{j,k}$ and get many vanishing moments of $\psi_{j,k}$. The large number of vanishing moments of $\psi_{j,k}$ leads to better compressibility for reasonably smooth f. If compression has been done, formula (6.1) is much more useful than (6.2). The number of significant terms is much smaller, and, moreover, the better regularity of $\widetilde{\psi}_{j,k}$ helps reconstruct f more precisely. Biorthogonal bases are close to an orthonormal basis. Both wavelets can be made symmetric. Symmetric biorthogonal wavelets close to orthonormal basis are close to coiflets. The construction of biorthogonal wavelet bases is simpler than of orthonormal bases.

• The existence of two-scale relations is the main feature of the construction of wavelet packets. The general idea of the wavelet packets is to iterate further the splitting of the frequency band, still keeping the same pair of filters. The scaling function introduced above acquires the name w_0, and the packet is built up through the following iterations

$$w_{2n}(x) = \sum_k h_k w_n(2x - k) \,, \tag{6.3}$$

$$w_{2n+1}(x) = \sum_k g_k w_n(2x - k) \,. \tag{6.4}$$

The usual mother wavelet is represented by w_1. This family of wavelets forms an orthonormal basis in $L^2(R)$ which is called the fixed scale *wavelet packet basis*. Figure 8 demonstrates the whole construction.

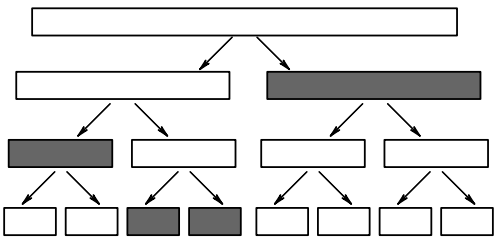

Figure 8. Wavelet packet construction.\

• One can abandon the orthonormality property and construct non-orthogonal wavelets called *frames*. An important special class of frames is given by the Riesz bases of $L^2(R)$. A Riesz basis is a frame, but the converse is not true in general. The frames satisfy the following requirement:

$$A||f||^2 \leqslant \sum_{j \in J} |\langle f, \varphi_j \rangle|^2 \leqslant B||f||^2 \,. \tag{6.5}$$

The constants A and B are called the *frame bounds*. For $A = B$ one calls them *tight frames*. The case $A = B = 1$ corresponds to *orthonormal wavelets*.

• When acting by (singular) operators, one sometimes gets infinities when usual wavelets are considered. Some suitable function $b(x)$ may be used to specify the extra conditions which will be necessary and sufficient for the result of the (singular integral) operator action to be continuous in L^2. In this case one chooses the so-called 'wavelets which are adapted to b'. Any function f is again decomposed as

$$f(x) = \sum_{\lambda} a(\lambda)\psi_{\lambda}^{(b)}(x)\,, \tag{6.6}$$

but the wavelet coefficients are now calculated according to

$$a_{\lambda} = \int b(x) f(x)\, \psi_{\lambda}^{(b)}(x)\, \mathrm{d}x\,. \tag{6.7}$$

They satisfy the normalization condition

$$\int b(x)\, \psi_{\lambda}^{(b)}(x)\, \psi_{\lambda'}^{(b)}(x)\, \mathrm{d}x = \delta_{\lambda,\lambda'}\,. \tag{6.8}$$

The cancellation condition now reads

$$\int b(x)\, \psi_{\lambda}^{(b)}(x)\, \mathrm{d}x = 0\,. \tag{6.9}$$

As we see, the cancellation is also adapted to the function b (in general, to the 'complex measure' $b(x)\,\mathrm{d}x$).

• Up to now we have considered wavelets with the dilation factor equal to 2. It is most convenient for numerical calculations. However, it can be proved [2, 21] that within the framework of a multiresolution analysis, the dilation factor must be rational, and no other special requirements are imposed. Therefore one can construct schemes with other integer or fractional dilation factors. Sometimes they may provide a sharper frequency localization. For wavelets with the dilation factor 2, their Fourier transform is essentially localized within one octave between π and 2π, whereas the fractional dilation wavelet bases may have a bandwidth narrower than one octave.

• Moreover, one can use the continuous wavelets as described at some length in Ref. [10]. The element of the wavelet basis is written in the form

$$\psi_{a,b}(x) = |a|^{-1/2}\, \psi\left(\frac{x-b}{a}\right). \tag{6.10}$$

The direct and inverse formulas of the wavelet transform look like

$$W_{a,b} = |a|^{-1/2} \int f(x)\, \psi\left(\frac{x-b}{a}\right) \mathrm{d}x\,, \tag{6.11}$$

$$f(x) = C_{\psi}^{-1} \int W_{a,b}\, \psi_{a,b}(x)\, \frac{\mathrm{d}a\, \mathrm{d}b}{a^2}\,. \tag{6.12}$$

Here

$$C_{\psi} = \int |\psi(\omega)|^2\, \frac{\mathrm{d}\omega}{|\omega|} = \int \left| \int \exp(-\mathrm{i}x\omega)\, \psi(x)\, \mathrm{d}x \right|^2 \frac{\mathrm{d}\omega}{|\omega|}\,. \tag{6.13}$$

From here one easily recognizes that the oscillation of wavelets required by Eqn (3.16) is a general property. The vanishing Fourier transform of a wavelet as $\omega \to 0$, which is just directly the condition (3.16), provides a finite value of C_{ψ} in (6.13). One of the special and often used examples of continuous wavelets is given by the second derivative of the Gaussian function which is called the Mexican hat (MHAT) wavelet after its shape. Actually, it can be considered as a special frame as shown by Daubechies. The reconstruction procedure (synthesis) is complicated and can become unstable in this case. However it is widely applied for the analysis of signals. Formula (6.11) is a kind of a convolution operation. That is why the general theory of so-called Calderon–Zygmund operators [3] (see Section 17.2) is, in particular, applicable to problems of wavelet decomposition.

7. Multidimensional wavelets

Multiresolution analysis can be performed in more than 1 dimensions. There are two ways [4] to generalize it to the two-dimensional case, for example, but we will consider the most often used construction given by tensor products. The tensor product method is a direct way to construct an r-regular multiresolution approximation which produces multidimensional wavelets of compact support. This enables us to analyze every space of functions or distributions in n dimensions whose regularity is bound by r.

The trivial way of constructing a two-dimensional orthonormal basis starting from a one-dimensional orthonormal wavelet basis $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ is simply to take the tensor product functions generated by two one-dimensional bases:

$$\Psi_{j_1,k_1;j_2,k_2}(x_1, x_2) = \psi_{j_1,k_1}(x_1)\psi_{j_2,k_2}(x_2)\,. \tag{7.1}$$

In this basis the two variables x_1 and x_2 are dilated independently.

More interesting for many applications is another construction, in which dilations of the resulting orthonormal wavelet basis control both variables simultaneously, and the two-dimensional wavelets are given by the following expression:

$$2^j \Psi(2^j x - k,\, 2^j y - l)\,, \qquad j, k, l \in Z\,, \tag{7.2}$$

where Ψ is no longer a single function: on the contrary, it consists of three elementary wavelets. To get an orthonormal basis of W_0 one has to use in this case three families

$$\varphi(x-k)\psi(y-l), \quad \psi(x-k)\varphi(y-l), \quad \psi(x-k)\psi(y-l)\,.$$

Then the two-dimensional wavelets are

$$2^j\varphi(2^j x - k)\psi(2^j y - l), \quad 2^j\psi(2^j x - k)\varphi(2^j y - l),$$

$$2^j\psi(2^j x - k)\psi(2^j y - l)\,.$$

In the two-dimensional plane, the analysis is done along the horizontal, vertical and diagonal strips with the same resolution in accordance with these three wavelets.

Figure 9 shows how this construction looks. The schematic representation of this procedure in the left-hand side of the figure demonstrates how the corresponding wavelet coefficients are distributed for different resolution levels ($j = 1$ and $j = 2$). In the figure, a set of geometrical objects is decomposed into two layers. One clearly sees how vertical, horizontal and diagonal edges are emphasized in the corresponding regions. One should also notice that the horizontal strip is resolved into two strips at a definite resolution level.

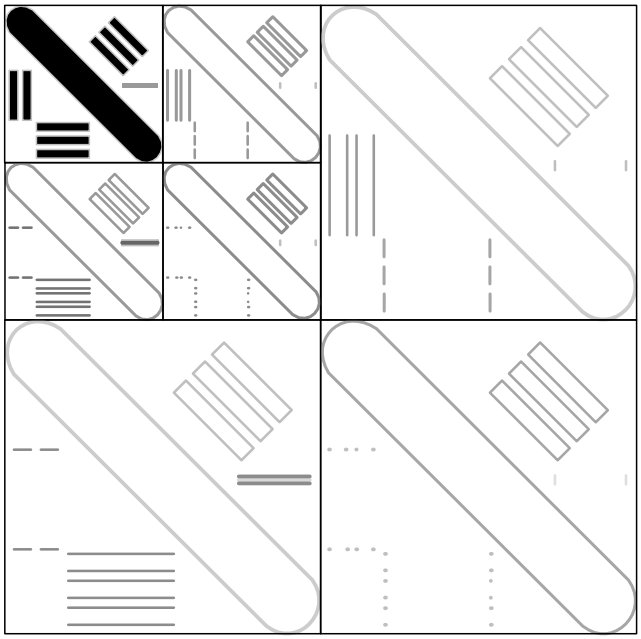

Figure 9. Example of the wavelet analysis of a two-dimensional plot. One sees that either horizontal or vertical details of the plot are more clearly resolved in the corresponding coefficients. Also, the small or large size (correlation length) details are better resolved depending on the level chosen.

In the general n-dimensional case, there exist 2^{n-1} functions which form an orthonormal basis and admit the multiresolution analysis of any function from $L^2(R^n)$. The normalization factor is equal to $2^{nj/2}$ in front of the function, as can be guessed already from the above two-dimensional case with this factor equal to 2^j in contrast to $2^{j/2}$ in one dimension.

There also exists a method to form wavelet bases which are not reducible to tensor products of one-dimensional wavelets (see [4]). In dimension 1, every orthonormal basis arises from a multiresolution approximation. In dimensions greater than 1, it is possible to form an orthonormal basis such that there is no r-regular multiresolution approximation ($r \geqslant 1$) from which these wavelets can be obtained [3].

8. The Fourier and wavelet transforms

In many applications (especially, for non-stationary signals), one is interested in the frequency content of a signal locally in time, i.e., tries to learn which frequencies are important at some particular moment. As has been stressed already, the wavelet transform is superior to the Fourier transform, first of all, due to the locality property of wavelets. The Fourier transform uses sine, cosine or imaginary exponential functions as the main basis. It is spread over the entire real axis whereas the wavelet basis is localized. It helps analyze the local properties of a signal using wavelets while the Fourier transform does not provide any information about the location where the scale (frequency) of a signal changes.

The necessity to use different functions was mentioned by L I Mandelstam [9] as early as in 1920 when he stated that "the physical meaning of the Fourier transform is to a high extent related to the resonance properties of linear systems with constant parameters; when considering linear systems with variable parameters, the Fourier expansion becomes inadequate and the functions cos and sin should be replaced by other functions".

Decomposition into wavelets allows singularities to be located by observing the places where the wavelet coefficients are (abnormally) large. Obviously, nothing of the kind happens for the Fourier transform. Once the wavelets have been constructed they perform incredibly well in situations where Fourier series and integrals involve subtle mathematics or heavy numerical calculations. But wavelet analysis cannot entirely replace Fourier analysis, indeed, the latter is often used in constructing the orthonormal bases of wavelets needed for analysis with wavelet series. Many theorems of wavelet analysis are proven with the help of the Fourier decomposition. The two kinds of analysis are thus complementary rather than always competing.

The Fourier spectrum f_ω of a one-dimensional signal $f(t)$ having finite energy (i.e., square-integrable) is given by

$$f_\omega = \int_{-\infty}^{\infty} f(t) \exp(-\mathrm{i}\omega t)\, \mathrm{d}t\,. \tag{8.1}$$

The inverse transform restores the signal

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_\omega \exp(\mathrm{i}\omega t)\, \mathrm{d}\omega\,. \tag{8.2}$$

It is an unitary transformation

$$\int |f(t)|^2\, \mathrm{d}t = \frac{1}{2\pi} \int |f_\omega|^2\, \mathrm{d}\omega\,. \tag{8.3}$$

This is the so-called *Parseval identity* which states the conservation of energy between the time and the frequency domains. Formula (8.1) asks for information about the signal $f(t)$ from both past and future times. It is suited for a stationary signal when the frequency ω does not depend on time. Thus its time-frequency band is well located in frequency and practically unlimited in time, i.e., Eqn (8.1) gives a representation of the frequency content of f but not of its time-localization properties. Moreover, the signal $f(t)$ should decrease fast enough at future and past infinities for the integral (8.1) to be meaningful.

An attempt to overcome these difficulties and improve time-localization while still using the same basis functions is made by the so-called *windowed Fourier transform*. The signal $f(t)$ is considered within some time interval (window) only. In practice, one has to restrict the search for the optimal window to the easily generated windows. A simple way of doing so amounts to multiplying $f(t)$ by a simple compactly supported window function, e.g., by $g_w = \theta(t - t_\mathrm{i})\theta(t_\mathrm{f} - t)$, where θ is the commonly used Heaviside step-function different from zero at positive values of its argument only, t_i, t_f are the initial and final cut-offs of the signal (more complicated square-integrable window functions g, well concentrated in time, e.g., Gaussian or canonical coherent states [10], can be used as well).

[9] We are grateful to E L Feĭnberg for pointing this out to us. (See: Mandel'stam L I *Collection of papers* Vol. I (Ed. S M Rytov) (Moscow: Izd. AN SSSR, 1948) p. 46.)

[10] In quantum mechanics, they are introduced for quantizing the classical harmonic oscillator. In signals they are known by the name of *Gabor functions*.

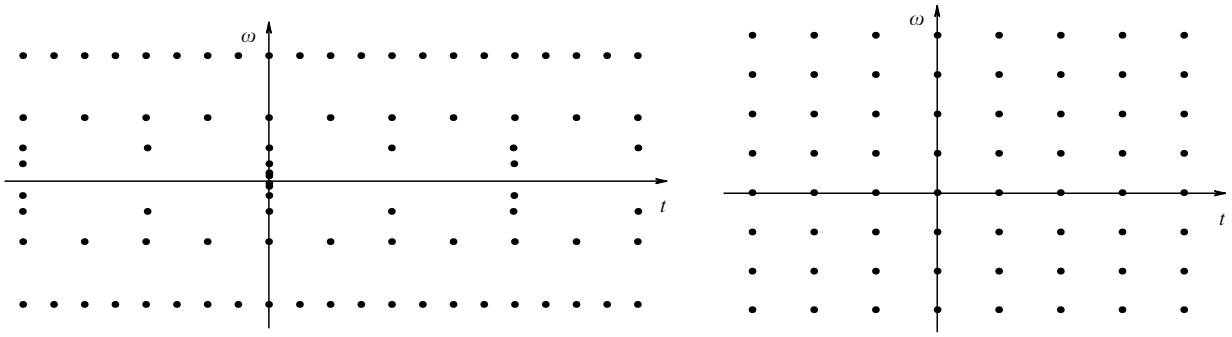

Figure 10. The lattices of time-frequency localization for the wavelet transform (left) and windowed Fourier transform (right).

Thus one gets

$$f_{\omega,w} = \int_{-\infty}^{\infty} f(t)\, g(t) \exp(-\mathrm{i}\omega t)\, \mathrm{d}t\,. \tag{8.4}$$

In its discrete form it can be rewritten as

$$f_{\omega,w} = \int f(t)\, g(t - nt_0) \exp(-\mathrm{i}m\omega_0 t)\, \mathrm{d}t\,, \tag{8.5}$$

where ω_0, $t_0 > 0$ are fixed, and m, n are some numbers which define the scale and location properties.

The time localization of the transform is limited now but the actual window is fixed by different functions in time and frequency which do not depend on the resolution scale and have fixed widths. Moreover, the orthonormal basis of the windowed Fourier transform can be constructed only for the so-called Nyquist density (corresponding to the condition $\omega_0 t_0 = 2\pi$; see Section 14), whereas there is no such restriction imposed on wavelets. At this critical value, frames are possible, but the corresponding functions are either poorly localized, or have poor regularity. This result is known as the Balian – Low phenomenon. In practical applications of the windowed Fourier transform, to achieve better localization one should choose $\omega_0 t_0 < 2\pi$ thus destroying the orthonormality.

The difference between the wavelet and windowed Fourier transforms lies in the shapes of the analyzing functions ψ and g. All g, regardless of the value of ω, have the same width. In contrast, the wavelets ψ automatically provide the time (or spatial location) resolution window adapted to the problem studied, i.e., to its essential frequencies (scales). Namely, let t_0, δ and ω_0, δ_ω be the centers and the effective widths of the wavelet basic function $\psi(t)$ and its Fourier transform. Then for the wavelet family $\psi_{j,k}(t)$ (3.15) and, correspondingly, for wavelet coefficients, the center and the width of the window along the t-axis are given by $2^j(t_0 + k)$ and $2^j\delta$. Along the ω-axis they are equal to $2^{-j}\omega_0$ and $2^{-j}\delta_\omega$. Thus the ratios of widths to the center position along each axis do not depend on the scale. This means that the wavelet window resolves both the location and the frequency in fixed proportions to their central values. For the high-frequency component of the signal it leads to a quite large frequency extension of the window whereas the time location interval is squeezed so that the Heisenberg uncertainty relation is not violated. That is why wavelet windows can be called *Heisenberg windows*. Correspondingly, the low-frequency signals do not require small time intervals and admit a wide window extension along the time axis. Thus wavelets well localize the low-frequency 'details' on the frequency axis and the high-frequency ones on the time axis. This ability of wavelets to find a perfect compromise between the time localization and the frequency localization by automatically choosing the widths of the windows along the time and frequency axes well adjusted to their centers location is crucial for their success in signal analysis. The wavelet transform cuts up the signal (functions, operators etc) into different frequency components, and then studies each component with a resolution matched to its scale providing a good tool for time-frequency (position-scale) localization. That is why wavelets can zoom in on singularities or transients (an extreme version of very short-lived high-frequency features!) in signals, whereas the windowed Fourier functions cannot. In terms of traditional signal analysis, the filters associated with the windowed Fourier transform are *constant bandwidth filters* whereas the wavelets may be seen as *constant relative bandwidth filters* whose widths in both variables linearly depend on their positions.

In Figure 10 we show the difference between these two approaches. It demonstrates the constant shape of the windowed Fourier transform region and the varying shape (with a constant area) of the wavelet transform region. The density of localization centers is homogeneous for the windowed Fourier transform whereas it changes for the wavelet transform so that at low frequencies the centers are far apart in time and become much denser for high frequencies.

From the mathematical point of view, it is important that orthonormal wavelets give good unconditional[11] bases for other spaces than that of the square integrable functions, outperforming the Fourier basis functions in this respect. It is applied in the subsequent sections to a characterization of such functions using only the absolute values of wavelet coefficients. In other words, by looking only at the absolute values of wavelet coefficients we can determine to which space this function belongs. This set of spaces is much wider than in the case of the Fourier transform by which only Sobolev spaces[12] can be completely characterized.

As we mentioned already, the wavelet analysis concentrates near the singularities of the function analyzed. The corresponding wavelet coefficients are negligible in the regions where the function is smooth. That is why wavelet series with plenty of non-zero coefficients represent really pathological functions, whereas 'normal' functions have

[11] For unconditional bases, the order in which the basis vectors are taken does not matter. All of the known constructions of unconditional bases of wavelets rely on the concept of multiresolution analysis.

[12] The function f belongs to the Sobolev space $W^s(R)$ if its Fourier transform provides the finite integrals $\int(1 + |\omega|^2)^s |f(\omega)|^2\, \mathrm{d}\omega$.

'sparse' or 'lacunary' wavelet series and are easy to compress. On the other hand, the Fourier series of the usual functions have a lot of non-zero coefficients, whereas 'lacunary' Fourier series represent pathological functions. Let us note finally that, nevertheless, the Fourier transform is systematically used in the proof of many theorems in the theory of the wavelet transform. It is not at all surprising because they constitute the stationary signals by themselves.

9. Wavelets and operators

The study of many operators acting on a space of functions or distributions becomes simple when suitable wavelets are used because these operators can be approximately diagonalized with respect to this basis. Orthonormal wavelet bases provide a unique example of a basis with non-trivial diagonal, or almost-diagonal, operators. The operator action on the wavelet series representing some function does not have uncontrollable sequences, i.e., wavelet decompositions are robust. One can describe precisely what happens to the initial series under the operator action and how it is transformed. In a certain sense, wavelets are stable under the operations of integration and differentiation. That is why wavelets, used as a basis set, allow us to solve differential equations characterized by widely different length scales found in many areas of physics and chemistry. Moreover, wavelets reappear as eigenfunctions of certain operators.

To deal with operators in the wavelet basis it is convenient, as usual, to use their matrix representation. For a given operator T it is represented by the set of its matrix elements in the wavelet basis:

$$T_{j,k;\,j',k'} = \langle \psi^*_{j,k} | T | \psi_{j',k'} \rangle . \tag{9.1}$$

For linear homogeneous operators, their matrix representation can be explicitly calculated [22].

It is extremely important that it is sufficient to first calculate the matrix elements at some (j-th) resolution level. All other matrix elements can be obtained from it using the standard recurrence relations. Let us derive the explicit matrix elements $r_{j,l;j,l'}$ of the homogeneous operator T of the order α:

$$r_{j,l;\,j,l'} = \langle \varphi_{j,l} | T | \varphi_{j,l'} \rangle . \tag{9.2}$$

Using the recurrence relations between the scaling functions at the given and finer resolution levels, one gets the following equation relating the matrix elements at neighboring levels:

$$\begin{aligned} r_{j,l;\,j,l'} &= \left\langle \Big(\sum_k h_k \, \varphi_{j+1,\,2l-k} \Big) | T | \Big(\sum_{k'} h_{k'} \, \varphi_{j+1,\,2l'-k'} \Big) \right\rangle \\ &= \sum_k \sum_{k'} h_k h_{k'} r_{j+1,\,2l-k;\,j+1,\,2l'-k'} . \end{aligned} \tag{9.3}$$

For the operator T having the homogeneity index α one obtains

$$r_{j,l;\,j,l'} = 2^{\alpha} \sum_k \sum_{k'} h_k h_{k'} r_{j,\,2l-k;\,j,\,2l'-k'} . \tag{9.4}$$

The solution of this equation defines the required coefficients up to the normalization constant which can be easily obtained from the results of the action by the operator T on a polynomial of a definite rank. For non-integer values of α, this is an infinite set of equations.

The explicit equation for the n-th order differentiation operator is

$$\begin{aligned} r^{(n)}_k &= \left\langle \varphi(x) \left| \frac{\mathrm{d}^n}{\mathrm{d}x^n} \right| \varphi(x-k) \right\rangle \\ &= \sum_{i,m} h_i h_m \left\langle \varphi(2x-i) \left| \frac{\mathrm{d}^n}{\mathrm{d}x^n} \right| \varphi(2x-m-2k) \right\rangle \\ &= 2^n \sum_{i,m} h_i h_m r^{(n)}_{2k-i+m} . \end{aligned} \tag{9.5}$$

This leads to a finite system of linear equations for r_k (the index n is omitted):

$$2^{-n} r_k = r_{2k} + \sum_m a_{2m-1} \big(r_{2k-2m+1} + r_{2k+2m-1} \big), \tag{9.6}$$

where both r_k and $a_m = \sum_i h_i h_{i+m}$ ($a_0 = 1$) are rational numbers in the case of Daubechies wavelets.[13] The wavelet coefficients can be found from these equations up to a normalization constant. The normalization condition reads [22]:

$$\sum_k^n r_k = n! . \tag{9.7}$$

For the support region of length L, the coefficients r_k differ from zero for $-L+2 \leqslant k \leqslant L-2$, and the solution exists for $L \geqslant n+1$. These coefficients possess the following symmetry properties:

$$r_k = r_{-k} \tag{9.8}$$

for even n, and

$$r_k = -r_{-k} \tag{9.9}$$

for odd values of n.

Here, as an example, we show in Table the matrix elements of the first and second order differential operators in the Daubechies wavelet basis with four vanishing moments (D^8).

Table.

k	h_k	$\langle \varphi(x) \vert \nabla \vert \varphi(x-k) \rangle$	$\langle \varphi(x) \vert \nabla^2 \vert \varphi(x-k) \rangle$
−6	0	0.00000084	0.00001592
−5	0	−0.00017220	−0.00163037
−4	0	−0.00222404	−0.01057272
−3	0	0.03358020	0.15097289
−2	−0.07576571	−0.19199897	−0.69786910
−1	−0.02963552	0.79300950	2.64207020
0	0.49761866	0	−4.16597364
1	0.80373875	−0.79300950	2.64207020
2	0.29785779	0.19199897	−0.69786910
3	−0.09921954	−0.03358020	0.15097289
4	−0.01260396	0.00222404	−0.01057272
5	0.03222310	0.00017220	−0.00163037
6	0	−0.00000084	0.00001592

[13] The matrix elements r_k are the same for all Daubechies wavelets with a fixed number of vanishing moments M, while there are several wavelet bases for a given M depending on the choice of roots of polynomials as has been demonstrated in formulas (4.12), (4.13). Formulas (9.5), (9.6) have been corrected in the English proofs. (This footnote has been added to the English proofs.)

For a continuous linear operator T represented by a singular integral

$$Tf(x) = \int K(x,y) f(y)\,\mathrm{d}y\,, \tag{9.10}$$

with some definite conditions imposed on the kernel K (see Section 17.2), there exists an important theorem [called the $T(1)$ *theorem*] which states a necessary and sufficient condition for the extension of T as a continuous linear operator on $L^2(R^n)$ (more details and the elegant wavelet proof of the theorem are given in Ref. [3]).

Let us note that a standard difficulty for any spectral method is the representation of the operator of multiplication by a function. As an example for physicists, we would mention the operator of the potential energy in the Schrödinger equation. However, it is well known that this operation is trivial and diagonal in real space. Therefore for such operations one should deal with a real space, and after all the operations are done there, return. Such an algorithm would need $O(N)$ operations only.

10. Nonstandard matrix multiplication

There are two possible ways to apply operators to functions within wavelet theory. They are called the *standard* and *nonstandard* matrix forms.

For smooth enough functions most wavelet coefficients are rather small. For a wide class of operators, most of their matrix elements are also quite small. Let us consider the structure of the elements of the matrix representation of some operator T that are large enough. The matrix elements satisfy the following relations

$$T_{j,k;\,j',k'} \to 0 \quad \text{at} \quad |k-k'| \to \infty\,, \tag{10.1}$$

$$T_{j,k;\,j',k'} \to 0 \quad \text{at} \quad |j-j'| \to \infty\,. \tag{10.2}$$

The topology of the distribution of these matrix elements within the matrix can be rather complicated. The goal of the nonstandard form is to replace the latter equation (10.2) by another, more rigorous one:

$$T_{j,k;\,j',k'} = 0 \quad \text{at} \quad j \neq j'\,. \tag{10.3}$$

It avoids taking the matrix elements between the different resolution levels. To deal with it, one should consider, instead of the wavelet space, the overfull space with the basis containing both wavelets and scaling functions at various resolution levels.

Let us consider the action of the operator T on the function f which transforms it into the function g:

$$g = Tf. \tag{10.4}$$

Both g and f have wavelet representations with the wavelet coefficients $({}_f s_{j,k}; {}_f d_{j,k})$ and $({}_g s_{j,k}; {}_g d_{j,k})$. At the finest resolution level j_n only the s-coefficients differ from zero, and the transformation looks like

$${}_g s_{j_n,k} = \sum_{k'} T_{SS}(j_n,k;j_n,k')\,{}_f s_{j_n,k'}\,. \tag{10.5}$$

At the next level, in both the standard and nonstandard approaches one gets

$${}_g s_{j_n-1,k} = \sum_{k'} T_{SS}(j_n-1,k;j_n-1,k')\,{}_f s_{j_n-1,k'} + \sum_{k'} T_{SD}(j_n-1,k;j_n-1,k')\,{}_f d_{j_n-1,k'}, \tag{10.6}$$

$${}_g d_{j_n-1,k} = \sum_{k'} T_{DS}(j_n-1,k;j_n-1,k')\,{}_f s_{j_n-1,k'} + \sum_{k'} T_{DD}(j_n-1,k;j_n-1,k')\,{}_f d_{j_n-1,k'}\,, \tag{10.7}$$

where

$$T_{SS}(j_n,k;j_n,k') = \int \bar{\varphi}_{j_n,k}(x)\, T\varphi_{j_n,k'}(x)\,\mathrm{d}x$$

and the replacement of subscripts $S \to D$ corresponds to the substitution $\varphi \to \psi$ in the integrals.

There is coupling between all resolution levels because all s-coefficients at this (j_n-1)-th level should be decomposed by the fast wavelet transform into s- and d-coefficients at higher levels. Therefore even for the almost diagonal initial step the standard matrix acquires a rather complicated form as demonstrated in Fig. 11 for 4-level operations (similar to those discussed above in the case of the Haar wavelets). It is thus inefficient for numerical purposes.

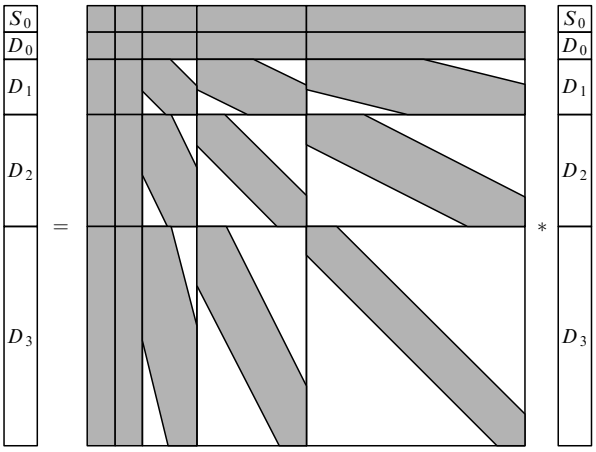

Figure 11. Matrix representation of the standard approach to the wavelet analysis. The parts containing non-zero wavelet coefficients are shaded.

As we see in Fig. 11, in the final stage of the standard approach we have to deal with the wavelet representation corresponding to formula (3.17) with only one s-coefficient left in the vectors which represents the overall weighted average of the functions (the SS-transition from f to g is given by the upper left box in the figure). At the same time, in the process of approaching it from the scaling-function representation (3.18), (2.1) we had to deal with average values at intermediate levels decomposing them at each step into s and d parts of further levels. These intermediate s-coefficients have been omitted since they were replaced by s- and d-coefficients at the next levels. That is why the matrix of the standard approach looks so complicated.

To simplify the form of the matrix, it was proposed [23] to use the redundant set of wavelet coefficients. Now, let us keep

these average values in the form of the corresponding *s*-coefficients of the intermediate levels both in the initial and final vectors representing functions f and g. Surely, we will deal with redundant vectors which are larger than necessary for the final answer. However, it should not bother us because we know the algorithm to reduce the redundant vector to a non-redundant form. At the same time, this trick simplifies both the form of the transformation matrix and the computation. This non-standard form is shown in Fig. 12. Different levels have been completely decoupled because there are no blocks in this matrix which would couple them. The block with *SS*-elements has been separated, and in its place the zero matrix is inserted. The whole matrix becomes artificially enlarged. Correspondingly, the vectors, characterizing the functions f and g, are also enlarged. Here all intermediate *s*-coefficients are kept for the function f (compare the vectors in the right-hand sides of Figs 11 and 12). Each S_{j+1} is generated from S_j and D_j. That is where the coupling of different levels is still present. In the transformation matrix all *SS*-elements are zero except for the lowest one S_0S_0. All other *SD*, *DS*, *DD* matrices are almost diagonal due to the finite support of scaling functions and wavelets. The redundant form of the g-function vector of Fig. 12 may be reduced to its usual wavelet representation of Fig. 11 by splitting up any S_j into S_{j-1} and D_{j-1} by a standard method. Then these S_{j-1} and D_{j-1} are added to the corresponding places in the vector. This iterative procedure allows one, by going from S_{j-1} down to S_0, to get the usual wavelet series of the function g. We get rid of all *s*-coefficients apart from S_0. The computation becomes fast.

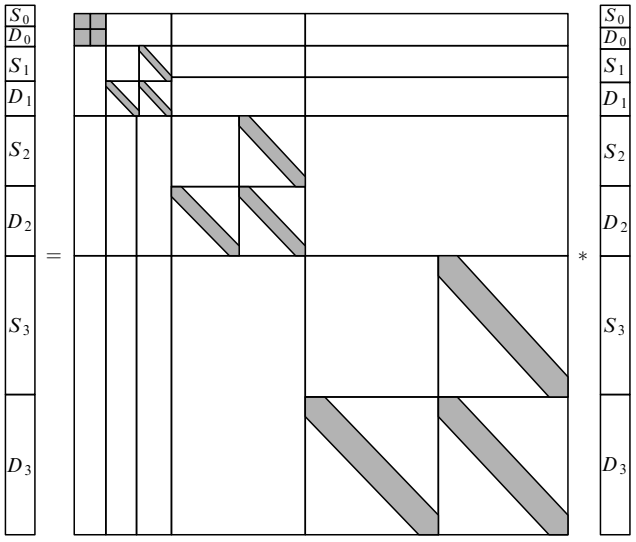

Figure 12. Nonstandard matrix multiplication in the wavelet analysis.

11. Regularity and differentiability

The analysis of any signal includes finding the regions of its regular and singular behavior. One of the main features of wavelet analysis is its capacity of doing a very precise analysis of the regularity properties of functions. When representing a signal by a wavelet transform, one would like to know if and at which conditions the corresponding series is convergent and, therefore, where the signal is described by the differentiable function or where the singularities appear. For certain singular functions, essential information is carried by a limited number of wavelet coefficients. Such information may be used in the design of numerical algorithms. We start with the traditional Hölder conditions of regularity and, in the next section, proceed to their generalization following Ref. [24].

The Hölder definition of the *pointwise regularity at a point* x_0 of a real-valued function f defined on R^n declares that this function belongs to the space $C^\alpha(x_0)$ ($\alpha > -n$) if there exists a polynomial P of order at most α such that

$$|f(x) - P(x - x_0)| \leqslant C|x - x_0|^\alpha, \tag{11.1}$$

where C is a constant.

The supremum of all values of α such that (11.1) holds is called the *Hölder exponent* of f at x_0 and denoted $\alpha(x_0)$ [14].

Correspondingly, a point x_0 is called a *strong α-singularity* ($-n < \alpha \leqslant 1$) of f if, at small intervals, the following inequality is valid

$$|f(x) - f(y)| \geqslant C|x - y|^\alpha \tag{11.2}$$

for a relatively large set of x's and y's close to x_0.

Such definitions work quite well for a rather smooth function. However, in the case of a function drastically fluctuating from point to point they are difficult to handle at each point because, e.g., the derivative of f may have no regularity at x_0, condition (11.1) may become unstable under some operators such as the Hilbert transform etc. The generalized definition of pointwise regularity will be given in the next section by introducing two-microlocal spaces. Here, we concentrate on global properties of the function f.

The uniform regularity of a function f at positive non-integer values of α consists in the requirement for Eqn (11.1) to hold for all real n-dimensional x_0 with a uniform value of the constant C. At first sight, this condition looks rather trivial because for smooth functions the uniform regularity coincides with the pointwise regularity, which is everywhere the same. To understand that this is non-trivial, one could consider, e.g., the function $f(x) = x\sin(1/x)$ with pointwise exponents $\alpha(0) = 1$, $\alpha(x_0) = \infty$ for any fixed x_0 different from 0, and uniform Hölder exponent $\alpha = 1/2$ which appears for the set of points $x_0 \propto |x - x_0|^{1/2} \gg |x - x_0|$.

The requirement of a definite uniform regularity of an n-dimensional function $f(x)$ on the whole real axis at positive non-integer values of α can be stated in terms of wavelet coefficients as an inequality

$$|d_{j,k}| \leqslant C2^{-(n/2+\alpha)j}. \tag{11.3}$$

Thus from the scale behavior of wavelet coefficients one gets some characteristics of the uniform regularity of the function. In particular, the linear dependence of the logarithms of wavelet coefficients on the scale j would indicate the scaling properties of a signal, i.e., the fractal behavior whose parameters are determined from higher moments of wavelet coefficients (see below). For pointwise regularity determining

[14] In principle, the expression $|x - x_0|^\alpha$ on the right-hand side of Eqn (11.1) can be replaced by a more general function satisfying definite conditions and called a modulus of continuity but we shall use only the above definition. In general, a modulus of continuity defines the largest deviation from that best polynomial approximation of a function f which is characterized by the set of smallest deviations compared to other polynomial approximations.

the local characteristics, the similar condition is discussed in the next section.

Now, let us formulate the conditions under which wavelet series converge at some points, or are differentiable. It has been proven that

• if f is square integrable, the wavelet series of f converges almost everywhere;

• the wavelet series of f is convergent on a set of points equipotent to R if

$$|d_{j,k}| = 2^{-j/2}\eta_j\,, \tag{11.4}$$

where η_j tends to 0 when j tends to $+\infty$;

• the function f is almost everywhere differentiable and the derivative of the wavelet series of f converges almost everywhere to f', if the following condition is satisfied

$$\sum_j |d_{j,k}|^2 2^{2j} \leqslant \infty; \tag{11.5}$$

• the function f is differentiable on a set of points equipotent to R and, at these points the derivative of its wavelet series converges to f', if

$$|d_{j,k}| \leqslant 2^{-3j/2}\eta_j\,, \tag{11.6}$$

where η_j tends to 0 when j tends to $+\infty$.

Limitations imposed on wavelet coefficients are generalized in the next section to include the pointwise regularity condition when two-microlocal spaces are considered. These better estimates do not hold just uniformly, but hold locally (possibly, apart from a set of points of small Hausdorff dimension which can be determined by wavelet analysis).

Note finally that one can also derive the global regularity of f from the decay in ω of the absolute value of its windowed Fourier transform, if the window function g is chosen to be sufficiently smooth. In most cases, however, the value of the uniform Hölder exponent computed from the Fourier coefficients will not be optimal. Nothing can be said from Fourier analysis about the local regularity, in contrast to two-microlocal wavelet analysis discussed below.

12. Two-microlocal analysis

The goal of two-microlocal analysis is to reveal the pointwise behavior of any function from the properties of its wavelet coefficients. The local regularity of the function is thus established.

The scalar product of an analyzed one-dimensional function and a wavelet maps this function into two-dimensional wavelet coefficients which reveal both the scale and location properties of a signal. In the above definitions of the pointwise regularity of the function at x_0 determined by the Hölder exponent only local but not scaling properties of the signal are taken into account. The problem of determining the exact degree of the Hölder regularity of a function is easily solved if this regularity is everywhere the same, because in such a case it usually coincides with the uniform regularity. The determination of the pointwise Hölder regularity, however, becomes much harder if the function changes wildly from point to point, i.e., its scale (frequency) characteristics depend strongly on the location (time). In this case we have to deal with a very non-stationary signal. To describe the singularity of $f(x)$ at x_0 only by the Hölder exponent $\alpha(x_0) \leqslant 1$ one looks for the order of magnitude of the difference $|f(x) - f(x_0)|$ when x tends to x_0, without taking into account, e.g., the possible high-frequency oscillations of $f(x) - f(x_0)$, i.e., its scale behavior.

The properties of the wavelet coefficients as functions of both scale and location [15] provide a unique way of describing the pointwise regularities. It is more general than the traditional Hölder approach because it allows us to investigate, characterize and easily distinguish some specific local behaviors such as approximate selfsimilarities and very strong oscillatory features like those of indefinitely oscillating functions which are closely related to so-called 'chirps' of the form $x^\alpha \sin(1/x^\beta)$ (reminiscent of bird, bat or dolphin sonar signals with very sharp oscillations which accelerate at some point x_0). Chirps are well known to everybody dealing with modern radar and sonar technology. In physics, they are known, e.g., in theoretical considerations of dark matter radiation and of gravitational waves. Sometimes, a similar dependence can reveal itself even in the more traditional correlation analysis [25–27]. A large value of the second derivative of the phase function of a frequency modulated signal is a typical feature. Such special behavior with frequency modulation laws hidden in a given signal can be revealed [24] with the help of two-microlocal analysis. Actually, there is no universally accepted definition of a chirp. Sometimes, any sine of a non-linear polynomial function of time is also called a chirp.

The two-microlocal space $C^{s,s'}(x_0)$ of the real-valued n-dimensional functions f (distributions) is defined by the following simple decay condition on their wavelet coefficients $d_{j,k}$

$$|d_{j,k}(x_0)| \leqslant C2^{-(n/2+s)j}\left(1 + |2^j x_0 - k|\right)^{-s'}, \tag{12.1}$$

where s and s' are two real numbers. This is a very important extension of the Hölder conditions. The two-microlocal condition is a local counterpart of the uniform condition (11.1). It is very closely related to the pointwise regularity condition (11.1) because it expresses, e.g., the singular behavior of the function itself at the point x_0 in terms of the k-dependence of its wavelet coefficients at the same point. Such an estimate is stable under derivation and fractional integration.

For $s' = 0$, the space thus defined is the global Hölder space $C^s(R^n)$.

If $s' > 0$, these conditions are weaker at x_0 than far from x_0 so that they describe the behavior of a function which is irregular at x_0 in a smooth environment, i.e., the regularity of f gets worse as x tends to x_0. If s' is positive, the inequality (12.1) implies that

$$|d_{j,k}| \leqslant C2^{-(n/2+s)j}\,,$$

so that f belongs to $C^s(R^n)$. The Hölder norms do not give any information about this singularity because one has to consider them in domains that exclude x_0, and the Hölder regularity of f deteriorates when x gets close to x_0. That is why we need a uniform regularity assumption when s' is positive.

If $s' < 0$, the converse situation persists when x_0 is a point of regularity in a nonsmooth environment.

[15] Let us note that the word 'location' (and, correspondingly, 'scale') does not necessarily imply a single dimension but could describe the position (and size or frequency) in any n-dimensional space.

Thus the inequality Eqn (12.1) clarifies the relationship between the pointwise behavior of a function and the size properties of its wavelet coefficients. If $s > 0$ and $s' > s$, then $C^{s,-s}(x_0)$ is included in $C^s(x_0)$. One can prove that the elements of the space $C^{s,s'}(x_0)$, for $s' > -s$, are functions whereas the elements of $C^{s,-s}(x_0)$ are (in general) not functions but distributions (and, moreover, quite 'wild distributions' for which the local Hölder condition (11.1) cannot hold and should be generalized by multiplying the right-hand side by the factor $\log 1/|x - x_0|$). However, if f belongs to the space $C^s(x_0)$ then it belongs to $C^{s,-s}(x_0)$ too. In particular, if at x close to x_0 the following inequality

$$|f(x)| \leqslant C|x - x_0|^s \quad \text{is valid for} \quad s \leqslant 0\,,$$

imposing some limit on the nature of the singularity at this point, then the estimates for wavelet coefficients are given by

$$|d_{j,k}| \leqslant C2^{-(n/2+s)j}|k - 2^j x_0|^s\,, \tag{12.2}$$

if the support of $\psi_{j,k}$ is at least at a distance 2^{-j} from x_0, and

$$|d_{j,k}| \leqslant C2^{-(n/2+s)j}\,, \tag{12.3}$$

if the support of $\psi_{j,k}$ is at a distance less than 2^{-j} from x_0, which demonstrate the above statement.

The two-microlocal spaces have some stability properties. In particular, the pseudodifferential operators of order 0 are continuous in these spaces (in contrast with the usual pointwise Hölder regularity condition which is not preserved under the action of these operators, for example, of the Hilbert transform). The position of the points of regularity of a function f which belongs to the space $C^{s,s'}(x_0)$ is essentially preserved under the action of singular integral operators such as the Hilbert transform. This property leads to a pointwise regularity result for solutions of partial differential equations which is formulated [24] by the following theorem.

Let Λ be a partial differential operator of order m, with smooth coefficients and elliptic at x_0. If $\Lambda f = g$ and g belongs to $C^{s,s'}(x_0)$, then f belongs to $C^{s+m,s'}(x_0)$.

Thus, if $\Lambda f = g$, and if g is a function that belongs to $C^s(x_0)$, then there exists a polynomial P of degree less than $s + m$ such that, for $|x - x_0| \leqslant 1$,

$$|f(x) - P(x)| \leqslant C\,|x - x_0|^{s+m}\,.$$

All the above conditions look as if the wavelets were eigenvectors of the differential operators d^α, $|\alpha| \leqslant r$, with corresponding eigenvalues $2^{j|\alpha|}$.

The more general functional spaces which admit the power dependence of wavelet coefficients on the scale j defined by an extra parameter p are considered in Ref. [24]. In particular, these are the Sobolev spaces where the integrability is required up to some power of both the function itself and its derivatives up to a definite order. In this case, the additional factor $j^{2/p}$ appears in the right-hand side of the inequality (11.1), or, in other words, in addition to the linear term in the exponent one should consider another logarithmically dependent term. Thus there is no rigorous selfsimilarity (fractality) of the function any more. However, the wavelet analysis allows us to determine the fractal dimensions of the sets of points where the function f is singular.

For continuous wavelets, the condition equivalent to Eqn (11.1) takes the form

$$|W(a,b)| \leqslant Ca^s\left(1 + \frac{|b - x_0|}{a}\right)^{-s'}\,. \tag{12.4}$$

The definite upper limits on the behavior of $|W(a,b)|$ can also be obtained if f is integrable in the neighborhood of the origin (see [24]). They are somewhat complicated, and we do not show them here.

It is instructive to examine the meaning of the conditions of the type (12.4). Let us consider the cone in the (b,a) half-plane defined by the condition $|b - x_0| < a$. Within this cone we have $|W(a,b)| = O(a^s)$ as $a \to 0$. Outside the cone, the behavior is governed by the distance of b to the point x_0. These two behaviors are generally different and have to be studied independently. However, it is shown in [28] that non-oscillating singularities may be characterized by the behavior of the absolute values of their wavelet coefficients within the cone. It is also shown in [28] that rapidly oscillating singularities, which we consider below, cannot be characterized by the behavior of their wavelet transform in the cone.

The two-microlocal methods proved especially fruitful in the analysis of oscillating trigonometric and logarithmic chirps.

A simple definition of trigonometric chirps at the origin $x = 0$ could simply be read

$$f(x) = x^\alpha g(x^{-\beta})\,, \tag{12.5}$$

where g is a 2π-periodic C^r function with vanishing integral. This type of behavior leads to the following expansion of continuous wavelet coefficients at small positive $b \leqslant \delta$ on curves $a = \lambda b^{1+\beta}$

$$W(\lambda b^{1+\beta}, b) = b^\alpha m_\lambda(b^{-\beta})\,, \tag{12.6}$$

where m_λ is a 2π-periodic function with a definite norm. It shows that rapidly oscillating singularities cannot be characterized by the behavior of their wavelet transform in the cone as in the previous examples. In this case, the wavelet coefficients should be carefully analyzed not in the cone but on definite curves because they are concentrated along these ridges. For small enough scales, such a ridge lies outside the influence cone of the singularity. A typical example is given by the function $f(x) = \sin(1/x)$ whose instantaneous frequency tends to infinity as $x \to 0$. The wavelet coefficient modulus is maximum on a curve of equation $b = Ca^2$ for some constant C depending only on the wavelet and this curve is not a cone. Therefore it is not sufficient to study the behavior of wavelet coefficients inside the cone to characterize the rapidly oscillating singularities. Let us note, that in case of noisy signals, the contribution of the deterministic part of the signal may be expected to be much larger than that of the noise just near the ridges of the wavelet transform because the wavelet transform of the noise is spread over the whole (j,k)-plane. It can be thus used for denoising the signal.

Logarithmic chirps have an approximate scaling invariance near the point x_0. A function f has a logarithmic chirp of order (α, λ) and of regularity $\gamma \geqslant 0$ at the origin $x = 0$ if for $x > 0$ there exists a $\log\lambda$-periodic function $G(\log x)$ in $C^\gamma(R)$ such that

$$f(x) - P(x) = x^\alpha G(\log x)\,. \tag{12.7}$$

The continuous wavelet coefficients of it satisfy the condition

$$W(a,b) = a^{\alpha} H\left(\log a,\ \frac{b}{a}\right), \tag{12.8}$$

where H is $\log\lambda$-periodic in the first argument and its behavior in b/a is restricted by some decay conditions. In other words, if the following scaling property is observed

$$C(\lambda a, \lambda b) = \lambda^{\alpha} C(a,b) \tag{12.9}$$

for $\lambda < 1$ and small enough a and b, then f has a logarithmic chirp of order (α, λ). More general and rigorous statements and proofs can be found in Ref. [24]. Thus we conclude that the type and behavior of a chirp can be determined from the behavior of its wavelet coefficients.

As an example of the application of the above statements, let us mention the continuous periodic function $\sigma(x)$ represented by the Riemann series

$$\sigma(x) = \sum_{n=1}^{\infty} \frac{1}{n^2} \sin(\pi n^2 x), \tag{12.10}$$

which belongs to the space $C^{1/2}$. It is best analyzed by a continuous wavelet transform using the specific complex wavelet $\psi(x) = (x+i)^{-2}$ proposed by Lusin in the 1930s in functional analysis. Then the wavelet transform of the Riemann series is given by the well known Jacobi Theta function. The behavior of wavelet coefficients determines the chirp type and its parameters. With the help of the two-microlocal analysis it was proven that this function has trigonometric chirps at some rational values of $x = x_0$ which are ratios of two odd numbers (also, its first derivative exists only at these points) and logarithmic chirps at quadratic irrational numbers (near these points its wavelet transform possesses scaling invariance properties).

For practical purposes, however, it may happen that very large values of j are needed to determine Hölder exponents reliably from the above conditions. For the global Hölder exponent, no assumptions about regularity of wavelets is required whereas determination of the local Hölder exponents requires a more detailed approach.

A nice illustration of the application of wavelet analysis for studies of local regularity properties of the function $f(x)$ is given in [2] for the following function

$$\begin{aligned} f(x) = {} & 2\exp(-|x|)\,\theta(-x-1) \\ & + \exp(-|x|)\,\theta(x+1)\theta(1-x) \\ & + \exp(-x)\left[(x-1)^2+1\right]\theta(x-1), \end{aligned} \tag{12.11}$$

which is infinitely differentiable everywhere except at $x = -1, 0, 1$ where, respectively, f, f', f'' are discontinuous. One computes for each of the three points the maxima of wavelet coefficients $A_j = \max_k |W_{j,k}|$ at various resolution levels $3 < j < 10$ and plots $\log A_j / \log 2$ versus j. The linear dependences on j are found at these points. The slopes at $x = -1, 0, 1$ are, correspondingly, -0.505; -1.495; -2.462, leading with a pretty good accuracy of 1.5% to rather precise estimates of the Hölder exponents 0, 1, 2.

Even better accuracy in determining the local regularity of a function can be achieved with the help of redundant wavelet families (frames) where the translational non-invariance is much less pronounced, the wavelet regularity plays no role and only the number of vanishing moments is important.

If orthonormal wavelets are used, then their regularity can become essential. Typically, they are continuous, have a non-integer uniform Hölder exponent and different local exponents at different points. In fact, there exists a whole hierarchy of (fractal) sets in which these wavelets have different Hölder exponents. There is a direct relation between the regularity of the function ψ and its number of vanishing moments. The more moments are vanishing, the smoother is the function ψ. For example, for the widely used Daubechies wavelets with a compact support the degree of regularity increases linearly with the maximum number of vanishing moments M for higher-tap wavelets as μM with $\mu = 0.2$ and, correspondingly, with the support width, as was already mentioned in Section 4. These properties justify the name 'mathematical microscope', which is sometimes bestowed on the wavelet transform. Let us note that the choice of the h_k that leads to maximal regularity is, however, different from the choice with maximal number of vanishing moments for ψ. How well the projections of a multiresolution approximation V_j converge to a given function f depends on the regularity of the functions in V_0.

There exists a special class of wavelets called *vaguelets*. They possess regularity properties characterized by the following restrictions:

$$\left|\psi_{j,k}(x)\right| \leqslant C 2^{nj/2}\left(1 + |2^j x - k|\right)^{-n-\alpha}, \tag{12.12}$$

$$\left|\psi_{j,k}(x') - \psi_{j,k}(x)\right| \leqslant C 2^{j(n/2+\beta)} |x' - x|^{\beta} \tag{12.13}$$

with $0 < \beta < \alpha < 1$ and a constant C. Surely, the cancellation condition (3.16) must be also satisfied. The norm of any function f is then limited by its wavelet coefficients as

$$||f|| \leqslant C'\left(\sum_{j,k} |d_{j,k}|^2\right)^{1/2}, \tag{12.14}$$

with a constant C'. Any continuous linear operator T on L^2 which satisfies the condition $T(1) = 0$ transforms an orthonormal wavelet basis into vaguelets.

In practice, the regularity of a wavelet can become especially important during the synthesis when after omission of small wavelet coefficients it is better to deal with a rather smooth ψ to diminish possible mistakes at the restoration stage. On the contrary, for analysis it seems to be more important to have wavelets with many vanishing moments to 'throw out' the smooth polynomial trends and reveal potential singularities. Also, the large number of vanishing moments leads to better compression but can enlarge mistakes in the inverse procedure of reconstruction because of the worsened regularity of wavelets. The use of the biorthogonal wavelets helps a lot at this stage because among two dual wavelets one has many vanishing moments whereas the other possesses good regularity properties. By choosing an appropriate pair of the biorthogonal wavelets one can minimize possible mistakes.

13. Wavelets and fractals

Some signals (objects) possess self-similar (fractal) properties (see, e.g., [29 – 32]). This means that by changing the scale one observes features at a new scale similar to those previously noticed at other scales. This property leads to power-like dependences. The formal definition of a (mono)fractal Hausdorff dimension D_F of a geometrical object is given by

the condition

$$0 < \lim_{\varepsilon \to 0} N(\varepsilon)\, \varepsilon^{D_F} < \infty\,, \tag{13.1}$$

which states that D_F is the only value for which the product of the minimal number of the (covering this object) hypercubes $N(\varepsilon)$ with a linear size $l = \varepsilon$ and of the factor ε^{D_F} stays constant for ϵ tending to zero. In the common school geometry of homogeneous objects, this coincides with the topological dimension. The probability $p_i(\varepsilon)$ to belong to a hypercube $N_i(\varepsilon)$ is proportional to ε^{D_F}, and the sum of moments is given by

$$\sum_i p_i^q(\varepsilon) \propto \varepsilon^{qD_F}\,. \tag{13.2}$$

The fractal dimension is directly related to the Hölder exponents.

Moreover, for more general objects called multifractals, the 'fractal exponents' $\alpha(x_0)$ [see Eqn (11.1)] vary from point to point. The Hausdorff dimension of the set of points x_0, where $\alpha(x_0) = \alpha_0$, is a function of $d(\alpha_0)$ whose graph determines the multifractal properties of a signal, let it be Brownian motion, fully developed turbulence or a purely mathematical construction of the Riemann series. Thus the weights of various fractal dimensions inside a multifractal differ for different multifractals, and the value D_F is now replaced by D_{q+1} which depends on q. It is called the *Renyi* (or *generalized*) *dimension* [33]. Usually, it is a decreasing function of q. The fractal (Hausdorff) information and correlation dimensions are, correspondingly, obtained from the Renyi dimension at $q = -1$, 0, 1. The difference between the topological and Renyi dimensions is called the *anomalous dimension* (or *codimension*). Fractals and multifractals are common among the purely mathematical constructions (the Cantor set, the Serpinsky carpet etc) and in nature (clouds, lungs etc).

The pointwise Hölder exponents are now determined using wavelet analysis. As we have seen, all wavelets of a given family $\psi_{j,k}(x)$ are similar to the basic wavelet $\psi(x)$ and derived from it by dilations and translations. Since wavelet analysis just consists in studying the signal at various scales by calculating the scalar product of the analyzing wavelet and the signal explored, it is well suited to revealing the fractal peculiarities. In terms of wavelet coefficients it implies that their higher moments behave in a power-like manner with the scale changing. The wavelet coefficients are less sensitive to noise because they measure, at different scales, the average fluctuations of the signal.

Namely, let us consider the sum Z_q of the q-th moments of the coefficients of the wavelet transform at various scales j

$$Z_q(j) = \sum_k |d_{j,k}|^q\,, \tag{13.3}$$

where the sum is over the maxima of $|d_{j,k}|$. Then it was shown [34, 35] that for a fractal signal this sum should behave as

$$Z_q(j) \propto 2^{j[\tau(q)+q/2]}\,, \tag{13.4}$$

i.e.,

$$\log Z_q(j) \propto j\left[\tau(q) + \frac{q}{2}\right]. \tag{13.5}$$

Thus the necessary condition for a signal to possess fractal properties is the linear dependence of $\log Z_q(j)$ on the level number j. If this requirement is fulfilled the dependence of τ on q shows whether the signal is monofractal or multifractal. Monofractal signals are characterized by a single dimension and, therefore, by a linear dependence of τ on q, whereas multifractal ones are described by a set of such dimensions, i.e., by non-linear functions $\tau(q)$. Monofractal signals are homogeneous, in the sense that they have the same scaling properties throughout the entire signal. Multifractal signals, on the other hand, can be decomposed into many subsets characterized by different local dimensions, quantified by a weight function. The wavelet transform, if done with wavelets possessing the appropriate number of vanishing moments, removes lowest polynomial trends that could cause the traditional box-counting techniques to fail in quantifying the local scaling of the signal. The function $\tau(q)$ can be considered as a scale-independent measure of the fractal signal. It can be further related to the Renyi dimensions, and Hurst and Hölder (at $q = 1$ as is clear from the examples in the previous sections) exponents (for more detail, see Refs. [36, 37]). The range of validity of the multifractal formalism for functions can be elucidated [38] with the help of the two-microlocal methods generalized to the higher moments of wavelet coefficients. Thus, wavelet analysis goes very far beyond the limits of the traditional analysis which uses the language of correlation functions (see, e.g., [36]) in approaching much deeper correlation levels.

Let us note that $Z_2(j)$ is just the dispersion (variance) of wavelet coefficients whose average is equal to zero. For positive values of q, $Z_q(j)$ reflects the scaling of the large fluctuations and strong singularities, whereas for negative q it reflects the scaling of the small fluctuations and weak singularities, thus revealing different aspects of underlying dynamics.

14. Discretization and stability

In signal analysis, real-life applications produce only sequences of numbers due to the discretization of continuous time signals. This procedure is called the *sampling* of analog signals. Below we consider its implications. At first sight, it seems that in this case the notions of singularities and Hölder exponents are meaningless. Nevertheless, one can say that the behavior of wavelet coefficients across scales provides a good way of describing the regularity of functions whose samples coincide with the observations at a given resolution.

First, we treat the doubly infinite sequence

$$f^{(d)} = \{f_n\} = \{f(n\,\Delta t)\} \qquad (-\infty < n < \infty)\,,$$

obtained by sampling a continuous (analog) signal at the regularly spaced values $t_n = n\Delta t$. If within each n-th interval Δt the function f can be replaced by the constant value $f(n\Delta t)$ then for small enough Δt one gets

$$f_\omega = \int_{-\infty}^{\infty} f(t)\exp(-\mathrm{i}\omega t)\,\mathrm{d}t \approx \Delta t \sum_{n=-\infty}^{\infty} f(n\,\Delta t)\exp(-\mathrm{i}n\omega\,\Delta t)\,. \tag{14.1}$$

The inversion formula reads

$$f_n = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_\omega \exp(\mathrm{i}n\omega\Delta t)\,\mathrm{d}\omega\,. \tag{14.2}$$

The function f_ω is periodic with the period $2\pi/\Delta t$. This means that one can consider it only within the frequency interval $[-\pi/\Delta t, \pi/\Delta t)$. Moreover, for real signals even the interval $[0, \pi/\Delta t)$ is sufficient. For time-limited signals the summation in Eqn (14.1) is done from $n_{\min} = 0$ to $n_{\max} = N - 1$ for N sampled values of a signal. For fast Fourier transform algorithms, it requires $O(N \log N)$ computations.

To get the relation between Fourier transforms of the continuous time signal f and of the sequence of samples $f^{(d)}$ we rewrite formula (14.2)

$$\begin{aligned} f_n &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \int_{(2k-1)\pi/\Delta t}^{(2k+1)\pi/\Delta t} f_\omega \exp(\mathrm{i}n\omega\Delta t)\, \mathrm{d}\omega \\ &= \frac{1}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} \sum_{k=-\infty}^{\infty} f_{\omega+2k\pi/\Delta t} \exp(\mathrm{i}n\omega\Delta t)\, \mathrm{d}\omega \\ &= \frac{1}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} f_\omega^{(d)} \exp(\mathrm{i}n\omega\Delta t)\, \mathrm{d}\omega\,, \end{aligned} \tag{14.3}$$

from which one gets

$$f_\omega^{(d)} = \sum_{k=-\infty}^{\infty} f_{\omega+2k\pi/\Delta t}\,. \tag{14.4}$$

Thus the Fourier transform of the discrete sample contains, in general, contributions from the Fourier transform of the continuous signal not only at the same frequency ω but also at the countably infinite set of frequencies $\omega + 2k\pi/\Delta t$. These frequencies are called *aliases of the frequency* ω. For 'undersampled' sets, the aliasing phenomenon appears, i.e., the admixture of high frequency components to lower frequencies. The frequency $\omega_{(N)} = \pi/\Delta t$ is called the *Nyquist frequency*. To improve the convergence of the series, 'oversampling' is used, i.e., f is sampled at a rate exceeding the Nyquist rate.

The band-limited function f can be recovered from its samples f_n's whenever the sampling frequency $\omega_s \propto (\Delta t)^{-1}$ is not smaller than the band-limit frequency ω_f, i.e., whenever $\omega_s \geqslant \omega_f$. This statement is known as the sampling (or *Shannon – Kotel'nikov*) *theorem*.

In this case, there is no aliasing since only one aliased frequency is in the band limits $[-\omega_f, \omega_f]$. If $\omega_s < \omega_f$, then the function f cannot be recovered without additional assumptions.

A rigorous proof of the theorem exists, but it is already quite clear from intuitive arguments that one cannot restore the high-frequency content of a signal by sampling it with lower frequency. Therefore, to get knowledge of high-frequency components one should sample a signal with a frequency exceeding all the frequencies important for a given physical process. Only then is the restoration of a signal stable. One can reduce the sampling frequency ω_s all the way down to ω_f without losing any information. This allows one to subsample the signal by keeping only smaller sets of data, i.e., to get a shorter sampled signal.

In the case of wavelet analysis, to have a numerically stable reconstruction algorithm for f from $d_{j,k}$ one should be sure that $\psi_{j,k}$ constitute a frame. For better convergence, one needs frames close to tight frames, i.e., those satisfying the condition $(B/A - 1) \ll 1$. The orthonormal wavelet bases have good time-frequency localization. In principle, if ψ itself is well localized both in time and in frequency, then the frame generated by ψ will share that property as well.

Discrete wavelets are quite well suited for the description of functions by their mean values at equally spaced points. However in practical real-life applications, apart from this projection of a function, one has to deal with the finite interval of its definition and with a finite number of resolution levels. As was mentioned in Section 5, one usually rescales the 'units' of levels by assuming that the label of the finest available scale is $j = 0$, and the coarser scales have positive labels $1 \leqslant j \leqslant J$. The fine-level coefficients are defined by the sampled values of $f(x)$ as $s_{0,k} = f(k)$, i.e., instead of the function $f(x)$ one considers its projection $P_0 f$. The higher-level coefficients are found from the iterative relations (5.2), (5.3), i.e., with the help of the fast wavelet transform without direct calculation of integrals $\int f(x)\psi_{j,k}(x)\,\mathrm{d}x$ and $\int f(x)\varphi_{j,k}(x)\,\mathrm{d}x$. Therefore the approximate representation of the function $f(x)$ which corresponds to the redefined versions of Eqns (2.5) and (3.17) can be written as

$$f(x) \approx P_0 f = \sum_{j=1}^{J} \sum_k d_{j,k}\, \psi_{j,k} + \sum_k s_{J,k}\, \varphi_{J,k}\,, \tag{14.5}$$

where the sum over k is limited by the interval in which $f(x)$ is defined. Moreover, to save computing time, one can use not the complete set of wavelet coefficients $d_{j,k}$ but only a part of them omitting small coefficients not exceeding some threshold value ϵ. This standard estimation method is called *estimation by coefficient thresholding*. If the sum in (14.5) is taken only over such coefficients $|d_{j,k}| > \varepsilon$ and the number of omitted coefficients is equal to n_0, then the function f_ε which approximates $f(x)$ in this situation will differ from $f(x)$ by the norm as

$$||f(x) - f_\varepsilon(x)|| < \varepsilon n_0^{1/2}\,. \tag{14.6}$$

Thus for a function, which is quite smooth in the main domain of its definition and changes drastically only in very small regions, the threshold value ε can be extremely small. Then it admits a large number of wavelet coefficients to be omitted with rather low errors in the final approximations. Instead of this so-called *hard-shrinkage procedure*, one can use a *soft-shrinkage thresholding* [39] which shifts positive and negative coefficients to their common origin after the omission procedure, i.e., replaces non-omitted $d_{j,k}$ in formula (14.5) by

$$d_{j,k}^{(\varepsilon)} = \operatorname{sign}(d_{j,k})(|d_{j,k}| - \varepsilon)\,. \tag{14.7}$$

It has been proven that such an approach leads to optimal min-max estimators.

One can find the coefficients of expansion (14.5) using the fast wavelet transform since the coefficients $s_{0,k}$ are fixed as the discrete values of $f(x)$. In iterative schemes, the error will however accumulate and their precision will not be sufficiently high. Much better accuracy can be achieved if interpolation wavelets are used. In this case, the values of the function on the homogeneous grid $f(k)$ are treated as s-coefficients for the interpolation basis, and the initial values of $s_{0,k}$ are formed by some linear combinations of them with coefficients which depend on the shapes of the wavelets considered.

The more elaborate procedure of weighting different wavelet coefficients before their omission called *quantization* is often used instead of this simplified procedure of coefficient

thresholding (see also Section 15.4). According to expert estimates, different weights (importance) are ascribed to different coefficients from the very beginning. These weights depend on the final goals of compression. Only then is the thresholding applied. Many orthogonal wavelet bases are infinitely more robust with respect to this quantization procedure than the Fourier trigonometric basis. In spite of this, the lack of symmetry for Daubechies wavelets with compact support does not always satisfy the experts because some visible defects appear. This problem can be cured when one uses symmetric biorthogonal wavelets having compact support.

15. Some applications

In this section we describe mostly those applications of wavelets which are closest to our personal interests (a brief summary is given in Ref. [40]). Even among them we have to choose those where, in our opinion, the use of wavelets was crucial for obtaining principally new information unavailable using other methods (see Web site www.awavelet.com).

15.1 Physics

The physics applications of wavelets are so numerous that it is impossible to describe even the most important of them here in detail (see, e.g., [10, 14, 41–43]). They are used both in purely theoretical studies in functional calculus, renormalization in gauge theories, conformal field theories, nonlinear chaoticity, and in more practical fields like quasicrystals, meteorology, acoustics, seismology, nonlinear dynamics of accelerators, fluid dynamics and turbulence, the structure of surfaces, cosmic ray jets, solar wind, galactic structure, cosmological density fluctuations, dark matter, gravitational waves etc. This list can easily be made longer. We discuss here two problems related to use of wavelets for solving differential equations with the aim of getting the electronic structure of an extremely complicated system, and for pattern recognition in multiparticle production processes in high energy collisions.

15.1.1 Solid state and molecules. The exact solution of a many-body problem is impossible, and one applies various approximate methods for solid state problems, e.g., the density functional theory [44]. However, the electronic spectra of complex atomic systems are so complicated that, even within this approach, it is impossible, in practice, to decipher them by commonly used methods. For example, it would require the calculation of about 2^{100} Fourier coefficients to represent the effective potential of the uranium atom (and even more, in case of uranium dimers). This is clearly an unrealistic task. The application of wavelet analysis methods makes it possible to resolve this problem [20, 45]. The potential of the uranium dimer is extremely singular. It varies by more than 10 orders of magnitude. Its reconstruction has now become a realistic problem. Quite high precision has been achieved with the help of wavelets as seen in Fig. 13.

The solution of the density functional theory equations by wavelet methods has been obtained by several groups [46–50]. The method described in Refs [49, 50] has been applied to a large variety of different materials. It possesses good convergence properties as is shown in Figs 14 and 15. They demonstrate how fast is the decrease of energy (or wave function) recipies with respect to the number of iterations (Fig. 14) or of the absolute values of the wavelet coefficients

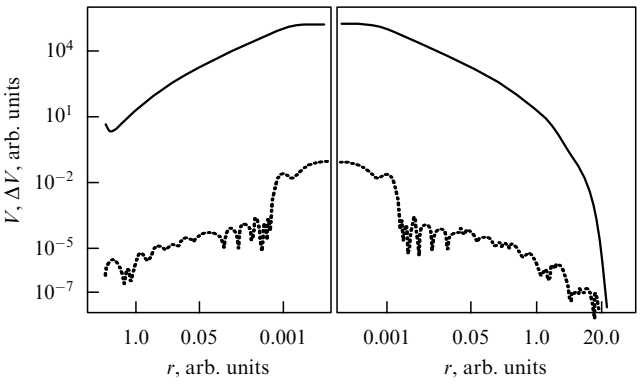

Figure 13. Coulomb potential V (the solid line) as a function of the distance and the precision of its calculation ΔV (the dashed line) for the Uranium dimer in arbitrary units. It is drawn for the direction from one atom to another one on the left-hand side and for the opposite direction on the right-hand side. The complete symmetry is seen with the larger distances shown on the right-hand side.

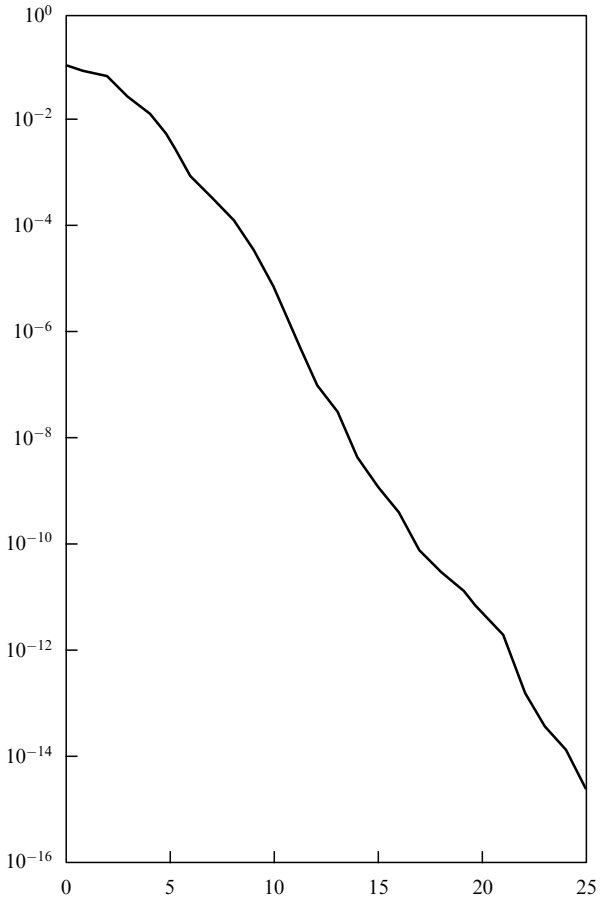

Figure 14. Fast decrease of the energy recipies with respect to the number of iterations demonstrates the good convergence of the solution of the density functional theory equations using the wavelet methods.

(Fig. 15) with respect to their index (when the coefficients are ordered according to their magnitude). Such a fast decrease allows us to consider much more complicated systems than it was possible before. This method has been successfully tested for solid hydrogen crystals at high pressure, manganate and hydrogen clusters [49, 50], 3*d*-metals and their clusters etc.

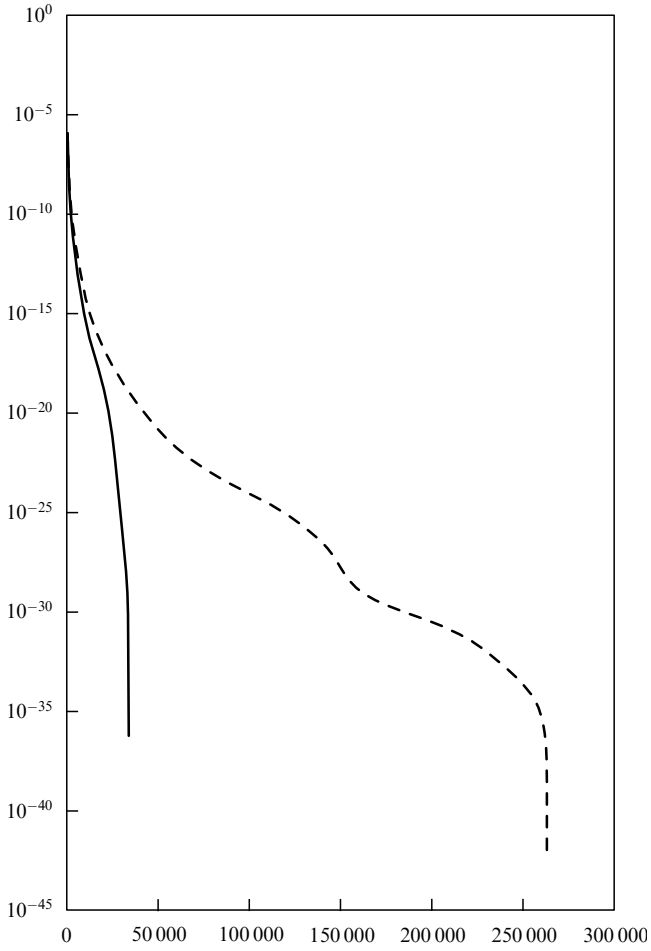

Figure 15. Same as in Fig. 14 for the absolute values of the wavelet coefficients.

The density matrix can be effectively represented with the help of Daubechies wavelets [51].

15.1.2 Multiparticle production processes. The first attempts to use wavelet analysis in multiparticle production go back to P Carruthers [52 – 54] who used wavelets for diagonalisation of covariance matrices of some simplified cascade models. The proposals for correlation studies in high multiplicity events with the help of wavelets were promoted [55, 56], and also, in particular, for special correlations typical for disoriented chiral condensate [57, 58]. It was recognized [59] that wavelet analysis can be used for pattern recognition in individual high multiplicity events observed in experiment.

High energy collisions of elementary particles result in the production of many new particles in a single event. Each newly created particle is depicted kinematically by its momentum vector, i.e., by a dot in the three-dimensional phase space. Different patterns formed by these dots in the phase space correspond to different dynamics. Understanding the dynamics is the main goal of all studies done at accelerators and in cosmic rays. Especially intriguing is the problem of quark-gluon plasma, the state of matter with deconfined quarks and gluons which could exist for extremely short time intervals. One hopes to create it in collisions of high energy nuclei. Nowadays, data about Pb – Pb collisions are available where, in a single event, more than 1000 charged particles are produced. We are waiting for the RHIC accelerator in Brookhaven and LHC in CERN to provide events with up to 20000 new particles created. However we do not know yet which patterns will be drawn by nature in individual events. Therefore the problem of phase space pattern recognition in an event-by-event analysis becomes meaningful.

It is believed that the detailed characterization of each collision event could reveal rare new phenomena, and it will be statistically reliable due to the large number of particles produced in a single event.

When individual events are imaged visually, the human eye has a tendency to observe different kinds of intricate patterns with dense clusters and rarefied voids. Combined with the search for maxima (spikes) in pseudorapidity [16] (polar angle) distributions, this has lead to indications of the so-called ring-like events in cosmic ray studies [60 – 64] and at accelerators [65, 66]. However, the observed effects are often dominated by statistical fluctuations. The method of factorial moments was proposed [67] to remove the statistical background but it is hard to apply in an event-by-event approach. Wavelet analysis avoids smooth polynomial trends and underlines the fluctuation patterns. By choosing the strongest fluctuations, one hopes to get those which exceed the statistical component. The wavelet transform of the pseudorapidity spectra of cosmic ray high multiplicity events of the JACEE collaboration was done in Ref. [60].

In Ref. [59] wavelets were first applied to analyze patterns formed in the phase space of the accelerator data on individual high multiplicity events of Pb-Pb interaction at an energy of 158 GeV per nucleon. With the emulsion technique used in the experiment only the angles of emission of particles are measured, and therefore the two-dimensional (polar + azimuthal angles) phase space is considered. The experimental statistics are rather low but acceptance is high and homogeneous which is important for the proper pattern recognition. To simplify the analysis, the two-dimensional target diagram representing the polar and azimuthal angles of created charged particles was split into 24 azimuthal angle sectors of $\pi/12$ and in each of them particles were projected onto the polar angle θ axis. Thus one-dimensional functions of the polar angle (pseudorapidity) distribution of these particles in 24 sectors were obtained. Then the wavelet coefficients were calculated in all of these sectors and connected together afterwards (the continuous MHAT wavelet was used). The resulting pattern showed that many particles are concentrated close to some value of the polar angle, i.e., reveal the ring-like structure in the target diagram. The interest in such patterns is related to the fact that they can result from so-called gluon Cherenkov radiation [68, 69] or, more generally, from gluon bremsstrahlung for a finite free path within quark-gluon medium (plasma, in particular).

More elaborate two-dimensional analysis with Daubechies (D^8) wavelets was done recently [70] and confirmed these conclusions with jet regions tending to lie on some ring-like formations. This is seen, e.g. in Fig. 16 where dark regions correspond to large wavelet coefficients of the large-scale particle fluctuations in two of the events analyzed. Only the resolution levels $6 \leqslant j \leqslant 10$ were left after the event analysis was done to store the long-range correlations in the events and get rid of short-range ones and background noise. Then the inverse restoration was done to get the event images only

[16] Pseudorapidity is defined as $\eta = -\log\tan(\theta/2)$, where θ is the polar angle of particle emission.

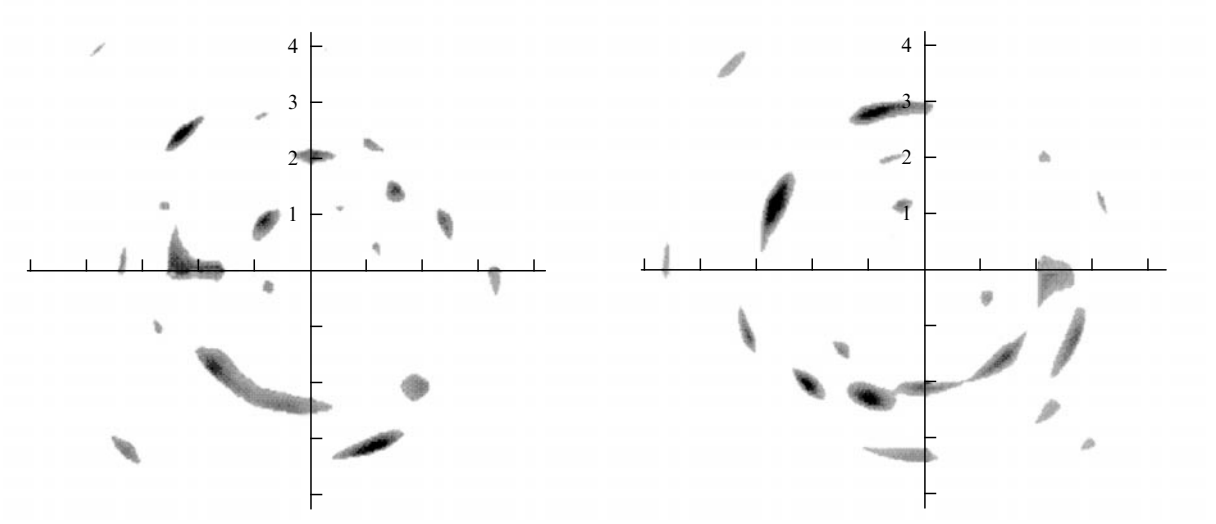

Figure 16. Restored images of long-range correlations in experimental target diagrams. They show the typical ring-like structure in some events of central Pb – Pb interactions.

with these dynamic correlations left, and this is what is seen in Fig. 16. It directly demonstrates that the large-scale correlations chosen have a ring-like (ridge) pattern. With larger statistics, one will be able to say if they correspond to theoretical expectations. However preliminary results favor positive conclusions [70]. This is due to the two-dimensional wavelet analysis that for the first time the fluctuation structure of an event is shown in a way similar to the target diagram representation of events on the two-dimensional plot.

Previously, some attempts [60, 71, 72] to consider such events with different methods of treating the traditional projection and correlation measures just revealed that such substructures lead to spikes in the polar angle (pseudorapidity) distribution and are somewhat jetty. Various Monte Carlo simulations of the process were compared to the data and failed to describe this jettiness in its full strength. More careful analysis [73, 74] of large statistics data on hadron-hadron interactions (unfortunately, however, for rather low multiplicity) with dense groups of particles separated showed some 'anomaly' in the angular distribution of these groups awaited from the theoretical side. Further analysis using the results of the wavelet transform is needed when many high multiplicity events become available. A more detailed review of this topic is given in Ref. [75].

15.2 Aviation (engines)

Multiresolution wavelet analysis has proved to be an extremely useful mathematical method for analyzing complicated physical signals at various scales and definite locations. It is tempting to begin with an analysis of signals depending on a single variable [17]. The time variation of the pressure in an aircraft compressor is one such signals. The aim of the analysis of this signal is motivated by the desire to find the precursors of a very dangerous effect (stall + surge) in engines leading to their destruction.

An axial multistage compressor is susceptible to the formation of a rotating stall which may be precipitated by a distorted inlet flow induced by abrupt manoeuvres of an aircraft (helicopter) or flight turbulence. The aerodynamic instability behind the rotating blades is at the origin of this effect. The instability regions called stall cells separate from the blades and rotate with a speed of about 60% of the rotor speed. Thus they are crossed by the blades approximately once every 1.6 rotor revolutions. This abruptly limits the flow and thus the pressure delivery to the downstream plenum (the combustion chamber) where the fuel is burnt and the gas jet is formed. The high prestall pressure build-up existing in the combustion chamber tends to push the flow backwards through the compressor. If the flow is reversed one calls it a 'deep surge'. In many cases this is attested by flames emerging from the engine inlet. It can have serious consequences for the engine life and operation, not to mention the aircraft and its passengers if it happens during flight. That is why the search for precursors of these dangerous phenomena is very important. Attempts to predict the development of a rotating stall and a subsequent surge with a velocity measuring probe such as a hot wire anemometer [76] provide a precursor or warning of only about 10 μs which is not enough for performing any operations which would preclude the surge development.

Multiresolution wavelet analysis of pressure variations in a gas turbine compressor reveals much earlier precursors of stall and surge processes [77]. Signals from 8 pressure sensors positioned at various places within the compressor were recorded and digitized for 3 different modes of operation (76, 81, 100% of nominal rotation speed) in stationary conditions with an interval of 1 ms for 5 – 6 s before the stall and the surge developed. An instability was induced by a slow injection of extra air into the compressor inlet. After a few minutes, this led to a fully blown instability in the compressor. The last 5 – 6 s interval was wavelet-analyzed. Since the signal fluctuates in time, so too does the sequence of wavelet coefficients at any given scale. A natural measure for this variability is the standard deviation (dispersion) of wavelet coefficients as a function of scale. A scale of $j = 5$ showed a remarkable drop of about 40% of the dispersion (variance) of the wavelet coefficients for more than 1 s prior to the malfunction's development. The dispersion is calculated according to the

[17] The trick of reducing the two-dimensional function to a set of one-dimensional ones was described in the previous subsection devoted to particle production.

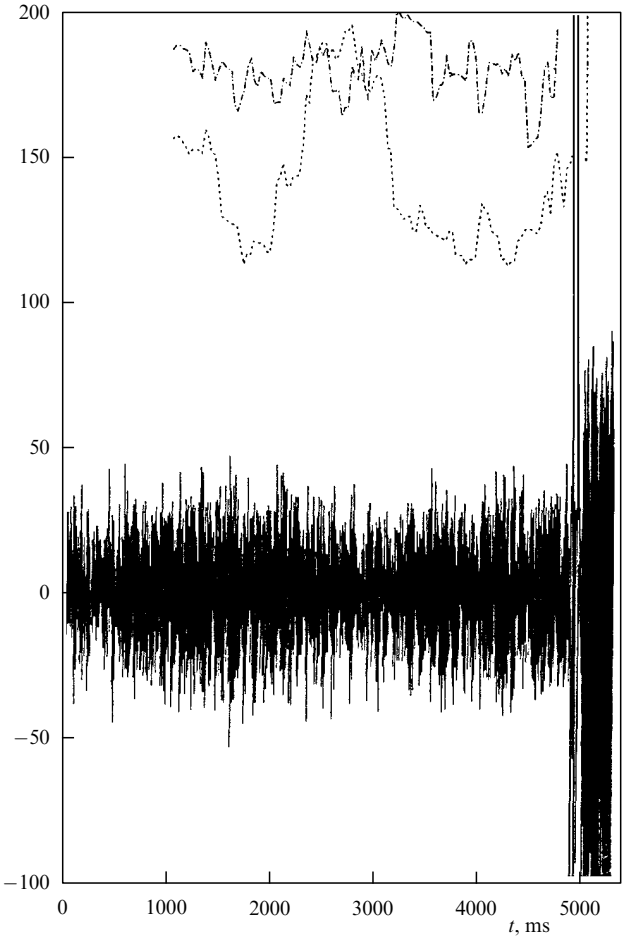

Figure 17. Signal of the pressure sensor (the solid line) and the dispersion of its wavelet coefficients (the dashed line). The time variation of the pressure in the engine compressor (the irregular solid line) has been wavelet analyzed. The dispersion of the wavelet coefficients (the dashed line) shows a maximum and a significant drop prior to the drastic increase of the pressure providing the precursor of this malfunction. The shuffled set of data does not show such an effect for the dispersion of the wavelet coefficients (the upper curve) pointing to its dynamic origin.

standard expression

$$\sigma(j, M) = \sqrt{\frac{1}{M-1} \sum_{k=0}^{M-1} \left[d_{j,k} - \langle d_{j,k} \rangle\right]^2}, \qquad (15.1)$$

where M is the number of wavelet coefficients at the scale j within some time interval which was chosen to be 1 s.

In Figure 17 the time variation of the pressure in the compressor in one of the modes of engine operation indicated above is shown by a very irregular line. The large fluctuations of the pressure on the right-hand side denote the surge onset. It is hard to see any warning of this drastic instability from the shape of this curve. The wavelet coefficients of this function were calculated with Daubechies 8-tap (D^8) wavelets. The dashed curve in the same figure shows the behavior of the dispersion of wavelet coefficients as a function of time at the scale level $j = 5$ which happens to be most sensitive to the surge onset [18]. Its precursor is seen as the maximum and the subsequent drop of the standard deviation by about 40% which appears about 1 – 2 s prior to the malfunction denoted by the large increase of both the pressure and dispersion on the right-hand side. The initial part of the dispersion plot is empty because the prescribed interval for the compilation of the representative initial distribution of wavelet coefficients is chosen to be 1 s. Namely, the time evolution of this distribution has been studied.

The randomly reordered (shuffled) sample of the same values of the pressure within the pre-surge time interval does not show such a drop of the dispersion as demonstrated by the dot-dashed line in the figure. This proves the dynamic origin of the effect. Further analyzed characteristics of this process (fractality, high-moments behavior) are discussed in [77]. No fractal properties of the signal have been found. The scale $j = 5$ violates the linearity of the in the sum $Z_q(j)$ function of j necessary for fractal behavior. Let us note that the multifractality of the signal may fail because of accumulation of points where chirp-like behavior happens. If it is the origin of this effect, it should provide a guide to the description of the dynamics of the analyzed processes. The time intervals before and after the appearance of the precursor were analyzed separately. Higher moments of wavelet coefficients as functions of their rank q behave in a different way at pre- and post-precursor time. This indicates the different dynamics in these two regions.

This method of wavelet analysis can be applied to any engines, motors, turbines, pumps, compressors etc. It provides a significant improvement in the diagnostics of the operating regimes of existing engines, which is important for preventing their failure and, consequently, for lowering associated economic losses. In particular, the effectiveness of stationary energetic installations is strongly lowered due to the necessity of limiting the range of their parameters because of combustion problems. The use of wavelet analysis can help in diagnosis and the prevention of undesirable effects. Two patents for diagnostics and automatic regulation of engines dated 19.03.1999 have been obtained by the authors of Ref. [77].

There are other problems in aviation which could be approached with wavelet analysis , e.g., combustion instabilities, analysis of ions in the jet from a combustion chamber or the more general problem of metal aging and cracks etc.

15.3 Medicine and biology

Applications of wavelet analysis to medicine and biology (see, e.g., [17, 78, 79]) also follow the same procedures as discussed above. They are either deciphering information hidden in one-dimensional functions (analysis of heartbeat intervals, ECG, EEG, DNA sequences etc) or pattern recognition (shapes of biological objects, blood cell classification etc).

15.3.1 Hearts. Intriguing results were obtained [80] when multiresolution wavelet analysis was applied to the sequence of time intervals between human heartbeats. It has been claimed that a *clinically* significant measure of the presence of heart failure can be found just from analysis of heartbeat intervals alone without knowledge of the full electrocardiogram plot while the previous approaches provided *statistically* significant measures only. Series of about 70000 interbeat intervals were collected for each of 27 patients treated. They were plotted versus interval number. Signal fluctuations were wavelet transformed and the dispersions of wavelet coefficients at different scales were calculated as in the case

[18] This shows that the correlations among successive values of the sampled signal are likely to occur at intervals of the order of 30 μs. Let us note that similar correlation lengths are typical for a sampled speech signal as well.

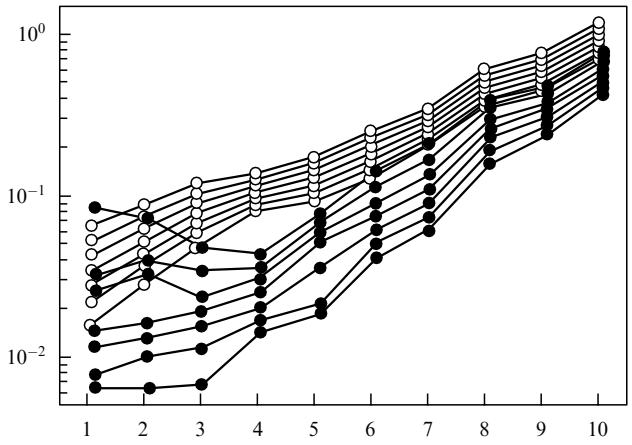

Figure 18. Sets of the values of the dispersions of wavelet coefficients for the heartbeat intervals of healthy (white circles) and heart failure (black circles) patients. They do not overlap at $j = 4$.

of aviation engines considered above but with averaging over the whole time interval because it is not now necessary to study the evolution during this time interval. Thus each patient was characterized by a single number (dispersion) at a given scale. It happened that the sets of numbers for healthy and heart failure patients did not overlap at the scale [19] $j = 4$ as seen in Fig. 18. This is considered as a clinically significant measure in distinction to the statistically significant measure when these sets overlap partly. The dispersions for healthy patients are larger probably corresponding to the higher flexibility of healthy hearts. The fluctuations are less anti-correlated for heart failure than for healthy dynamics.

This analysis was questioned by another group [81, 82] stating that the above results on the scale-dependent separation between healthy and pathologic behavior depend on the database analyzed and on the analyzing wavelet. They propose new scale-independent measures — the exponents characterizing the scaling of the partition function of the wavelet coefficients of the heartbeat records. These exponents reveal the multifractal behavior for a healthy human heartbeat and the loss of multifractality for a life-threatening condition, congestive heart failure. This distinction is ascribed to nonlinear features of the healthy heartbeat dynamics. The authors claim that the multifractal approach to wavelet analysis robustly discriminates healthy subjects from heart-failure subjects. In our opinion, further studies are needed.

Earlier, Fourier analysis of the heart rate variability disclosed the frequency spectrum of the process. Three frequency regions play the main role. The high frequency peak is located at about 0.2 Hz, and the low frequency peak is about 0.1 Hz. The superlow frequency component is reminiscent of $1/f$-noise with its amplitude strongly increasing at low frequencies. There have been some attempts [83] to develop theoretical models of such a process with the help of the wavelet transform.

The wavelet analysis of one-dimensional signals is also useful for deciphering the information contained in ECGs and EEGs. The functions to be analyzed there are more complicated than those in the above studies. Some very promising results have been obtained already. In particular, it was shown that anomalous effects in ECGs appear mostly at rather large scales (low frequencies) whereas the normal structures are inclined to comparatively small scales (high frequencies). This corresponds to the above results on heartbeat intervals. Such an analysis of ECGs was reported, e.g., in Ref. [84]. We shall not describe it in detail here because these studies are still only in their initial stages. The time-frequency analysis of EEGs can be found, e.g., in Ref. [85]. It can locate the source of epileptic activity and its propagation in the brain. For a general description of EEG methodics see, e.g., Refs [86, 87].

Wavelet analysis has been used also for diagnostics of the embryonic state during pregnancy [79]. A special pursuit method was developed to fit the properties of the signal as well as possible.

15.3.2 DNA sequences. The wavelet transform with continuous wavelets has been used for fractal analysis of DNA sequences [88, 89] in an attempt to reveal the nature and the origin of long range correlations in these sequences. It is still debated whether such correlations are different for protein-coding (exonic) and noncoding (intronic, intergenic) nucleotide sequences. To graphically portray these sequences in the form of one-dimensional functions, the so-called 'DNA walk' analysis was applied with the conversion of the 4-letter text of DNA into a binary set [90]. Applying the wavelet transform to 121 DNA sequences (with 47 coding and 74 noncoding regions) selected in the human genome, the authors of Refs [88, 89] have been able to show that there really exists a difference between the two subsequences. This was demonstrated by the presence of long range correlations in noncoding sequences while the coding sequences look like uncorrelated random walks. To show this, the averaged partition (generating) functions of wavelet coefficients [see Eqn (13.3)] over these two statistical samples were calculated. Both of them scaled with the exponent predicted for homogeneous Brownian motion, i.e., $\tau(q) = qH - 1$. The main difference between the noncoding and coding sequences is the value of H, namely, $H_{\mathrm{nc}} = 0.59 \pm 0.02$ and $H_{\mathrm{c}} = 0.51 \pm 0.02$ which distinguish uncorrelated and correlated subsamples. Moreover, it has been shown that $\tau(q)$ spectra extracted from both sets are surprisingly in remarkable agreement with the theoretical prediction for Gaussian processes if the probability density of wavelet coefficients at fixed scales are plotted versus these coefficients scaled with their r.m.s. value at the same scale. The results of wavelet analysis clearly show that the purine (A, G) versus pyrimidine (C, T) content [20] of DNA is likely to be relevant to the long range correlation properties observed in DNA sequences.

15.3.3 Blood cells. Another problem which can be solved with the help of wavelet analysis is the recognition of different shapes of biological objects. By itself, this has very wide range of applicability. Here, we consider the blood cell classification scheme according to the automatic wavelet analysis developed by us, and a particular illustration is given for erythrocyte classification. An automatic search, stability in determining the cells shapes and high speed of processing can be achieved with computer wavelet analysis. The main idea of the method relies on the fact that at a definite resolution scale wavelet analysis clearly reveals the contours of the blood cells

[19] By coincidence this is the same as for aviation engines.

[20] A, C, G, T denote the usual four letter alphabet of any DNA text: Adenine, Cytosine, Guanine, Thymine.

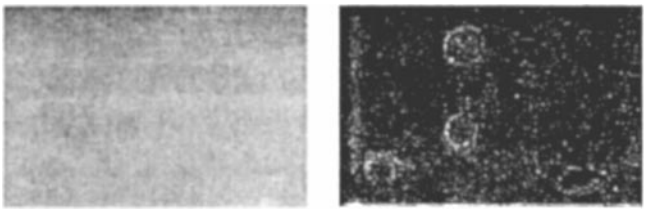

Figure 19. (a) The photo of blood cell, obtained from the microscope. (b) The same photo after the wavelet analysis done. The blurred image of a blood cell (left) becomes clearly visible one (right) after the wavelet transform.

which allows them to be classified. In Figure 19 we demonstrate how it became possible to improve the image resolution by such a method. The low quality image of blood cells has been transformed into its wavelet image with clearly seen cells contours.

The shapes of erythrocytes differ for various types. Depending on their shape, they can play either a positive or negative role for a human being. Therefore their differentiation is very important. After microscope analysis of a dry blood smear and registration of blood cells in a computer, the wavelet analysis of the set of registered blood cells was done.

However the extreme random irregularities at some points of a cell contour can prevent one from performing the analysis. Therefore a special smearing procedure was invented to avoid such points without loss of crucial peculiarities typical for a definite type of cell. After that, the set of blood cells with somewhat smeared contours is ready for wavelet analysis. It consists in wavelet correlation analysis which shows the different behavior of the correlation measures depending on the particular cell shapes. Using an expert classification of cells into a definite sample, the wavelet characteristics of correlations typical for a particular class were found and the then inserted in the computer program. This software was used for classification of blood cells obtained from other patients, and results were cross-checked by experts. Their positive conclusions are encouraging. The whole procedure is now done fast and automatically without human intervention. In Figure 20 blood cells of various kinds are shown.

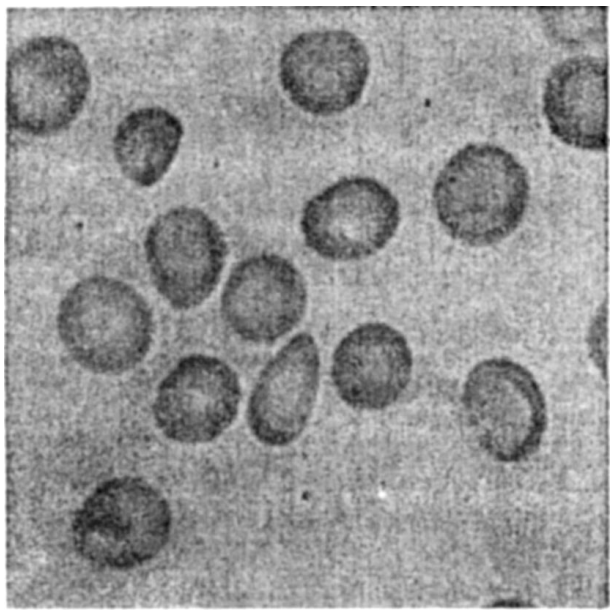

Figure 20. Classification of erythrocyte cells.

15.4 Data compression

Data compression is needed if one wants, e.g., to store data spending as little memory capacity as possible or to transfer it at a low cost using smaller packages. The example of the FBI using wavelet analysis for pattern recognition and saving in that way a lot of money on computer storage of fingerprints is well known. This is done by omitting small wavelet coefficients after the direct wavelet transform was applied. Surely, to restore the information one should be confident in the stability and good quality of the inverse transformation. Therefore, both analysis and synthesis procedures are necessary for data compression and its subsequent reconstruction. Above, we have shown how successful wavelet analysis has been in solving many problems. Due to the completeness of the wavelet system it is well suited for the proper inverse transform (synthesis) as well (see, e.g., Ref. [91]).

The approach to the solution of the problem strongly depends on the actual requirements imposed on the final outcome. There are at least three of them. If one wants to keep the quality of the restored image (film) practically as good as the initial one, the compression should not be very strong. This is required, e.g., if experts should not be able to distinguish the compressed and uncompressed copies of a movie when they are shown on two screens in parallel. Another requirement would be important if one wants to compress an image as strongly as possible leaving it still recognizable. This is required, e.g., if one needs to transfer the information in a line with limited capacity. Finally, one can require the whole procedure of analysis and synthesis to be done as fast as possible. This is necessary, e.g., if the information must be obtained immediately but at lower cost. These three situations require a different choice of wavelets to optimize analysis and synthesis. In all cases wavelet analysis has an advantage over the coding methods which use the windowed Fourier transform but the quantitative estimate of this advantage varies with the problem solved.

Let us recall that any image in a computer must be digitized and saved as a bitmap or, in other words, as a matrix, each element of which describes the color of the point in the original image. The number of elements of the matrix (image points) depends on the resolution chosen in the digital procedure. It is a bitmap that is used for the subsequent reproduction of the image on a screen, a printer etc. However, it is not desirable to store it in such a form because it would require a huge computer capacity. That is why, at present, numerous coding algorithms (compression) for a bitmap have been developed, whose effectiveness depends on the image characteristics [21]. All these algorithms belong to two categories — they either code with a loss of information or without any loss of it (in that case the original bitmap can be completely recovered by the decoding procedure). In more general applications, the latter algorithm is often called as the *data archival*.

[21] It is evident that to store a 'black square', one does not have to deal with a matrix of all black dots but just store the three numbers showing the width, the height and the color. This is the simplest example of image compression.

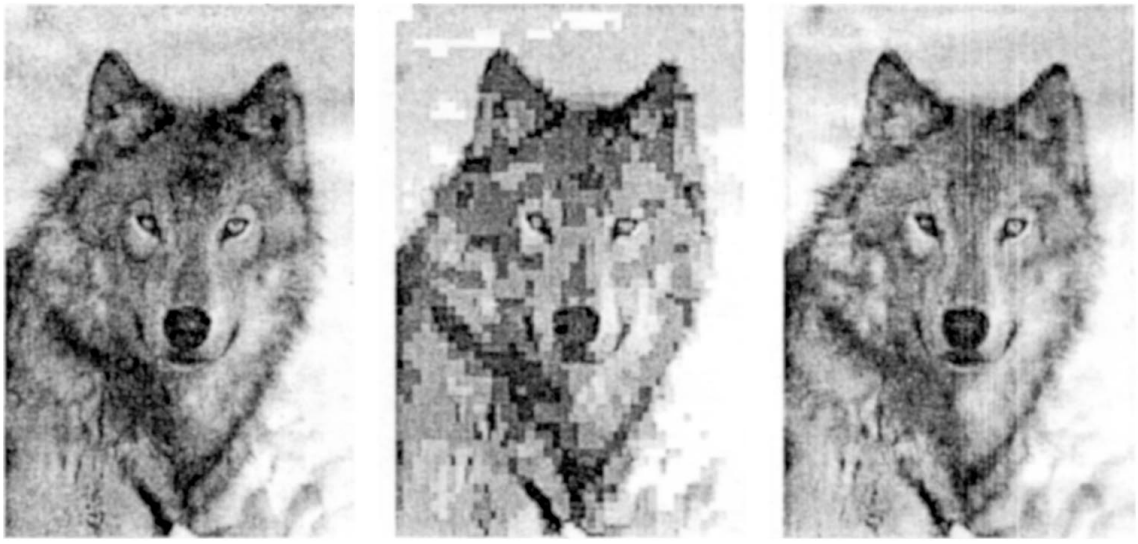

Figure 21. (a) Original photo (the file size is 461760 bytes). (b) Photo reconstructed after compression according to the JPEG-algorithm. (the file size is 3511 bytes). (c) Photo reconstructed after compression according to the wavelet algorithm (the file size is 3519 bytes). The better quality of the wavelet transform is clearly seen when comparing the original image (left) and the two images restored after similar compression by the windowed Fourier transform (middle) and the wavelet transform (right).

As an example, we consider a compression algorithm with the loss of information. The 'loss of information' means in this case that the restored image is not completely identical to the original one but this difference is practically indistinguishable by the human eye [22].

At present, most computer stored images (in particular, those used in Internet) with continuous tone [23] are coded with the help of the algorithm JPEG (for a detailed description, see [92]). The main stages of this algorithm are as follows. One splits the image into matrices with 8 by 8 dots. For each matrix, a discrete cosine-transform is performed. The obtained frequency matrices are subjected to a so-called *quantization* procedure when the most crucial elements for the visual recognition are chosen according to a special 'weight' table compiled beforehand by specialists after their collective decision was achieved. This is the only stage where the 'loss' of information occurs. Then the transformed matrix with the chosen frequencies (scales) is compacted and coded by the so-called entropy method (also called arithmetic or Huffmann method).

The algorithm applied above differs from that described by use of the wavelets instead of the windowed cosine-transform and by the transform of the whole image instead of an 8 by 8 matrix only. Figure 21 demonstrates the original image and the two final images restored after similar compression according to JPEG and wavelet algorithms. It is easily seen that the quality of the wavelet image is noticeably higher than for JPEG for practically the same size of the coded files. The requirement of the same quality for both algorithms leads to file sizes 1.5 – 2 times smaller for the wavelet algorithm that could be crucial for the transmission of the image, especially if the transmission line capacity is limited.

[22] The 'compression quality' is usually characterized by a parameter which varies from 0 to 100. Here 100 implies minimal compression (the best quality), and the restored image is practically indistinguishable from the initial one, while 0 means the maximum compression which still allows one to distinguish some details of the original image in the recovered one.

[23] That is the images contain many slightly different colors. Ordinarily to store such images in a computer, 16 million colors per pixel are used.

15.5 Microscope focusing

Surely, the problem of microscope focusing is tightly related to pattern recognition. One should resolve a well focused image from that with diffused contours. It is a comparatively easy task for wavelets because in the former case the image gradients at the contour are quite high while in the latter they become rather vague. Therefore the wavelet coefficients are larger when the microscope is well focused on the object and drastically decrease with defocusing. At a definite resolution level corresponding to the contour scale the defocusing effect is strongest. This is demonstrated in Fig. 22. The peak of the wavelet coefficients shows the most focused image. After doing the wavelet analysis of an image, the computer sends a command to shift the microscope so that larger values of the wavelet coefficients and thus better focusing are attained. At other levels it is somewhat less pronounced and, moreover, it is asymmetric depending on whether the microscope is positioned above the focus location or below it. This asymmetry has been used for automatic microscope focusing with a well defined direction of movement toward the focus location.

16. Conclusions

The beauty of the mathematical construction of the wavelet transformation and its utility in practical applications attract researchers from both pure and applied science. Moreover, the commercial outcome of this research has become quite important. We have outlined a minor part of the activity in this field. However we hope that the general trends in the development of this subject become comprehended and appreciated.

The unique mathematical properties of wavelets make them a very powerful tool in analysis and subsequent synthesis of any signal. The orthogonality property allows one to get independent information from different scales.

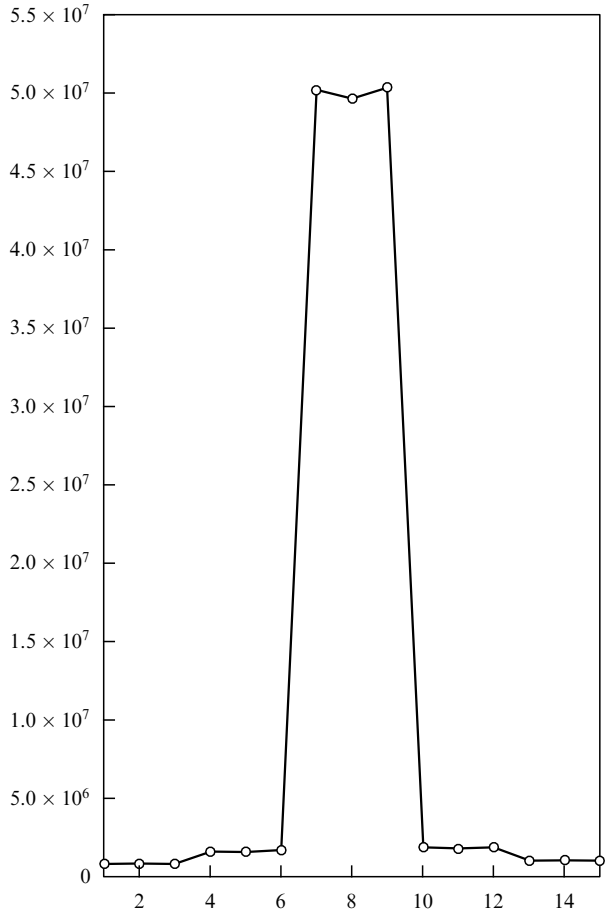

Figure 22. Focusing line. The values of the wavelet coefficients calculated for different positions of the microscope are shown on the vertical axis. The corresponding positions are numbered on the horizontal axis. The best focusing is obtained at the maximum of the curve (positions 6–10). Large wavelet coefficients correspond to a better focused image. The microscope moves to the focus position being driven by computer commands to increase of the wavelet coefficients of the analyzed objects.

Normalization assures that at different steps of this transformation the value of information is not changed and mixed up. The property of localization helps get knowledge about those regions in which definite scales (frequencies) play the most important role. Finally, the completeness of the wavelet basis formed by dilations and translations results in the validity of the inverse transformation.

Multiresolution analysis leads to the fast wavelet transform and together with the procedure of nonstandard matrix multiplication to rather effective computing. Analytic properties of functions, local and global Hölder indices, multifractal dimensions etc. are subject to wavelet analysis. The natural accommodation of differential operators in this environment opens the way to an effective solution of differential equations.

All these properties enable us to analyze complex signals at different scales and locations through the wavelet transform, and to solve equations describing extremely complicated nonlinear systems involving interactions at many scales, to study strongly singular functions etc. The wavelet transform is easily generalized to sets of any dimension, and multidimensional objects can be analyzed as well. Therefore, wavelets are indispensable for pattern recognition.

Thus the wavelet applications in various fields are numerous and nowadays give a very fruitful outcome. In this review paper we managed to describe only some of these applications leaving aside the main bulk of them. The potentialities of wavelets are still not used at their full strength.

However one should not cherish vain hopes that this machinery works automatically in all situations using its internal logic and does not require any intuition. According to Meyer [4], "no 'universal algorithm' is appropriate for the extreme diversity of the situations encountered". Actually it needs a lot of experience in choosing the proper wavelets, in a suitable formulation of the problem under investigation, in considering the most important scales and characteristics describing the analyzed signal, in the proper choice of the algorithms (i.e., the methodology) used, in studying the intervening singularities, in avoiding possible instabilities etc. By this remark we would not like to discourage newcomers from entering the field but, quite to the contrary, to attract those who are not afraid of hard but exciting research and experience.

17. Appendix

17.1 Multiresolution analysis

The general approach which respects all properties required from wavelets is known as the multiresolution approximation and is defined mathematically in the following way [4]:

A multiresolution approximation of $L^2(R^n)$ is, by definition, an increasing sequence V_j, $j \in Z$, of closed linear subspaces of $L^2(R^n)$ with the following properties:

(1) $\bigcap_{-\infty}^{\infty} V_j = \{0\}$, $\bigcup_{-\infty}^{\infty} V_j$ is dense in $L^2(R^n)$;

(2) for all $f \in L^2(R^n)$ and all $j \in Z^n$

$$f(x) \in V_j \leftrightarrow f(2x) \in V_{j+1}\,;$$

(3) for all $f \in L^2(R^n)$ and all $k \in Z^n$

$$f(x) \in V_0 \leftrightarrow f(x-k) \in V_0\,;$$

(4) there exists a function, $g(x) \in V_0$, such that the sequence $g(x-k)$, $k \in Z^n$, is an orthonormal (or Riesz) basis of the space V_0. It is clear, that scaling versions of the function $g(x)$ form bases for all V_j.

Thus the multiresolution analysis consists of a sequence of successive approximation spaces V_j which are scaled and invariant under integer translation versions of the central space V_0. There exists an orthonormal or, more generally, Riesz basis in this space. The Haar multiresolution analysis can be written in these terms as

$$V_j = \left\{ f \in L^2(R)\,; \text{ for all } k \in Z \;\; f\big|_{[k2^{-j},\,(k+1)2^{-j}]} = \text{const} \right\}. \tag{17.1}$$

In Figure 3 we showed what the projections of some f on the Haar spaces V_0, V_1 might look like. The general distributions are decomposed into a series of correctly localized fluctuations of a characteristic form defined by the wavelet chosen.

The functions $\varphi_{j,k}$ form an orthonormal basis of V_j. The orthogonal complement of V_j in V_{j+1} is called W_j. The subspaces W_j form a mutually orthogonal set. The sequence of $\psi_{j,k}$ constitutes an orthonormal basis for W_j at any definite

j. The whole collection of $\psi_{j,k}$ and $\varphi_{j,k}$ for all j is an orthonormal basis for $L^2(R)$. This ensures us that we have constructed a multiresolution analysis approach, and the functions $\psi_{j,k}$ and $\varphi_{j,k}$ constitute the small and large scale filters, correspondingly. The whole procedure of multiresolution analysis has been demonstrated in the graphs of Fig. 4.

In accordance with the above formulated goal, one can define the notion of wavelets in the following way [4]:

A function $\psi(x)$ of a real variable is called a (basic) *wavelet of class m* if the following four properties hold:

(1) if $m = 0$, $\psi(x)$ and $\varphi(x)$ belong to $L^\infty(R)$; if $m \geqslant 1$, $\psi(x)$, $\varphi(x)$ and all their derivatives up to order m belong to $L^\infty(R)$;

(2) $\psi(x)$, $\varphi(x)$ and all their derivatives up to order m decrease rapidly as $x \to \pm\infty$;

(3) $\int_{-\infty}^{\infty} x^n \psi(x)\,\mathrm{d}x = 0$ for $0 \leqslant n \leqslant m$, and $\int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x = 1$;

(4) the collection of functions $2^{j/2}\psi(2^j x - k)$, $2^{j/2}\varphi(2^j x - k)$, $j, k \in Z$, is an orthonormal basis of $L^2(R)$.

Then equation (3.17) is valid. If ψ and φ both have compact support, then it gives a decomposition of any distribution of order less than m. Moreover, the order of the distribution f (the nature of its singularities) can be calculated exactly and directly from the size of its wavelet coefficients as has been shown in Sections 11, 12. The functions $2^{j/2}\psi(2^j x - k)$ are the wavelets (generated by the 'mother' ψ), and the conditions 1), 2), 3) express, respectively, the regularity, the localization and the oscillatory character. One sees that property 3) is satisfied for the Haar wavelets for $m = 0$ only, i.e., its regularity is $r = 0$. In general, for each integer $r \geqslant 1$, there exists a multiresolution approximation V_j of $L^2(R)$ which is r-regular and such that the associated real-valued functions φ and ψ have compact support. As the regularity increases, so do the supports of φ and ψ. The wavelet $2^{j/2}\psi(2^j x - k)$ is 'essentially concentrated' on the dyadic interval $I = [k2^{-j}, (k+1)2^{-j}]$ for $j, k \in Z$. Its Fourier transform is supported by $2^j(2\pi/3) \leqslant |\omega| \leqslant 2^j(8\pi/3)$. In fact, it has a one octave frequency range.

17.2. Calderon – Zygmund operators

In the wavelet analysis of operator expressions the integral Calderon – Zygmund operators are often used. There are several definitions of them (see, e.g., the monograph [3]). We give here the definition used by Daubechies [2].

A *Calderon – Zygmund operator* T on R is an integral operator

$$Tf(x) = \int K(x,y) f(y)\,\mathrm{d}y\,, \tag{17.2}$$

for which the integral kernel satisfies

$$|K(x,y)| \leqslant \frac{C}{|x-y|}\,, \tag{17.3}$$

$$\left|\frac{\partial}{\partial x} K(x,y)\right| + \left|\frac{\partial}{\partial y} K(x,y)\right| \leqslant \frac{C}{|x-y|^2} \tag{17.4}$$

and which defines a bounded operator on $L^2(R)$.

With such a definition, special care should be taken at $x = y$.

17.3. Relation to the Littlewood – Paley decomposition

In the literature devoted to signal processing, the so-called Littlewood-Paley decomposition is often used. It is closely related to the wavelet transform. Therefore we give here the dictionary which relates the Littlewood – Paley coefficients denoted as $\Delta_j(f)(x)$ to both discrete $(d_{j,k})$ and continuous $(W(a,b))$ wavelet coefficients.

$$\Delta_j(f)(x) = 2^{nj/2} d_{j,2^j x} = W(2^{-j}, x)\,, \tag{17.5}$$

$$d_{j,k} = 2^{-nj/2} W(2^{-j}, 2^{-j}k) = 2^{-nj/2}\Delta_j(f)(2^{-j}k)\,. \tag{17.6}$$

Acknowledgements. We are grateful to all our colleagues with whom we worked on wavelet problems. We especially thanks A V Leonidov who read this paper and made important suggestions. We thank A Martin who pointed out some omissions in the References. O Ivanov is especially grateful to S Goedecker with whom many methods were developed and discussed. V Nechitaĭlo is also grateful to INTAS for support (grant 97-103).

References

1. Meyer Y *Wavelets and Operators* (Cambridge: Cambridge Univ. Press, 1992)
2. Daubechies I *Ten Lectures on Wavelets* (Philadelphia: SIAM, 1991)
3. Meyer Y, Coifman R *Wavelets, Calderon – Zygmund and Multilinear Operators* (Cambridge: Cambridge Univ. Press, 1997)
4. Meyer Y *Wavelets: Algorithms and Applications* (Philadelphia: SIAM, 1993)
5. *Progress in Wavelet Analysis and Applications* (Eds Y Meyer, S Roques) (Gif-sur-Yvette: Editions Frontiers, 1993)
6. Chui C K *An Introduction to Wavelets* (San Diego: Academic Press, 1992)
7. Hernandez E, Weiss G *A First Course on Wavelets* (Boca Raton: CRC Press, 1997)
8. Kaiser G A *Friendly Guide to Wavelets* (Boston: Birkhauser, 1994)
9. *Wavelets: An Elementary Treatment of Theory and Applications* (Ed T Koornwinder) (Singapore: World Scientific, 1993)
10. Astaf'eva N M *Usp. Fiz. Nauk* **166** 1145 (1996) [*Phys. Usp.* **39** 1085 (1996)]
11. Carmona R, Hwang W-L, Torresani B *Practical Time-Frequency Analysis* (San Diego: Academic Press, 1998)
12. Grossman A, Morlet J "Decomposition of functions into wavelets of constant shape, and related transforms", in *Mathematics + Physics, Lectures on Recent Results* Vol. 1 (Ed. L Streit) (Singapore: World Scientific, 1985)
13. Morlet J, Arens G, Fourgeau E, Giard D *Geophysics* **47** 203, 222 (1982)
14. *Wavelets in Physics* (Ed. J C Van den Berg) (Cambridge: Cambridge Univ. Press, 1999)
15. Mallat S *A Wavelet Tour of Signal Processing* (San Diego: Academic Press, 1998)
16. Erlebacher G, Hussaini M Y, Jameson L M *Wavelets: Theory and Applications* (New York: Oxford Univ. Press, 1996)
17. *Wavelets in Medicine and Biology* (Eds A Aldroubi, M Unser) (Boca Raton: CRC Press, 1996)
18. Haar A *Math. Ann.* **69** 331 (1910)
19. Sweldens W *Appl. Comput. Harmon. Anal.* **3** 186 (1996)
20. Goedecker S, Ivanov O V *Comput. Phys.* **12** 548 (1998)
21. Ausher P *Ondelettes Fractales et Applications* (Paris: Univ. Paris, Dauphine, 1989)
22. Beylkin G *SIAM J. Numer. Anal.* **6** 1716 (1992)
23. Beylkin G, Coifman R, Rokhlin V *Commun. Pure Appl. Math.* **44** 141 (1991)
24. Jaffard S, Meyer Y *Mem. Am. Math. Soc.* **123** 587 (1996)
25. Dremin I M *Phys. Lett. B* **313** 209 (1993)
26. Dremin I M, Hwa R C *Phys. Rev. D* **49** 5805 (1994)
27. Dremin I M, Gary J W *Phys. Rep.* **349** 301 (2001); hep-ph/0004215
28. Mallat S, Hwang W L *IEEE Trans. Inform. Theory* **38** 617 (1992)
29. Mandelbrot B B *The Fractal Geometry of Nature* (San Francisco: W.H. Freeman, 1982)
30. Feder J *Fractals* (New York: Plenum Press, 1988)
31. Paladin G, Vulpiani A *Phys. Rep.* **156** 147 (1987)

32. Dremin I M *Usp. Fiz. Nauk* **160** (8) 105 (1990) [*Sov. Phys. Usp.* **33** 647 (1990)]
33. Renyi A *Probability Theory* (Amsterdam: North-Holland, 1970)
34. Muzy J F, Bacry E, Arneodo A *Phys. Rev. Lett.* **67** 3515 (1991); *Int. J. Bifurcat. Chaos* **4** 245 (1994)
35. Arneodo A, d'Aubenton-Carafa Y, Thermes C *Physica D* **96** 291 (1996)
36. De Wolf E A, Dremin I M, Kittel W *Phys. Rep.* **270** 1 (1996)
37. Takayasu H *Fractals in the Physical Sciences* (Manchester: Manchester Univ. Press, 1990)
38. Jaffard S *Multifractal Formalism for Functions* (Philadelphia: SIAM, 1997)
39. Donoho D, Johnstone I *J. Am. Stat. Assoc.* **90** 1200 (1995)
40. Dremin I M, Ivanov O V, Nechitaĭlo V A *Nauka – Proizvodstvu* (6) 13 (2000)
41. Spiridonov V P *Komp'yuterra* (8) 38 (1998)
42. Torresani B, in *Self-similar Systems* (Eds V B Priezzhev, V P Spiridonov) (Dubna: JINR, 1999)
43. Bowman C, Newell A C *Rev. Mod. Phys.* **70** 289 (1998)
44. Parr R G, Yang W *Density-Functional Theory of Atoms and Molecules* (New York: Oxford Univ. Press, 1989)
45. Goedecker S, Ivanov O V *Solid State Commun.* **105** 665 (1998)
46. Wei S, Chou M Y *Phys. Rev. Lett.* **76** 2650 (1997)
47. Han S, Cho K, Ihm J *Phys. Rev. B* **60** 1437 (1999)
48. Tymczak C J, Wang X *Phys. Rev. Lett.* **78** 3654 (1997)
49. Ivanov O V, Antropov V P *J. Appl. Phys.* **85** 4821 (1999)
50. Ivanov O V *Phys. Rev. B* (2001) (to be published)
51. Goedecker S, Ivanov O V *Phys. Rev. B* **59** 7270 (1999)
52. Carruthers P, in *Proc. of Hot and Dense Matter, Bodrum, 1993* (Singapore: World Scientific, 1994) p. 65
53. Lipa P, Greiner M, Carruthers P, in *Proc. of Soft Physics and Fluctuations, Crakow, 1993* (Singapore: World Scientific, 1994) p. 105
54. Greiner M et al. *Z. Phys. C* **69** 305 (1996)
55. Suzuki N, Biyajima M, Ohsawa A *Prog. Theor. Phys.* **94** 91 (1995)
56. Huang D *Phys. Rev. D* **56** 3961 (1997)
57. Sarcevic I, Huang Z, Thews R *Phys. Rev. D* **54** 750 (1996)
58. Nandi B K et al. (WA98 Collab.), in *Proc. of 3rd Int. Conference on Physics and Astrophysics of Quark-Gluon Plasma, Jaipur, 1997* (Singapore: World Scientific, 1998) p. 12
59. Astafyeva N M, Dremin I M, Kotelnikov K A *Mod. Phys. Lett. A* **12** 1185 (1997)
60. Apanasenko A V, Dobrotin N A, Dremin I M et al. *Pis'ma Zh. Eksp. Teor. Fiz.* **30** 157 (1979) [*JETP Lett.* **30** 145 (1979)]
61. Alekseeva K I et al. *Izv. Akad. Nauk SSSR* **26** 572 (1962); *J. Phys. Soc. Jpn.* **17** 409 (1962)
62. Maslennikova N V et al. *Izv. Akad. Nauk SSSR* **36** 1696 (1972)
63. Arata N *Nuovo Cimento A* **43** 455 (1978)
64. Dremin I M, Orlov A M, Tretyakova M I *Proc. 17 ICRC* **5** 149 (1981); *Pis'ma Zh. Eksp. Teor. Fiz.* **40** 320 (1984) [*JETP Lett.* **40** 1115 (1984)]
65. Marutyan I A et al. *Yad. Fiz. 29* 1566 (1979)
66. Adamus M et al. (NA22 Collab.) *Phys. Lett. B* **185** 200 (1987)
67. Bialas A, Peschanski R *Nucl. Phys. B* **273** 703 (1988)
68. Dremin I M *Pis'ma Zh. Eksp. Teor. Fiz.* **30** 152 (1979) [*JETP Lett.* **30** 140 (1979)]
69. Dremin I M *Yad. Fiz.* **33** 1357 (1981)
70. Dremin I M, Ivanov O V, Kalinin S A et al. *Phys. Lett. B* **499** 97 (2001)
71. Adamovich M I et al. (EMU01 Collab.) *J. Phys. G* **19** 2035 (1993)
72. Cherry M L et al. (KLM Collab.) *Acta Phys. Pol. B* **29** 2129 (1998)
73. Dremin I M, Lasaeva P L, Loktionov A A et al. *Yad. Fiz.* **52** 840 (1990) [*Sov. J. Nucl. Phys.* **52** 840 (1990)]; *Mod. Phys. Lett. A* **5** 1743 (1990)
74. Agababyan N M et al. *Phys. Lett. B* **389** 397 (1996)
75. Dremin I M *Usp. Fiz. Nauk* **170** 1235 (2000) [*Phys. Usp.* **43** 1137 (2000)]
76. Georgantas A A *Review of Compressor Aerodynamic Instabilities* (Canada: National Aeronautical Establishment, 1994)
77. Dremin I M, Furletov V I, Ivanov O V et al. *Control Eng. Pract.* (2001) (to be published)
78. Akay M *IEEE Spectrum* **34** 50 (1997)
79. Lambrou T, Linney A, Speller R *Komp'yuterra* (6) 50 (1998)
80. Thurner S, Feurstein M C, Teich M C *Phys. Rev. Lett.* **80** 1544 (1998)
81. Amaral L A N, Goldberger A L, Ivanov P C et al. *Phys. Rev. Lett.* **81** 2388 (1998)
82. Ivanov P C, Amaral L A N, Goldberger A L et al. *Nature* **399** 461 (1999)
83. Yang F, Liao W *IEEE Eng. Med. Biol.* **16** 17 (1997)
84. Ivanov P C et al. *Nature* **383** 323 (1996)
85. Blanco S et al. *IEEE Eng. Med. Biol.* **16** 64 (1997)
86. Schiff S J et al. *Electroen. Clin. Neurophysiol.* **91** 442 (1994)
87. Blanco S et al. *Phys. Rev. E* **54** 6661 (1996)
88. Arneodo A et al. J F *Phys. Rev. Lett.* **74** 3293 (1995)
89. Arneodo A et al. *Physica A* **254** 24 (1998)
90. Peng C K et al. *Nature* **356** 168 (1992)
91. Antonini M, Barlaud M, Mathieu P, Daubechies I *IEEE Trans. on Image Process.* **1** 205 (1992)
92. Wallace G K, in *Compression Standard, Communications of the ACM*, *April 1991*, ftp://ftp.uu.net/graphics/jpeg/wallace.ps.gz