Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 13 January 1999
Accepted 23 September 1999

Fixed-scale wavelet-type approximation of periodic density distributions

V. Y. Lunin

Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region, 142292 Russia. Correspondence e-mail: lunin@impb.psn.ru

For a chosen unit cell, a function defined in real space (a standard signal) is considered as a crystallographic wavelet-type function if it is localized in a small region of the real space, if its Fourier transform is likewise localized in reciprocal space, and if it is a periodical function which possesses a symmetry. The fixed-scale analysis consists in the decomposition of a studied distribution into a sum of copies of the same standard signal, but shifted into nodes of a grid in the unit cell. For a specified standard signal and grid of the permitted shifts in the unit cell, the following questions are discussed: whether an arbitrary function may be represented as the sum of the shifted standard signals; how the coefficients in the decomposition are calculated; what is the best fixed-scale approximation in the case that the exact decomposition does not exist. The interrelations between the fixed-scale decomposition and the phase problem, automatic map interpretation and density-modification methods are pointed out.

1. Introduction

The wavelet transform (see *e.g.* Combes *et al.*, 1988; Chu, 1992; Daubechies, 1993) was found to be a useful tool in different applications for signal analysis, data compression and image processing. Nevertheless, with a few exceptions (Pouligny *et al.*, 1991; Main, 1996; Ferrer *et al.*, 1998), this tool has not yet found the proper place in the field of crystallography. At the same time, the basic ideas of the wavelet transform are very natural for crystallography and are present implicitly in a number of crystallographic approaches. Hence, the use of the advanced mathematical tools for wavelet analysis may permit new possibilities for crystallographic applications and improve the corresponding computer procedures. The wavelet transform can be viewed (Daubechies, 1993) as a synthesis of ideas that have emerged since the 60's from mathematics, physics and electrical engineering. Nevertheless, precursors of wavelet-transform ideas can be found much earlier in different branches of science and in crystallography in particular. The basic idea of wavelet analysis, namely to use a set of standard blocks (wavelets) of different sizes to represent the object, is very close (if not identical) to that of representing an electron-density distribution as a sum of contributions of atoms. Some additional requirements which are usually applied to the set of chosen blocks refine this rather vague idea. The main demands that are usually applied to the wavelets is that they are localized or have a fast decay in both real and reciprocal spaces (or in both time and frequency in other applications). The fast decay in reciprocal space usually gives them their characteristic appearance with 'ripples' in real space which clarifies the use of the term wavelet.

Wavelet analysis is a technique that consists in decomposing an arbitrary signal into elementary contributions that are constructed from one standard signal (wavelet) by means of dilations and translations. The values of the dilations and translations can be chosen to vary continuously or to be restricted to a discrete lattice. We shall restrict our consideration to the decomposition of a real-space function into a weighted sum of shifted copies of the same standard signal (wavelet) without additional dilations. We use the term wavelet here to emphasize our interest to space-localized contributions. Nevertheless, most of the results are valid for an arbitrary standard signal too. Such a decomposition may be considered as one step in the full multiresolution wavelet analysis. A peculiarity of the use of wavelet analysis in crystallography (which distinguishes it from the usual applications of wavelet analysis) is that here we have to work simultaneously with two scales. The first one is the scale of the unit-cell dimensions – the object function is periodic and it is natural to require the same for the building blocks. On the other hand, the scale of the details we try to study is much finer than the unit-cell dimensions, so the blocks that are used must be localized in a region of the unit cell that is sufficiently small compared with the period. We use the term crystallographic wavelet analysis to emphasize that the standard signal is presumed to satisfy the periodicity of the crystal space and, when necessary, appropriate crystallographic symmetries. To satisfy these requirements, some modification is introduced into the basic approaches of wavelet analysis (Daubechies, 1993). We start in §§2 and 3 with the one-dimensional case to demonstrate the basic ideas and approaches. General multi-dimensional algorithms are presented and necessary theorems

are proved in Appendix *A*. §4 presents some applications of fixed-scale decompositions in macromolecular crystallography.

One of the most important advantages of wavelet analysis is that it provides an efficient means of data compression. When a suitable wavelet basis is chosen for a class of objects of similar nature, it often occurs that only a small number of wavelet coefficients have significant values, while the others are negligible. So a small number of strong wavelet coefficients need to be stored to represent the object accurately. These algorithms were applied recently to data compression for diffraction patterns (Ferrer *et al.*, 1998). We consider in §4 the problem of data compression from a different point of view with its relation to the phase problem. It is demonstrated with a test that it is enough to have the number of real wavelets close to the number of the measured intensities to represent a low-resolution Fourier synthesis with an accuracy close to that of the MIR-phased synthesis. It must be noted that, contrary to some other methods of parameterization of low-resolution syntheses (Lunin *et al.*, 1995, 1998), the wavelet coefficients are related to the structure factors by linear relationships.

Another area of wavelet transform applications is data filtering, where once again the idea that the strongest wavelet coefficients are mostly reliable is used. In such a case, suppression of small wavelet coefficients with subsequent calculation of the modified function may reduce the noise. A crystallographic example of such filtering consists of procedures for phase improvement in which a noisy Fourier synthesis is interpreted in terms of atomic coordinates and then the atom-based representation of the electron density is used to calculate improved phases (Agarwal & Isaacs, 1977; Lunin & Urzhumtsev, 1984; Lunin *et al.*, 1985; Lamzin & Wilson, 1993; Vellieux, 1998). It is worth noting that in usual applications of wavelet-based filtering the smallest wavelet coefficients are frequently the ones corresponding to the highest scale, and they are often assumed as corresponding to a noise. Nevertheless, for macromolecular Fourier syntheses, there exists another possibility: when small wavelet coefficients correspond to wavelets placed into the solvent region. Obviously, the second type of filtering may only be achieved by means of fixed-scale wavelet analysis. As is shown below, grid-restricted values in a Fourier synthesis may be considered to some extent as wavelet coefficients. So the methods that use these values as primary variables (Urzhumtsev, 1997; Somoza *et al.*, 1998) and different density-modification methods (Podjarny *et al.*, 1996) may be formulated in terms of wavelet analysis too.

If the object function possesses some crystallographic space-group symmetry, then the wavelet function may be required to have a symmetry too. It is shown in Appendix *A* that there exists a kind of duality in the symmetry of the wavelet function and the wavelet coefficients of the object function. If the wavelet has the space-group symmetry, then the wavelet coefficients have the point-group symmetry and *vice versa*.

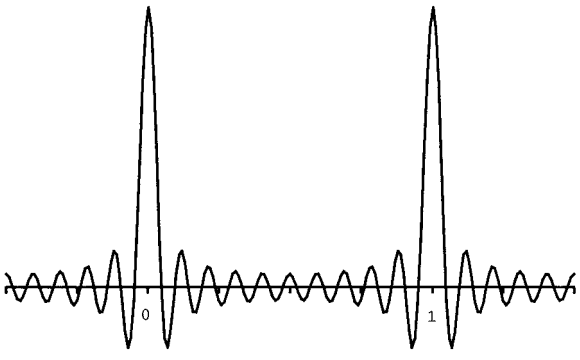

Figure 1
The wavelet (25) for $P = 10$.

2. One-dimensional fixed-scale decomposition

We start with the one-dimensional case to elucidate the main problems and ideas of fixed-scale analysis of the periodical functions. The multidimensional case is covered more thoroughly in Appendix *A*.

2.1. Definitions

We consider here functions $f(x)$ depending on one real variable x and suppose implicitly that every function considered is a periodic one with the period 1:

$$f(x+1) = f(x) \quad \text{for all real } x. \tag{1}$$

Furthermore, we suppose that $f(x)$ is a real-valued function, *i.e.* its structure factors display Hermitian symmetry:

$$\hat{f}(-h) = \overline{\hat{f}(h)} \quad \text{for all } h. \tag{2}$$

Here and below, $\hat{f}$ stands for structure factors of the function f, *i.e.*

$$\hat{f}(h) = \int_0^1 f(x) \exp[2\pi i h x]\, \mathrm{d}x, \quad h \in \mathbf{Z}, \tag{3}$$

$$f(x) = \sum_{h \in \mathbf{Z}} \hat{f}(h) \exp[-2\pi i h x], \quad x \in \mathbf{R}, \tag{4}$$

where $\mathbf{R}$ and $\mathbf{Z}$ denote the sets of all real and all integer numbers, respectively.

2.2. Wavelets

Let $\zeta(x)$ be a real periodic function which will be used below as a 'building block' in the decomposition of a signal $f(x)$. We call $\zeta(x)$ the standard signal in general, or the standard wavelet if $\zeta(x)$ may be considered to be localized in a vicinity of the origin, *i.e.* it has significant values at points close to the origin and negligible values at points remote from the origin. (Naturally, for periodic functions, we consider as the origin not the point $x = 0$ alone but all integer points.) Some examples of what we call wavelets are shown in Figs. 1 and 2. Similar to (3)–(4), the standard signal $\zeta(x)$ is defined uniquely by its structure factors $\hat{\zeta}(h)$, $h \in \mathbf{Z}$.

Let the regular grid with the number of divisions M be introduced in the interval (0, 1). The set of functions

$$W_{\zeta,M} = \{\zeta_n(x)\}_{n=0}^{M-1}, \tag{5}$$

where the

$$\zeta_n(x) = \zeta(x - n/M), \quad n = 0, \ldots, M-1, \tag{6}$$

are copies of the standard signal shifted to the points of the grid in the unit cell, is said to be the W system of order M generated by $\zeta(x)$.

2.3. Fixed-scale representation

Let the standard signal $\zeta(x)$ and the number of divisions M be chosen. We call the decomposition of the function $f(x)$ into a weighted sum of shifted copies of the standard signal:

$$f(x) = \sum_{n=0}^{M-1} A(n)\zeta(x - n/M) = \sum_{n=0}^{M-1} A(n)\zeta_n(x) \tag{7}$$

[*i.e.* the representation of $f(x)$ as a linear combination of the elements from the $W_{\zeta,M}$ system], the fixed-scale representation of $f(x)$. The real numbers $A(0), \ldots, A(M-1)$ (which may be positive, negative or zero) are called the wavelet coefficients. The 'fixed-scale' term is used to emphasize that every member in the sum (7) differs from the standard signal $\zeta(x)$ only by the multiplier $A(n)$ and the shift in real space by n/M.

The questions that arise immediately and which are discussed below are:

(i) whether a function $f(x)$ can be represented in the form (7) with some previously specified $\zeta(x)$ and M;

(ii) how the coefficients $A(n)$ can be calculated provided the function $f(x)$ is known;

(iii) whether the presentation (7) of $f(x)$ is unique;

(iv) what is the best approximation by the sum (7) for a function $f(x)$ which cannot be expanded into the sum (7) exactly?

Two more questions which are discussed in this section are:

(i) how large is the variety of functions that can be presented as the sum (7) provided the standard signal and the number of divisions M are fixed?;

(ii) whether the set $W_{\zeta,M}$ forms an orthonormal system.

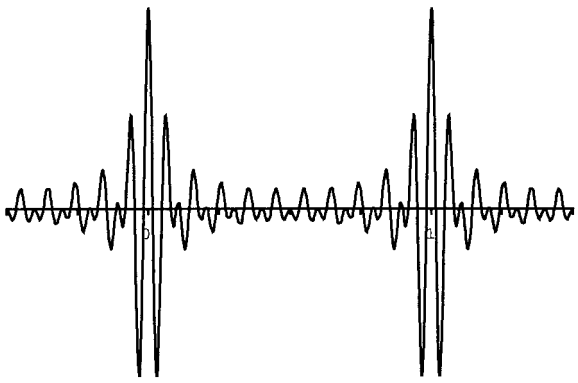

Figure 2
The wavelet (37) for $P = 10$.

2.4. The fixed-scale representation in reciprocal space

If there exists the representation (7) for the function $f(x)$, then, calculating structure factors for both sides of this identity, we get the equivalent representation in reciprocal space:

$$\begin{aligned}\hat{f}(h) &= \sum_{n=0}^{M-1} A(n)\hat{\zeta}_n(h) \\ &= \sum_{n=0}^{M-1} A(n)\hat{\zeta}(h)\exp[2\pi i(hn/M)], \quad h \in \mathbf{Z}.\end{aligned} \tag{8}$$

This identity may be written as

$$\hat{f}(h) = M\hat{A}(h)\hat{\zeta}(h), \quad h \in \mathbf{Z}, \tag{9}$$

where $\hat{A}(h)$ is an M-periodic function calculated as the inverse discrete Fourier transform (IDFT) of the wavelet coefficients

$$\hat{A}(h) = (1/M)\sum_{n=0}^{M-1} A(n)\exp[2\pi i(hn/M)], \quad h \in \mathbf{Z}. \tag{10}$$

It must be emphasized that (9) is equivalent to (7) and all the questions concerning the existence and uniqueness of the decomposition (7) may be equally reformulated as ones concerning the existence and uniqueness of the M-periodic function $\hat{A}(h)$ satisfying (9).

2.5. Necessary and sufficient conditions for the existence of the fixed-scale representation

2.5.1. Nonvanishing $\hat{\zeta}(h)$. Let us suppose first that $\hat{\zeta}(h) \neq 0$ for all h. Then it follows from (8) that

$$\hat{f}(h)/\hat{\zeta}(h) = \sum_{n=0}^{M-1} A(n)\exp[2\pi i(hn/M)] \tag{11}$$

and the necessary and sufficient condition for the existence of the representation (7) is that the ratio $\hat{f}(h)/\hat{\zeta}(h)$ be an M-periodic function of integer argument h, *i.e.*

$$\hat{f}(h)/\hat{\zeta}(h) = \hat{f}(h + kM)/\hat{\zeta}(h + kM), \quad h, k \in \mathbf{Z}. \tag{12}$$

The wavelet coefficients may be restored in this case uniquely from (11) by means of the discrete Fourier transform (DFT)

$$A(n) = \sum_{h=0}^{M-1}[\hat{f}(h)/M\hat{\zeta}(h)]\exp[-2\pi i(hn/M)], \quad n = 0, \ldots, M-1, \tag{13}$$

and may be imagined as the values of the weighted by $1/M\hat{\zeta}(h)$ truncated Fourier synthesis (4) calculated at the proper grid.

2.5.2. Vanishing $\hat{\zeta}(h)$; normalizing function. It was mentioned above that it is usually supposed for the wavelet that $\hat{\zeta}(h)$ is localized in both real and reciprocal spaces. So the condition $\hat{\zeta}(h) \neq 0$ for all h is too restrictive. We consider now a more general case when $\hat{\zeta}(h)$ may be equal to zero for some h. It follows from (9) that for all integers h and k

$$\hat{f}(h + kM) = M\hat{A}(h)\hat{\zeta}(h + kM) \tag{14}$$

and two necessary conditions for the existence of the representations (9) and (7) are

$$\hat{f}(h) = 0 \text{ for } h \text{ such that } \hat{\zeta}(h) = 0 \tag{15}$$

and

$$\hat{f}(h)\hat{\zeta}(h+kM) = \hat{f}(h+kM)\hat{\zeta}(h), \quad h,k \in \mathbf{Z}. \tag{16}$$

These conditions are also sufficient. Let us suppose that the conditions (15)–(16) are satisfied. We derive formulae for the wavelet coefficients satisfying (7).

We define the normalizing function which plays an essential role in the wavelet analysis

$$\hat{Z}_{\zeta,M}(h) = \sum_{k\in\mathbf{Z}} |\hat{\zeta}(h+kM)|^2, \quad h \in \mathbf{Z}, \tag{17}$$

and denote by $Z^0_{\zeta,M}$ the set of its zeros:

$$Z^0_{\zeta,M} = \{h \in \mathbf{Z} : \hat{Z}_{\zeta,M}(h) = 0\}. \tag{18}$$

It is shown below that the existence and uniqueness of the fixed-scale decomposition are conditioned by geometrical properties of the $Z^0_{\zeta,M}$ set rather than the particular values of $\hat{\zeta}(h)$.

If h does not belong to $Z^0_{\zeta,M}$, then the identity (14) determines the value of $\hat{A}(h)$ unambiguously. In fact, in this case, there exists an integer k^* such that $\hat{\zeta}(h+k^*M) \neq 0$ and, to satisfy (14), $\hat{A}(h)$ must be defined as

$$\hat{A}(h) = \hat{f}(h+k^*M)/M\hat{\zeta}(h+k^*M). \tag{19}$$

Owing to condition (16), this definition does not depend on the choice of the particular k^* satisfying $\hat{\zeta}(h+k^*M) \neq 0$.

If h belongs to the set $Z^0_{\zeta,M}$, then $\hat{\zeta}(h+kM) = 0$ for all k and, owing to (15), also $\hat{f}(h+kM) = 0$ for all k. Thus, the condition (14) is satisfied independently of the value of $\hat{A}(h)$. Let us now define $\hat{A}^0(h)$ by (19) for $h \notin Z^0_{\zeta,M}$ and equal to zero for $h \in Z^0_{\zeta,M}$. The identity (9) is now valid for all h and thus the representation (7) exists with wavelet coefficients equal to

$$A^0(n) = \sum_{h=0}^{M-1} \hat{A}^0(h)\exp[-2\pi i(hn/M)], \quad n = 0,\ldots,M-1. \tag{20}$$

The representation (7) is not unique if the set $Z^0_{\zeta,M}$ is not empty. If there exist integers h such that $\hat{Z}_{\zeta,M}(h) = 0$, then corresponding $\hat{A}(h)$ values are not defined uniquely but may be chosen arbitrarily. Similarly, the wavelet coefficients $\{A(n)\}_{n=0}^{M-1}$ in representation (7) may be chosen in this case in different ways. The general expression for admissible wavelet coefficients is

$$A(n) = A^0(n) + \sum_{\substack{h\in Z^0_{\zeta,M}\\ 0\le h<M}} C_h \exp[-2\pi i(hn/M)], \quad n = 0,\ldots,M-1, \tag{21}$$

where C_h values may be chosen arbitrarily. It is possible to reduce this ambiguity if the additional minimality restriction is applied to wavelet coefficients in (7):

$$\sum_{n=0}^{M-1} |A(n)|^2 \Rightarrow \min. \tag{22}$$

The last condition is equivalent to the condition

$$\sum_{h=0}^{M-1} |\hat{A}(h)|^2 \Rightarrow \min, \tag{23}$$

and thus zero values must be assigned to $\hat{A}(h)$ if h belongs to $Z^0_{\zeta,M}$.

2.6. Example 1. The decomposition of a finite resolution Fourier synthesis

Let us consider a Fourier synthesis with finite resolution $d = 1/P$, *i.e.* we consider a real periodic function $f(x)$ such that $\hat{f}(h) = 0$ for $|h| > P$ (P an integer). Let us consider the simplest wavelet, which is defined in reciprocal space by

$$\hat{\zeta}(h) = \begin{cases} 1 & \text{for} \quad |h| \le P \\ 0 & \text{for} \quad |h| > P. \end{cases} \tag{24}$$

or, equivalently, in real space by

$$\zeta(x) = \sin[\pi(2P+1)x]/\sin(\pi x). \tag{25}$$

The shape of the wavelet is shown in Fig. 1.

If the number, M, of the grid nodes is chosen to be equal to $2P+1$, then the conditions (15)–(16) are satisfied and the normalizing function is constant:

$$\hat{Z}_{\zeta,2P+1}(h) = 1 \quad \text{for all } h. \tag{26}$$

It now follows from (19) that

$$\hat{A}(h) = [1/(2P+1)]\hat{f}(h) \quad \text{for } |h| \le P, \tag{27}$$

$$\hat{A}(h+kM) = \hat{A}(h) \quad \text{for all } h,k \in \mathbf{Z}. \tag{28}$$

As a result, any one-dimensional Fourier synthesis with resolution $d = 1/P$ may be represented uniquely as the sum of shifted wavelets (25):

$$f(x) = \sum_{n=0}^{2P} A(n)\frac{\sin\left(\pi(2P+1)\{x-[n/(2P+1)]\}\right)}{\sin\{\pi[x-n/(2P+1)]\}}, \tag{29}$$

where the wavelet coefficients are

$$A(n) = \frac{1}{2P+1}\sum_{h=-P}^{P} \hat{f}(h)\exp\left[-2\pi i h\frac{n}{2P+1}\right]. \tag{30}$$

We see that, up to the multiplier $1/(2P+1)$, these coefficients are just the values of the d-resolution Fourier synthesis calculated at the grid points $n/(2P+1)$, $n = 0,\ldots,2P$.

If the number of wavelets M is chosen to be less than $2P+1$, then the condition (16) is not generally satisfied. For example, it follows in this case from (16) that $\hat{f}(M) = \hat{f}(0)$, which is not usually valid. Thus the exact decomposition (7) does not exist in this case.

If the number of wavelets M is chosen to be greater than $2P+1$, then the conditions (15)–(16) are satisfied, but the normalizing function

$$\hat{Z}_{\zeta,M}(h) = \begin{cases} 1 & \text{for} \quad 0 \le h \le P \quad \text{or} \quad M-P \le h \le M-1 \\ 0 & \text{for} \quad P+1 \le h \le M-P-1 \end{cases} \tag{31}$$

has zeros. As a result, while the representation of $f(x)$ as a sum of wavelets (25) still exists, it is now not unique. Fig. 4(*a*) shows

the number of independent wavelets in the $W_{\zeta,M}$ system for different numbers M of wavelet grid points.

2.7. Redundant *W* systems

We call a W system a redundant one if the standard signal may be expressed as linear combinations of its shifted copies, *i.e.* if there exist identities

$$\sum_{n=0}^{M-1} \alpha_n \zeta(x - n/M) \equiv 0, \tag{32}$$

where not all α_n are equal to zero. It follows from Theorem 1, proved in Appendix A, that the number of different identities (23) is $M - N^0_{\zeta,M}$, where $N^0_{\zeta,M}$ is the number of zeros of the normalizing function $\hat{Z}_{\zeta,M}(h)$ in each period M, *i.e.* the number of integers h such that $0 \le h \le M-1$ and $\hat{Z}_{\zeta,M}(h) = 0$. The identities (23) may be written in more explicit form as

$$\sum_{n=0}^{M-1} \exp[-2\pi i(hn/M)]\zeta(x - n/M) \equiv 0 \quad \text{if } \hat{Z}_{\zeta,M}(h) = 0. \tag{33}$$

It follows from this consideration that there exist sufficiently large systems $W_{\zeta,M}$ such that most of their members are just linear combinations of a small number of independent copies of the standard signal and only the weighted sums of these independent signals may be represented in the form (7). Formally, it is possible to exclude $N^0_{\zeta,M}$ copies from the system and get a linearly independent subsystem. But the centers of the left signals will no longer fill a regular grid. It is sometimes more convenient to leave a redundant but regular system $W_{\zeta,M}$.

It must be noted that, when defined by the additional requirement (22), wavelet coefficients $A^0(n)$ are not independent. They satisfy the additional restrictions

$$\sum_{n=0}^{M-1} \exp[2\pi i(hn/M)]A^0(n) = 0 \quad \text{if } \hat{Z}_{\zeta,M}(h) = 0. \tag{34}$$

2.8. Example 2. Resolution-shell-restricted syntheses

Consider now the space $T_{P,2P}$ of all Fourier syntheses calculated in the resolution shell $1/P \le d \le 1/2P$, *i.e.* consider all functions $f(x)$ such that

$$\hat{f}(h) = 0 \quad \text{for } |h| \le P \text{ and for } |h| > 2P, \tag{35}$$

where P is an integer. Such syntheses constitute corrections which must be added to 'low-resolution' syntheses (with $d_1 = 1/P$ resolution) to 'improve the resolution' up to $d_2 = 1/2P$.

We consider here the simplest standard wavelet defined in reciprocal space as

$$\hat{\zeta}(h) = \begin{cases} 1 & \text{for } P < |h| \le 2P \\ 0 & \text{otherwise,} \end{cases} \tag{36}$$

or, equivalently, in the space by

$$\zeta(x) = \{2\sin(P\pi x)\cos[(3P+1)\pi x]\}/\sin(\pi x). \tag{37}$$

The shape of the wavelet is shown in Fig. 2.

First, let the number of wavelets be chosen to be equal to the number of nonzero $\hat{\zeta}(h)$ values, *i.e.* to $M = 2P$. It is easy to see that the normalizing function in this case is given by

$$\hat{Z}_{\zeta,2P}(h) = \begin{cases} 2 & \text{for } h = 0 \\ 0 & \text{for } h = \pm P \\ 1 & \text{for } 1 \le h < P \text{ and } P < h < 2P, \end{cases} \tag{38}$$

and it follows from (33) that the wavelets are linearly dependent:

$$\sum_{n=0}^{2P-1} (-1)^n \zeta(x - n/2P) \equiv 0. \tag{39}$$

The system $W_{\zeta,2P}$ is now not sufficient to represent all functions from $T_{P,2P}$. It follows from the condition (16) that a necessary condition is

$$\hat{f}(-2P) = \hat{f}(-2P + 4P) = \hat{f}(2P), \tag{40}$$

which together with (2) means that $\hat{f}(2P)$ must be real. Thus, the functions from $T_{P,2P}$ that do not satisfy this condition cannot be represented as a sum of $W_{\zeta,2P}$ wavelets.

Let us now increase the number of wavelets by one and consider the wavelet system $W_{\zeta,2P+1}$. This system is redundant and contains more elements than the dimensionality of the $T_{P,2P}$ space. It is easy to see that in this case

$$Z_{\zeta,2P+1}(h) = \begin{cases} 0 & \text{for } h = 0 \\ 1 & \text{for } h = 1, \ldots, 2P, \end{cases} \tag{41}$$

so $W_{\zeta,2P+1}$ contains $2P$ linearly independent wavelets and thus every function from $T_{P,2P}$ may be represented as a weighted sum of wavelets from $W_{\zeta,2P+1}$. The linear dependence of the full $W_{\zeta,2P+1}$ system takes the form

$$\sum_{n=0}^{2P} \zeta[x - n/(2P+1)] = 0 \quad \text{for all } x. \tag{42}$$

The coefficients of the wavelet decomposition for $f(x)$ may be calculated in this case as the DFT (of length $2P+1$) of the values

$$\hat{A}(h) = \begin{cases} 0 & \text{for } h = 0 \\ \hat{f}(2P+1-h) & \text{for } 1 \le h \le P \\ \hat{f}(2P+1+h) & \text{for } -P \le h \le -1. \end{cases} \tag{43}$$

It is interesting to note that if we increase further the number of wavelets M then, so long as $2P+2 \le M \le 4P$ is satisfied, the number of linearly independent wavelets in $W_{\zeta,M}$ is less than $2P$ again and not all syntheses from $T_{P,2P}$ may be expanded into a sum of $W_{\zeta,M}$ wavelets. Only when M exceeds $4P$ does it then become possible again to represent every function from $T_{P,2P}$ as a sum of $W_{\zeta,M}$ wavelets. Fig. 4(*b*) shows the number of independent wavelets in the $W_{\zeta,M}$ system for different numbers M of the wavelet grid points. It follows from this example that the dependence of the variety of functions that may be represented exactly as sums of the wavelets on the number of the grid points is not straightforward. The transition to a finer grid may result in a reduced number of linearly independent functions which may be represented as a wavelet sum.

2.9. Orthogonal *W* systems

If the condition $\hat{Z}_{\zeta,M}(h) \neq 0$ is satisfied for all integer h, the signals $\{\zeta_n(x)\}_{n=0}^{M-1}$ are linearly independent but they are not, generally, orthogonal. The necessary and sufficient condition for the $W_{\zeta,M}$ to be an orthogonal system is given by Theorem 2, proved in Appendix A, and is simply that the normalizing function $\hat{Z}_{\zeta,M}(h)$ keeps a constant value for all h. It follows from this that the wavelet (25) in §2.6 generates an orthonormal basis in $T_{0,P}$.

If $W_{\zeta,M}$ is a linearly independent system but the orthogonality condition is not satisfied, it is always possible (see §A6 for details) to find another standard signal $\vartheta(x)$ such that it is a combination of elements from $W_{\zeta,M}$:

$$\vartheta(x) = \sum_{n=0}^{M-1} P(n)\zeta(x - n/M), \tag{44}$$

but which now generates an orthogonal system $W_{\vartheta,M} = \{\vartheta(x - n/M)\}_{n=0}^{M-1}$.

Unfortunately, the decay properties of the function $\vartheta(x)$ may be much worse than the ones of the initial signal $\zeta(x)$ (see Fig. 3, for example). So that, although orthogonal systems are more convenient in applications, the desire to have space-localized wavelets forces one to deal with non-orthogonal wavelet systems as well (Daubeshies *et al.*, 1986).

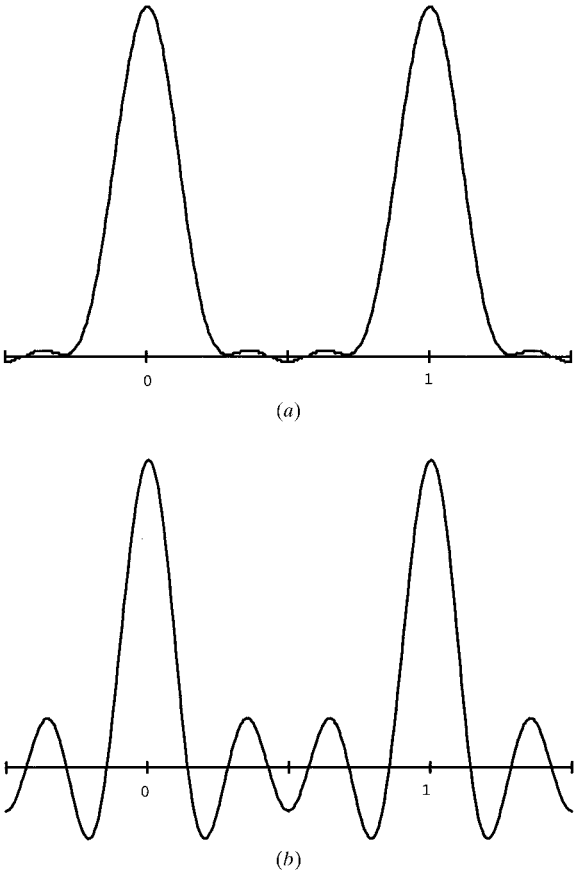

Figure 3
(*a*) The standard wavelet defined by $\hat{\zeta}(h) = \exp[-0.2h^2]$ for $|h| \leq 3$ and $\hat{\zeta}(h) = 0$ for $|h| > 3$. (*b*) The wavelet generating an orthonormal basis in $W_{\zeta,7}$.

3. The best approximation of a function by a sum of fixed-scale signals

The conditions (15)–(16) are not, generally, satisfied and so the exact representation of the function as the sum (7) is not possible. We consider here how the best approximation for $f(x)$ in the form of the weighted sum

$$f_L(x) = \sum_{n=0}^{M-1} A(n)\zeta(x - n/M) = \sum_{n=0}^{M-1} A(n)\zeta_n(x) \tag{45}$$

may be found. By the best approximation, we mean the one that minimizes the residual

$$\begin{aligned} R &= \int_0^1 [f(x) - f_L(x)]^2 \,\mathrm{d}x \\ &= \sum_{h\in\mathbf{Z}} |\hat{f}(h) - \hat{f}_L(h)|^2 \\ &= \sum_{h=0}^{M-1} \sum_{k\in\mathbf{Z}} |\hat{f}(h + kM) - M\hat{A}(h)\hat{\zeta}(h + kM)|^2. \end{aligned} \tag{46}$$

The task of the minimization of this residual may be split into M independent minimization problems

$$\sum_{k\in\mathbf{Z}} |\hat{f}(h + kM) - M\hat{A}(h)\hat{\zeta}(h + kM)|^2 \Rightarrow \min, \qquad h = 0, \ldots, M-1. \tag{47}$$

If $h \notin Z_{\zeta,M}^0$, the last sum is a quadratic function with respect to $\hat{A}(h)$ and its minimum is attained at

$$\hat{A}(h) = \frac{\sum_{k\in\mathbf{Z}} \hat{f}(h + kM)\overline{\hat{\zeta}(h + kM)}}{M\sum_{k\in\mathbf{Z}} |\hat{\zeta}(h + kM)|^2}. \tag{48}$$

For $h \in Z_{\zeta,M}^0$, the sum (47) does not contain $\hat{A}(h)$ and any $\hat{A}(h)$ value may be chosen without changing the sum. The wavelet coefficients for the best approximation satisfying in addition the minimal principle (22) may be found as the DFT (20) of the values $\hat{A}^0(h)$ defined by (48) for $h \notin Z_{\zeta,M}^0$ and as zero for $h \in Z_{\zeta,M}^0$. It is easy to see that the residual (46) may be calculated in this case by

$$\begin{aligned} R &= \sum_{h\in\mathbf{Z}} |\hat{f}(h)|^2 - M^2 \sum_{j=0}^{M-1} Z_{\zeta,M}(j)|\hat{A}(j)|^2 \\ &= \sum_{h\in\mathbf{Z}} |\hat{f}(h)|^2 - \sum_{\substack{0\leq j<M \\ j\notin Z_{\zeta,M}^0}} \left|\sum_{k\in\mathbf{Z}} \hat{f}(j + kM)\overline{\hat{\zeta}(j + kM)}\right|^2 \Big/ Z_{\zeta,M}(j). \end{aligned} \tag{49}$$

The formulae (48) and (20) perform the chain of calculations necessary to calculate wavelet coefficients for the best approximation of a given function $f(x)$. To calculate the approximation $f_L(x)$, the direct summation (45) may be employed. Nevertheless, a more effective procedure would be to calculate first the structure factors $\hat{f}_L(h)$ by means of

$$\hat{f}_L(h) = M\hat{A}(h)\hat{\zeta}(h), \quad h \in \mathbf{Z}, \tag{50}$$

and then calculate $f_L(x)$ using the fast Fourier transform algorithm.

4. Crystallographic examples

We consider here some test results on the application of the wavelet transform for crystallographic purposes. The experimentally observed magnitudes for RNAase *sa* (Sevcik *et al.*, 1991) were used in these tests. This protein crystallizes in the space group $P2_12_12_1$ with unit-cell dimensions 64.9 × 78.32 × 38.79 Å. The phases calculated from the refined atomic model were considered in the tests to be the true phases. When judging the quality of syntheses calculated with phases obtained by the multiple isomorphous replacement (MIR) method, the initially defined (unrefined) MIR phases were used.

4.1. Wavelet approximation of an electron-density distribution

The distribution of electron density in the unit cell of a crystal is customarily approximated by a finite Fourier series. The required information is both the magnitudes and the phases of the structure factors, although the magnitudes only may be obtained experimentally. The goal of the first test was to estimate with what accuracy a finite-resolution Fourier synthesis may be approximated by the sum of a relatively small number of wavelets. Table 1 presents test results for different resolution zones. For every resolution zone, the wavelet grid was chosen in such a way that the number of independent wavelets did not exceed the number of independent structure factors corresponding to this zone. In the considered case, the wavelet coefficients are real numbers (in contrast to the complex structure factors). This means that the number used to represent the synthesis parameters did not exceed the number of the measured structure-factor magnitudes. The map correlation coefficient (Lunin & Woolfson, 1993) was used as a measure of synthesis similarity. For comparison, the map correlation coefficients between exactly phased and MIR-phased maps were calculated too. It follows from Table 1 that it is possible to obtain map quality compatible with the quality of the MIR-phased map even when reducing the number of wavelets used to the number of experimentally measured magnitudes. Several sections of the 3 Å-resolution exactly phased Fourier synthesis as well as its wavelet approximation and the MIR-phased synthesis are shown in Fig. 5.

Table 1
Accuracy of wavelet approximation for Fourier syntheses.

Resolution zone (Å)	Number of reflections†	Number of wavelets†	Correlation with the exactly phased synthesis (%)	
			Wavelet approximation	MIR-phased synthesis
∞–16	39	36	89	23
∞–10	137	120	82	62
∞–5	977	900	83	58
∞–3	4263	4256	83	43

† Numbers of independent reflections and wavelets are shown.

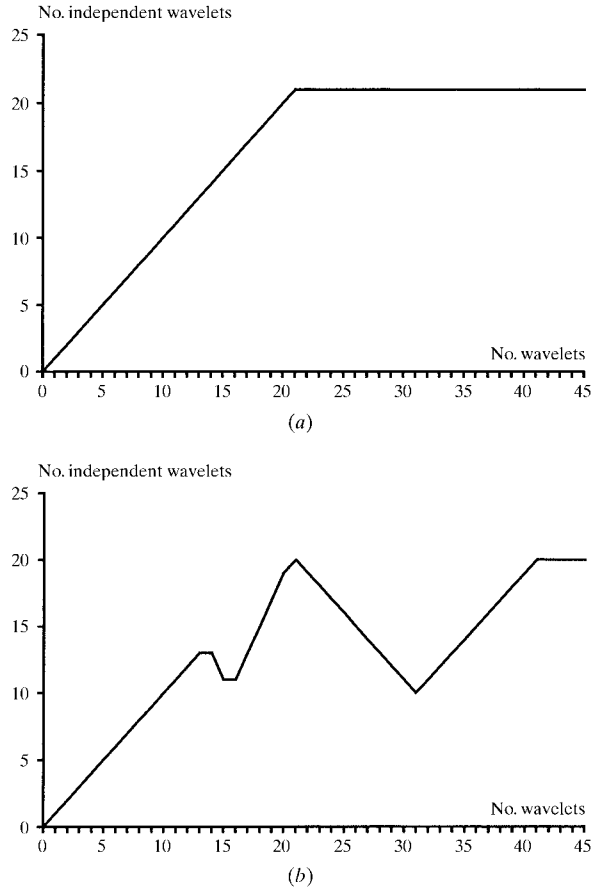

Figure 4
The number of independent wavelets in the $W_{\zeta,M}$ system for different numbers M of wavelet: (*a*) for the standard wavelet (25), $P = 10$; (*b*) for the standard wavelet (27), $P = 10$.

4.2. Refinement of low-resolution phases

It is known that the isomorphous replacement method often provides poor low-resolution phases. Nevertheless, low-resolution reflections play an essential role in imaging the molecular envelope. The simple wavelet-based filtration procedure permits restoration of the low-resolution component. The procedure consists in representing the initial synthesis as the sum of wavelets and calculating an improved image as the wavelet sum corresponding to the strongest wavelets only. Table 2 presents the results of the application of this procedure to the MIR-phased syntheses for RNAase *sa* in different resolution zones.

It must be noted that, as the wavelet coefficients may be considered to some extent as the values of a properly sampled Fourier synthesis, this filtration method is very close to that suggested by Wang (1985) and the double-step filtration methods (Urzhumtsev *et al.*, 1989) of bounding the molecular region in noisy syntheses as well as to the low-density elimination method (Shiono & Woolfson, 1992) and some other density-modification procedures (Podjarny *et al.*, 1996). Table 2 reflects that in our case the filtration had provided a significant improvement of the very low resolution phases but had not affected medium- and high-resolution phases.

APPENDIX *A*
Multidimensional crystallographic fixed-scale wavelet analysis

This section summarizes some results concerning multidimensional analogs of results presented in §§2 and 3 for the one-dimensional case. The term 'crystallographic' is used in the title to emphasize that we assume that the distribution $f(\mathbf{r})$ as well as the standard signal $\zeta(\mathbf{r})$ are real periodic functions. The term 'fixed scale' means that we study the possibility of representing the function as the sum of shifted copies of the same standard signal $\zeta(\mathbf{r})$ without additional dilation. We use the term wavelet to underline our interest in space-localized signals, while formally the decay properties of $\zeta(\mathbf{r})$ are not used in the proofs below.

Table 2
The wavelet-based filtration of MIR-phased syntheses.

Resolution zone (Å)	Number of reflections†	Correlation with the exactly phased synthesis (%)	
		MIR-phased synthesis‡	Improved synthesis§
∞–16	39	23	88
16–10	98	72	75
10–5	840	57	58
5–3	3286	44	45
∞–10	137	62	78
∞–5	977	58	63
∞–3	4263	47	49

† Numbers of independent reflections are shown. ‡ Syntheses calculated with the observed magnitudes and MIR phases were subjected to wavelet decomposition. § The negative wavelet coefficients were suppressed.

A1. Notation

Let $V = \{(x, \ldots, x_m) : 0 \le x_1 \le 1, \ldots, 0 \le x_m \le 1\}$ be a cube in m-dimensional real space $\mathbf{R}^m$. We consider the real linear space $L^2(V)$ composed of real functions $f(\mathbf{r})$, $\mathbf{r} = (x_1, x_2, \ldots, x_m) \in \mathbf{R}^m$ such that they are periodic with the period 1 with respect to any variable and are square integrable in V. We denote by $\hat{f}(\mathbf{h})$ the structure factors (Fourier coefficients) corresponding to $f(\mathbf{r})$, so that

$$\hat{f}(\mathbf{h}) = \int_V f(\mathbf{r}) \exp[2\pi i(\mathbf{h}, \mathbf{r})] \, \mathrm{d}V_{\mathbf{r}} \tag{51}$$

and

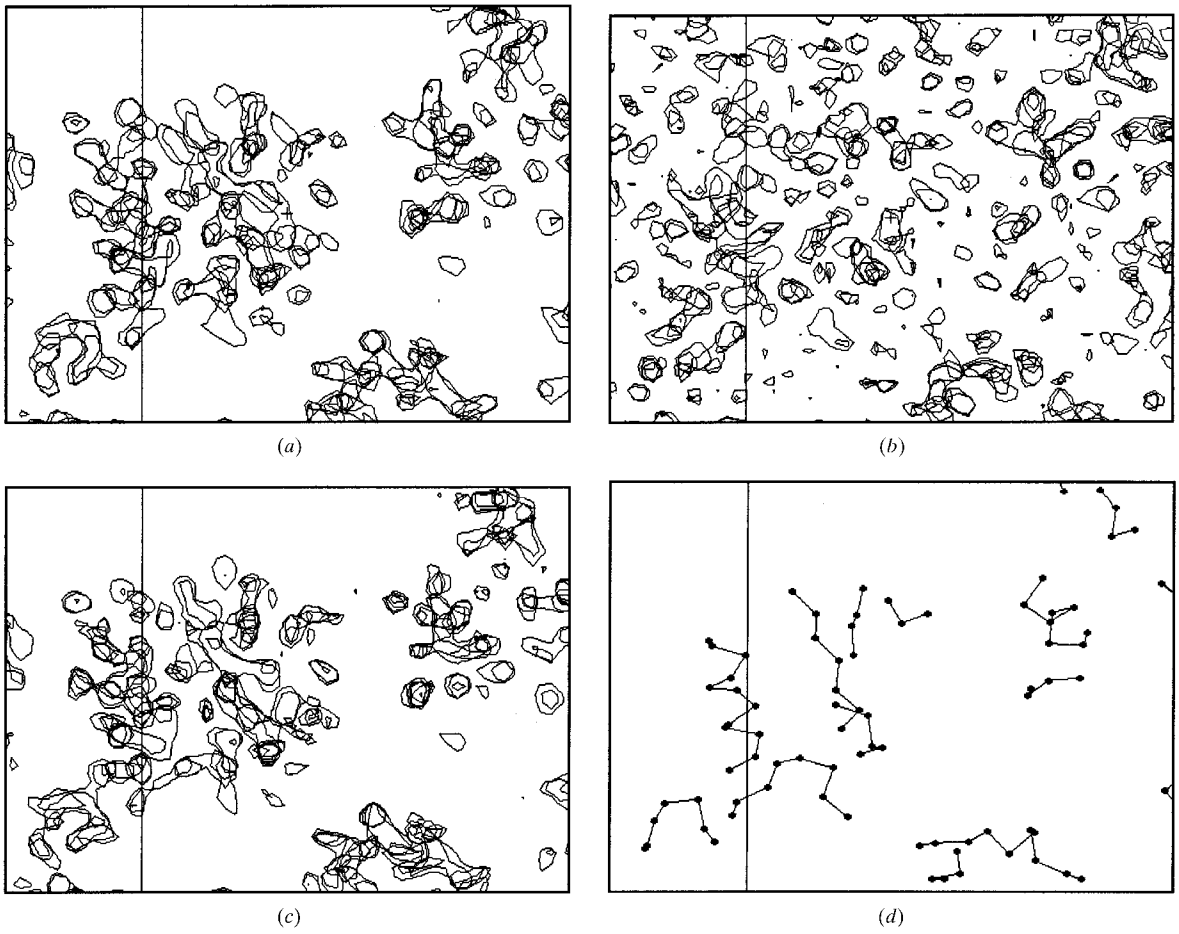

Figure 5
Fragments of Fourier syntheses (3 Å resolution, 4263 independent complex structure factors) for RNAase *sa* calculated with: (*a*) the experimental magnitudes and the phases calculated from the refined atomic model; (*b*) the experimental magnitudes and MIR phases; (*c*) wavelet approximation (4256 real independent wavelets); (*d*) the true atomic positions.

$$f(\mathbf{r}) = \sum_{\mathbf{h}\in\mathbf{Z}^m} \hat{f}(\mathbf{h}) \exp[-2\pi i(\mathbf{h}, \mathbf{r})], \tag{52}$$

where $\mathbf{h} = (h_1, \ldots, h_m) \in \mathbf{Z}^m$ and $\mathbf{Z}^m$ is the m-dimensional space of integers.

Let $\zeta(\mathbf{r})$ be the function that defines what we call the standard signal in real space and $\hat{\zeta}(\mathbf{h})$ be its structure factors. We define the regular grid for signal centers by the numbers of divisions along the axes $M_1, \ldots, M_m$ and denote by $\mathbf{M} = \mathrm{diag}(M_1, \ldots, M_m)$ the diagonal matrix with the principal diagonal formed by $M_1, \ldots, M_m$ and all other elements equal to zero. Obviously,

$$\det(\mathbf{M}) = M_1 M_2 \ldots M_m. \tag{53}$$

Let

$$\mathbf{P}(\mathbf{M}) = \{(h_1, \ldots, h_m) : 0 \le h_1 < M_1, \ldots, 0 \le h_m < M_m\} \tag{54}$$

be a parallelepiped in $\mathbf{Z}^m$. We denote by $\zeta_{\mathbf{n}}(\mathbf{r}) = \zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n})$ the standard signal shifted into the grid point $\mathbf{M}^{-1}\mathbf{n} = (n_1/M_1, \ldots, n_m/M_m)$ and by

$$W_{\zeta,\mathbf{M}} = \{\zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$$

the full set of signals generated by the standard one and all permitted shifts. Let $L_{\zeta,\mathbf{M}}$ be the linear envelope of $W_{\zeta,\mathbf{M}}$ and $f_L(\mathbf{r})$ denotes the orthogonal projection [in the $L^2(V)$ sense] of $f(\mathbf{r})$ onto $L_{\zeta,\mathbf{M}}$. The studied decomposition may be written in this notation as

$$f(\mathbf{r}) = \sum_{\mathbf{n}\in\mathbf{P}(\mathbf{M})} A(\mathbf{n})\, \zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n}). \tag{55}$$

A2. Discrete Fourier transform

For the set of $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$ (complex, in general) values, we denote by $\{\hat{A}(\mathbf{h})\}_{\mathbf{h}\in\mathbf{P}(\mathbf{M})}$ their inverse discrete Fourier transform (IDFT):

$$\hat{A}(\mathbf{h}) = [\det(\mathbf{M})]^{-1} \sum_{\mathbf{n}\in\mathbf{P}(\mathbf{M})} A(\mathbf{n}) \exp[2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})]. \tag{56}$$

The values $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$ may be recovered from (56) by means of the discrete Fourier transform (DFT):

$$A(\mathbf{n}) = \sum_{\mathbf{h}\in\mathbf{P}(\mathbf{M})} \hat{A}(\mathbf{h}) \exp[-2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})]. \tag{57}$$

These formulae are defined first for $\mathbf{h}, \mathbf{n} \in \mathbf{P}(\mathbf{M})$ but they may be extended to $\mathbf{Z}^m$ by supposing $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$ and $\{\hat{A}(\mathbf{h})\}_{\mathbf{h}\in\mathbf{P}(\mathbf{M})}$ to be $M_1, \ldots, M_m$-periodic functions, *i.e.* the functions with periods $M_1, \ldots, M_m$ in the corresponding directions.

A3. The reciprocal-space representation

The possibility of representing a function $f(\mathbf{r})$ as (55) may be formulated by an equivalent way in reciprocal space.

Lemma 1. The function $f(\mathbf{r})$ may be represented in the form (55) with some coefficients $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$ if and only if its structure factors may be represented in the form

$$\hat{f}(\mathbf{h}) = \det(\mathbf{M})\hat{A}(\mathbf{h})\hat{\zeta}(\mathbf{h}) \quad \text{for all } \mathbf{h} \in \mathbf{Z}^m \tag{58}$$

with an $M_1, \ldots, M_m$-periodic function $\hat{A}(\mathbf{h})$. The coefficients $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$ and $\{\hat{A}(\mathbf{h})\}_{\mathbf{h}\in\mathbf{P}(\mathbf{M})}$ are connected by the discrete Fourier transform (56)–(57).

Proof. The calculation of the structure factors for both sides of (55) leads to

$$\begin{aligned}\hat{f}(\mathbf{h}) &= \sum_{\mathbf{n}\in\mathbf{P}(\mathbf{M})} A(\mathbf{n})\hat{\zeta}_{\mathbf{n}}(\mathbf{h}) \\ &= \sum_{\mathbf{n}\in\mathbf{P}(\mathbf{M})} A(\mathbf{n})\hat{\zeta}(\mathbf{h}) \exp[2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})] \\ &= \hat{\zeta}(\mathbf{h}) \sum_{\mathbf{n}\in\mathbf{P}(\mathbf{M})} A(\mathbf{n}) \exp[2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})] \\ &= \det(\mathbf{M})\hat{A}(\mathbf{h})\hat{\zeta}(\mathbf{h}),\end{aligned} \tag{59}$$

which proves Lemma 1.

A4. The uniqueness of the fixed-scale representation

Let the standard signal $\zeta(\mathbf{r})$ and the grid matrix $\mathbf{M}$ be given. We define the normalizing function by

$$\hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h}) = \sum_{\mathbf{k}\in\mathbf{Z}^m} |\hat{\zeta}(\mathbf{h} + \mathbf{M}\mathbf{k})|^2, \tag{60}$$

the set of its zeros by

$$Z^0_{\zeta,\mathbf{M}} = \{\mathbf{h} \in \mathbf{Z}^m : \hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h}) = 0\} \tag{61}$$

and the number $N^0_{\zeta,\mathbf{M}}$ of zeros per unit cell [*i.e.* the number of points in the intersection $\mathbf{P}(\mathbf{M}) \cap Z^0_{\zeta,\mathbf{M}}$].

Let us suppose that the representation (55) exists with some coefficients $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$, *i.e.* there exist the representation (58) with some coefficients $\{\hat{A}(\mathbf{h})\}_{\mathbf{h}\in\mathbf{P}(\mathbf{M})}$. It is easy to see that if $\mathbf{h}' \in Z^0_{\zeta,\mathbf{M}}$ then (58) holds independently of the value of $\hat{A}(\mathbf{h}')$ and so this value may be changed to any other one. This means that if (55) is valid with $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$, then it is valid with any signal coefficients of the form

$$A'(\mathbf{h}) = \begin{cases} A(\mathbf{h}) & \text{for } \mathbf{h} \notin Z^0_{\zeta,\mathbf{M}} \\ A(\mathbf{h}) + C_{\mathbf{h}} & \text{for } \mathbf{h} \in Z^0_{\zeta,\mathbf{M}}, \end{cases} \tag{62}$$

where the $C_{\mathbf{h}}$ are arbitrary complex numbers. The real-space analog is that the coefficients

$$A'(\mathbf{n}) = A(\mathbf{n}) + \sum_{\substack{\mathbf{h}\in Z^0_{\zeta,\mathbf{M}} \\ \mathbf{h}\in\mathbf{P}(\mathbf{M})}} C_{\mathbf{h}} \exp[-2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})] \tag{63}$$

will satisfy (55) as well as $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$ do.

Lemma 2. Let the representation (55) exist. Then among all sets (63) of the signal coefficients such that (55) is valid, the minimal sum

$$\sum_{\mathbf{n}\in\mathbf{P}(\mathbf{M})} |A(\mathbf{n})|^2 \Rightarrow \min \tag{64}$$

is attained for wavelet coefficients satisfying additionally the condition

$$\hat{A}(\mathbf{h}) = 0 \quad \text{for} \quad \mathbf{h} \in Z^0_{\zeta,\mathbf{M}}. \tag{65}$$

Proof. The proof follows immediately from

$$\sum_{\mathbf{n}\in\mathbf{P(M)}} |A(\mathbf{n})|^2 = \det(\mathbf{M}) \sum_{\mathbf{h}\in\mathbf{P(M)}} |\hat{A}(\mathbf{h})|^2. \quad (66)$$

We suppose below that the condition (64) or (65) is applied additionally when the representation (55) is not unique.

A5. The dimensionality of the fixed-scale system

Let $L_{\zeta,\mathbf{M}}$ be the real linear envelope of $W_{\zeta,\mathbf{M}} = \{\zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n})\}_{\mathbf{n}\in\mathbf{P(M)}}$, *i.e.* the variety of functions that may be represented as (55) with specified $\zeta(\mathbf{r})$ and $\mathbf{M}$. The following theorem shows the wide variety of $L_{\zeta,\mathbf{M}}$.

Theorem 1. The dimensionality of the envelope $L_{\zeta,\mathbf{M}}$ may be calculated by means of

$$\dim(L_{\zeta,\mathbf{M}}) = \det(\mathbf{M}) - N^0_{\zeta,\mathbf{M}}. \quad (67)$$

The $N^0_{\zeta,\mathbf{M}}$ linear identities connecting shifted signals may be written as

$$\sum_{\mathbf{n}\in\mathbf{P(M)}} \exp[-2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})]\zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n}) \equiv 0 \quad \text{for } \mathbf{h} \in \mathbf{P(M)} \cap Z^0_{\zeta,\mathbf{M}}. \quad (68)$$

Proof. The set $\mathbf{P(M)}$ contains $\det(\mathbf{M}) - N^0_{\zeta,\mathbf{M}}$ points $\mathbf{j}$ such that $\hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h}) \neq 0$. For every such point, it is possible to construct the nonzero function in reciprocal space by means of

$$\hat{\mu}_{\mathbf{j}}(\mathbf{h}) = \begin{cases} \hat{\zeta}(\mathbf{h}) & \text{for } \mathbf{h} = \mathbf{j} + \mathbf{Mk}, \quad \mathbf{k} \in \mathbf{Z}^m \\ 0 & \text{for other points.} \end{cases} \quad (69)$$

Obviously, the functions corresponding to different $\mathbf{j}$ are linearly independent and every function $\hat{f}(\mathbf{h})$ that may be represented in the form (58) is a linear combination of them. Thus, the $\det(\mathbf{M}) - N^0_{\zeta,\mathbf{M}}$ functions $\mu_{\mathbf{j}}(\mathbf{r})$, obtained by means of (52), form the basis in $L_{\zeta,\mathbf{M}}$.

To prove the identity (68), consider the function

$$Q_{\mathbf{h}}(\mathbf{r}) = \sum_{\mathbf{n}\in\mathbf{P(M)}} \exp[-2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})]\zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n}) \quad (70)$$

and calculate its structure factors for all $\mathbf{j} \in \mathbf{P(M)}$, $\mathbf{k} \in \mathbf{Z}^m$:

$$\begin{aligned} \hat{Q}_{\mathbf{h}}(\mathbf{j} + \mathbf{Mk}) &= \sum_{\mathbf{n}\in\mathbf{P(M)}} \exp[-2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})]\hat{\zeta}_{\mathbf{n}}(\mathbf{j} + \mathbf{Mk}) \\ &= \hat{\zeta}(\mathbf{j} + \mathbf{Mk}) \sum_{\mathbf{n}\in\mathbf{P(M)}} \exp[2\pi i(\mathbf{j} - \mathbf{h}, \mathbf{M}^{-1}\mathbf{n})] \\ &= \det(\mathbf{M})\hat{\zeta}(\mathbf{h} + \mathbf{Mk}). \end{aligned} \quad (71)$$

If $\hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h}) = 0$, then $\hat{\zeta}(\mathbf{h} + \mathbf{Mk}) = 0$ for all $\mathbf{k}$ and thus $Q_{\mathbf{h}}(\mathbf{r}) \equiv 0$.

Remark. It follows from this theorem that the variety of functions that may be represented exactly as weighted sums of shifted copies of the same signal $\zeta(\mathbf{r})$ depends both on the chosen grid of permitted signal shifts and properties of the signal structure factors $\hat{\zeta}(\mathbf{h})$. The dependence on the number of grid points is not straightforward. As follows from Fig. 4(*b*), the transition to a finer grid may result in a reduced number of linearly independent functions, which may be represented as a fixed-scale sum. Concerning the properties of $\hat{\zeta}(\mathbf{h})$, it must be said that here the only essential for the dimensionality property is the distribution of the zeros of $\hat{\zeta}(\mathbf{h})$ in the reciprocal-space lattice, but not the particular values of the nonzero structure factors.

A6. Orthonormal $W_{\zeta,\mathbf{M}}$ systems

Theorem 2. The system $W_{\zeta,\mathbf{M}}$ is an orthonormal system if and only if

$$\hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h}) = [\det(\mathbf{M})]^{-1} \quad \text{for all } \mathbf{h} \in \mathbf{P(M)}. \quad (72)$$

Proof. Let

$$\delta(\mathbf{n}) = \begin{cases} 1 & \text{if } \mathbf{n} = \mathbf{0} \\ 0 & \text{if } \mathbf{n} \neq \mathbf{0}. \end{cases} \quad (73)$$

The system is an orthonormal one if and only if for every $\mathbf{n}, \mathbf{m} \in \mathbf{P(M)}$ the orthonormality condition is satisfied:

$$\begin{aligned} \delta(\mathbf{n} - \mathbf{m}) &= \langle \zeta_{\mathbf{n}}, \zeta_{\mathbf{m}} \rangle \\ &= \sum_{\mathbf{h}\in\mathbf{Z}^m} \hat{\zeta}_{\mathbf{n}}(\mathbf{h})\overline{\hat{\zeta}_{\mathbf{m}}(\mathbf{h})} \\ &= \sum_{\mathbf{h}\in\mathbf{Z}^m} |\hat{\zeta}(\mathbf{h})|^2 \exp[2\pi i(\mathbf{M}^{-1}\mathbf{h}, \mathbf{n} - \mathbf{m})] \\ &= \sum_{\mathbf{h}\in\mathbf{P(M)}} \sum_{\mathbf{k}\in\mathbf{Z}^m} |\hat{\zeta}(\mathbf{h} + \mathbf{Mk})|^2 \exp[2\pi i(\mathbf{M}^{-1}\mathbf{h}, \mathbf{n} - \mathbf{m})]. \end{aligned} \quad (74)$$

This is equivalent to the condition

$$\delta(\mathbf{m}) = \sum_{\mathbf{h}\in\mathbf{P(M)}} \hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h}) \exp[-2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{m})] \quad \text{for all } \mathbf{m} \in \mathbf{P(M)}. \quad (75)$$

Applying the IDFT to both sides of the last identity, we obtain the statement of Theorem 2.

If the system $W_{\zeta,\mathbf{M}}$ is composed of linearly independent but not orthonormal signals, it is possible to construct another standard signal $\theta(\mathbf{r})$ such that the system $W_{\theta,\mathbf{M}}$ forms an orthonormal basis in $L_{\zeta,\mathbf{M}}$. The new orthogonalized standard signal $\theta(\mathbf{r})$ may be defined by its structure factors

$$\hat{\theta}(\mathbf{h}) = \hat{\zeta}(\mathbf{h})/[\det(\mathbf{M})\hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h})]^{1/2} = \det(\mathbf{M})\hat{P}(\mathbf{h})\hat{\zeta}(\mathbf{h}), \quad (76)$$

where

$$\hat{P}(\mathbf{h}) = \{\det(\mathbf{M})[\det(\mathbf{M})\hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h})]^{1/2}\}^{-1}. \quad (77)$$

It is easy to see that the condition (72) is satisfied for $\hat{\theta}(\mathbf{h})$ and

$$\theta(\mathbf{r}) = \sum_{\mathbf{n}\in\mathbf{P(M)}} P(n)\zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n}), \quad (78)$$

where the signal coefficients $P(\mathbf{n})$ are obtained by the DFT from $\hat{P}(\mathbf{h})$. It follows from (78) that

$$L_{\theta,\mathbf{M}} = L_{\zeta,\mathbf{M}}, \quad (79)$$

so $W_{\theta,\mathbf{M}}$ is an orthonormal basis in $L_{\zeta,\mathbf{M}}$.

A7. The best fixed-scale approximation

For any function $f(\mathbf{r}) \in L^2(V)$, we denote by $f_L(\mathbf{r})$ the best approximation of $f(\mathbf{r})$ by signals from $W_{\zeta,\mathbf{M}}$. If $W_{\zeta,\mathbf{M}}$ is an orthonormal basis, then $f_L(\mathbf{r})$ may be calculated by means of

$$f_L(\mathbf{r}) = \sum_{\mathbf{n}\in\mathbf{P}(\mathbf{M})} A(\mathbf{n})\zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n}), \tag{80}$$

where

$$A(\mathbf{n}) = \langle f, \zeta_{\mathbf{n}}\rangle = \int_V f(\mathbf{r})\overline{\zeta(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n})}\,\mathrm{d}V_{\mathbf{r}}. \tag{81}$$

The formula (81) may be extended for redundant systems as follows.

Theorem 3. For any $f(\mathbf{r}) \in L^2(V)$, its projection $f_L(\mathbf{r})$ onto $L_{\zeta,\mathbf{M}}$ may be represented as (75) with the coefficients calculated by means of

$$A(\mathbf{n}) = \langle f, \zeta^*_{\mathbf{n}}\rangle = \int_V f(\mathbf{r})\overline{\zeta^*(\mathbf{r} - \mathbf{M}^{-1}\mathbf{n})}\,\mathrm{d}V_{\mathbf{r}}, \quad \mathbf{n} \in \mathbf{P}(\mathbf{M}), \tag{82}$$

where $\zeta^*(\mathbf{r})$ is the dual signal defined in the reciprocal space as

$$\hat\zeta^*(\mathbf{h}) = \begin{cases} \hat\zeta(\mathbf{h})/[\det(\mathbf{M})\hat Z_{\zeta,\mathbf{M}}(\mathbf{h})] & \text{if } \hat Z_{\zeta,\mathbf{M}}(\mathbf{h}) \neq 0 \\ 0 & \text{if } \hat Z_{\zeta,\mathbf{M}}(\mathbf{h}) = 0. \end{cases} \tag{83}$$

When being defined as (77), the signal coefficients satisfy the additional minimality condition (65).

Proof. It is easy to see that every function that can be represented in the form (58) with $\mathbf{M}$-periodic $\hat A(\mathbf{h})$ may be represented in the form

$$\hat f(\mathbf{h}) = \det(\mathbf{M})\hat A^*(\mathbf{h})\hat\zeta^*(\mathbf{h}) \tag{84}$$

with $\mathbf{M}$-periodic $\hat A^*(\mathbf{h})$ and *vice versa*. Hence, the linear envelopes of $W_{\zeta,\mathbf{M}}$ and $W_{\zeta^*,\mathbf{M}}$ coincide and

$$\begin{aligned} B(n) &= \langle f, \zeta^*_{\mathbf{n}}\rangle \\ &= \langle f_L, \zeta^*_{\mathbf{n}}\rangle \\ &= \sum_{\mathbf{m}\in\mathbf{P}(\mathbf{M})} A(\mathbf{m})\langle \zeta_{\mathbf{m}}, \zeta^*_{\mathbf{n}}\rangle \\ &= \sum_{\mathbf{m}\in\mathbf{P}(\mathbf{M})} A(\mathbf{m}) \sum_{\mathbf{h}\in\mathbf{Z}^m} \hat\zeta_{\mathbf{m}}(\mathbf{h})\overline{\hat\zeta^*_{\mathbf{n}}(\mathbf{h})} \\ &= \sum_{\mathbf{j}\in\mathbf{P}(\mathbf{M})} \left\{ \sum_{\mathbf{m}\in\mathbf{P}(\mathbf{M})} A(\mathbf{m}) \exp[2\pi i(\mathbf{m}, \mathbf{M}^{-1}\mathbf{j})] \right\} \\ &\quad \times \exp[-2\pi i(\mathbf{n}, \mathbf{M}^{-1}\mathbf{j})] \sum_{\mathbf{k}\in\mathbf{Z}^m} \hat\zeta(\mathbf{j} + \mathbf{M}\mathbf{k})\overline{\hat\zeta^*(\mathbf{j} + \mathbf{M}\mathbf{k})} \\ &= \det(\mathbf{M}) \sum_{\mathbf{j}\in\mathbf{P}(\mathbf{M})} \hat A(\mathbf{j}) \left\{ \sum_{\mathbf{k}\in\mathbf{Z}^m} \hat\zeta(\mathbf{j} + \mathbf{M}\mathbf{k})\overline{\hat\zeta^*(\mathbf{j} + \mathbf{M}\mathbf{k})} \right\} \\ &\quad \times \exp[-2\pi i(\mathbf{n}, \mathbf{M}^{-1}\mathbf{j})]. \end{aligned} \tag{85}$$

It follows from this and (60) and (83) that

$$\begin{aligned} \hat B(j) &= \det(\mathbf{M})\hat A(\mathbf{j}) \sum_{\mathbf{k}\in\mathbf{Z}^m} \hat\zeta(\mathbf{j} + \mathbf{M}\mathbf{k})\overline{\hat\zeta^*(\mathbf{j} + \mathbf{M}\mathbf{k})} \\ &= \begin{cases} \hat A(\mathbf{j}) & \text{for } \mathbf{j} \notin Z^0_{\zeta,\mathbf{M}} \\ 0 & \text{for } \mathbf{j} \in Z^0_{\zeta,\mathbf{M}}. \end{cases} \end{aligned} \tag{86}$$

Hence the result (82) is true if the condition (65) is satisfied.

The following theorem permits us to define the signal coefficients through the discrete Fourier transform.

Theorem 4. The coefficients $\{A(\mathbf{n})\}_{\mathbf{n}\in\mathbf{P}(\mathbf{M})}$ in the representation (80) that satisfy (65) may be calculated as the DFT of the values

$$\hat A(\mathbf{h}) = \begin{cases} [\det(\mathbf{M})\hat Z_{\zeta,\mathbf{M}}(\mathbf{h})]^{-1} \sum_{\mathbf{k}\in\mathbf{Z}^m} \hat f(\mathbf{h} + \mathbf{M}\mathbf{k})\overline{\hat\zeta(\mathbf{h} + \mathbf{M}\mathbf{k})} \\ \qquad\qquad \text{if } \hat Z_{\zeta,\mathbf{M}}(\mathbf{h}) \neq 0 \\ 0 \qquad\qquad \text{if } \hat Z_{\zeta,\mathbf{M}}(\mathbf{h}) = 0. \end{cases} \tag{87}$$

Proof. It follows from (82) that

$$\begin{aligned} A(\mathbf{n}) &= \langle f, \zeta^*_{\mathbf{n}}\rangle \\ &= \sum_{\mathbf{h}\in\mathbf{Z}^m} \hat f(\mathbf{h})\overline{\hat\zeta^*(\mathbf{h})} \exp[-2\pi i(\mathbf{h}, \mathbf{M}^{-1}\mathbf{n})] \\ &= \sum_{\mathbf{j}\in\mathbf{P}(\mathbf{M})} \left\{ \sum_{\mathbf{k}\in\mathbf{Z}^m} \hat f(\mathbf{j} + \mathbf{M}\mathbf{k})\overline{\hat\zeta^*(\mathbf{j} + \mathbf{M}\mathbf{k})} \right\} \\ &\quad \times \exp[-2\pi i(\mathbf{j}, \mathbf{M}^{-1}\mathbf{n})] \end{aligned} \tag{88}$$

and

$$\hat A(\mathbf{j}) = \sum_{\mathbf{k}\in\mathbf{Z}^m} \hat f(\mathbf{j} + \mathbf{M}\mathbf{k})\overline{\hat\zeta^*(\mathbf{j} + M\mathbf{k})}. \tag{89}$$

The formula (86) follows from the definition (83).

A8. The symmetry of the signal coefficients

Let $\Gamma = \{(\mathbf{R}_\nu, \mathbf{t}_\nu)\}_{\nu=1}^n$ be a crystallographic space group and $\Pi = \{(\mathbf{R}_\nu, \mathbf{0})\}_{\nu=1}^m$ be the corresponding point group of symmetries. We say that a function $f(\mathbf{r})$ possesses the Γ symmetry if for all $0 \leq \nu < n - 1$

$$f(\mathbf{R}_\nu\mathbf{r} + \mathbf{t}_\nu) = f(\mathbf{r}) \quad \text{for all } \mathbf{r} \in \mathbf{R}^m \tag{90}$$

or, equivalently,

$$\hat f(\mathbf{R}^T_\nu\mathbf{h}) = \hat f(\mathbf{h}) \exp[-2\pi i(\mathbf{h}, \mathbf{t}_\nu)] \quad \text{for all} \quad \mathbf{h} \in \mathbf{Z}^m. \tag{91}$$

Let us suppose that the grid $\mathbf{M}$ is consistent with the Γ symmetry, *i.e.* every symmetry operation transforms the grid points into grid points again. It follows, in particular, from the consistency that for every ν the inner product $(\mathbf{M}\mathbf{k}, \mathbf{t}_\nu)$ is an integer and the set of the points $\mathbf{R}^T_\nu\mathbf{M}\mathbf{k}$ coincides with $\mathbf{M}\mathbf{k}$ when $\mathbf{k}$ runs through $\mathbf{Z}^m$. The next theorem links the symmetry of the standard signal with the symmetry of the signal coefficients.

Theorem 5. Let $f(\mathbf{r})$ be a function that possesses Γ symmetry and the grid $\mathbf{M}$ be consistent with this symmetry.

If the standard signal $\zeta(\mathbf{r})$ possesses Γ symmetry, then the signal coefficients $A(\mathbf{n})$ possess the point-group symmetry, *i.e.*

$$A(\mathbf{R}_\nu\mathbf{n}) = A(\mathbf{n})|_{\mathrm{mod}\,\mathbf{M}} \quad \text{for all } 0 \leq \nu < n - 1 \text{ and } \mathbf{n} \in P(\mathbf{M}). \tag{92}$$

If the standard signal $\zeta(\mathbf{r})$ possesses the point-group symmetry, then the signal coefficients possess Γ symmetry, *i.e.*

$$A(\mathbf{R}_\nu\mathbf{n}+\mathbf{M}\mathbf{t}_\nu)=A(\mathbf{n})|_{\mathrm{mod}\,\mathbf{M}} \quad \text{for all } 0\le\nu<n-1 \text{ and } \mathbf{n}\in P(\mathbf{M}). \quad (93)$$

Proof. It is enough to show that the values $\hat{A}(\mathbf{h})$ possess the corresponding symmetry. It follows from (60) that if $\zeta(\mathbf{r})$ has Γ or Π symmetry, then

$$\begin{aligned}\hat{Z}_{\zeta,\mathbf{M}}(\mathbf{R}_\nu^T\mathbf{h}) &= \sum_{\mathbf{k}\in\mathbf{Z}^m}|\hat{\zeta}(\mathbf{R}_\nu^T\mathbf{h}+\mathbf{M}\mathbf{k})|^2\\ &= \sum_{\mathbf{k}\in\mathbf{Z}^m}|\hat{\zeta}(\mathbf{R}_\nu^T\mathbf{h}+\mathbf{R}_\nu^T\mathbf{M}\mathbf{k})|^2\\ &= \sum_{\mathbf{k}\in\mathbf{Z}^m}|\hat{\zeta}(\mathbf{h}+\mathbf{M}\mathbf{k})|^2\\ &= \hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h}), \end{aligned} \quad (94)$$

i.e. the normalizing function possesses the point-group symmetry.

It follows from (87) now that

$$\hat{A}(\mathbf{R}_\nu^T\mathbf{h}) = [\det(\mathbf{M})\hat{Z}_{\zeta,\mathbf{M}}(\mathbf{h})]^{-1} \times \sum_{\mathbf{k}\in\mathbf{Z}^m}\hat{f}(\mathbf{R}_\nu^T\mathbf{h}+\mathbf{R}_\nu^T\mathbf{M}\mathbf{k})\overline{\hat{\zeta}(\mathbf{R}_\nu^T\mathbf{h}+\mathbf{R}_\nu^T\mathbf{M}\mathbf{k})}. \quad (95)$$

If both $f(\mathbf{r})$ and $\zeta(\mathbf{r})$ possess Γ symmetry, then, owing to (91),

$$\hat{f}(\mathbf{R}_\nu^T\mathbf{h}+\mathbf{R}_\nu^T\mathbf{M}\mathbf{k})\overline{\hat{\zeta}(\mathbf{R}_\nu^T\mathbf{h}+\mathbf{R}_\nu^T\mathbf{M}\mathbf{k})} = \hat{f}(\mathbf{h}+\mathbf{M}\mathbf{k})\overline{\hat{\zeta}(\mathbf{h}+\mathbf{M}\mathbf{k})}, \quad (96)$$

and as the result we get

$$\hat{A}(\mathbf{R}_\nu^T\mathbf{h}) = \hat{A}(\mathbf{h}), \quad (97)$$

which means that the signal coefficients $A(\mathbf{n})$ have the point-group symmetry.

If $\zeta(\mathbf{r})$ possess the point-group symmetry, then

$$\begin{aligned}&\hat{f}(\mathbf{R}_\nu^T\mathbf{h}+\mathbf{R}_\nu^T\mathbf{M}\mathbf{k})\overline{\hat{\zeta}(\mathbf{R}_\nu^T\mathbf{h}+\mathbf{R}_\nu^T\mathbf{M}\mathbf{k})}\\ &\quad= \hat{f}(\mathbf{h}+\mathbf{M}\mathbf{k})\overline{\hat{\zeta}(\mathbf{h}+\mathbf{M}\mathbf{k})}\exp[-2\pi i(\mathbf{h}+\mathbf{M}\mathbf{k},\mathbf{t}_\nu)]\\ &\quad= \hat{f}(\mathbf{h}+\mathbf{M}\mathbf{k})\overline{\hat{\zeta}(\mathbf{h}+\mathbf{M}\mathbf{k})}\exp[-2\pi i(\mathbf{h},\mathbf{t}_\nu)] \end{aligned} \quad (98)$$

and so

$$\hat{A}(\mathbf{R}_\nu^T\mathbf{h}) = \hat{A}(\mathbf{h})\exp[-2\pi i(\mathbf{h},\mathbf{t}_\nu)], \quad (99)$$

which means that the signal coefficients $A(\mathbf{n})$ have Γ symmetry.

I wish to express my gratitude to the referees for their valuable advice and to C. Lecomte for his interest in my work. The work was supported by grant No. RFBR 97-04-48319.

References

Agarwal, R. C. & Isaacs, N. W. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2835–2839.

Chu, C. K. (1992). *An Introduction to Wavelets.* New York: Academic Press.

Combes, J. M., Grossmann, A. & Tchamitchian, P. (1988). Editors. *Wavelets.* Berlin: Springer-Verlag.

Daubechies, I. (1993). Editor. *Different Perspectives on Wavelets. Proceedings of Symposia in Applied Mathematics*, Vol. 47. New York: American Mathematical Society.

Daubeshies, I., Grossman, A. & Meyer, Y. (1986). *J. Math. Phys.* **27**, 1271–1283.

Ferrer, J.-L., Roth, M. & Antoniadis, A. (1998). *Acta Cryst.* D**54**, 184–199.

Lamzin, V. S. & Wilson, K.S. (1993). *Acta Cryst.* D**49**, 129–147.

Lunin, V. Y., Lunina, N. L., Petrova, T. E., Urzhumtsev, A. G. & Podjarny, A. D. (1998). *Acta Cryst.* D**54**, 726–734.

Lunin, V. Y., Lunina, N. L., Petrova, T. E., Vernoslova, E. A., Urzhumtsev, A. G. & Podjarny, A. D. (1995). *Acta Cryst.* D**51**, 896–903.

Lunin, V. Yu. & Urzhumtsev, A. G. (1984). *Acta Cryst.* A**40**, 269–277.

Lunin, V. Yu., Urzhumtsev, A. G., Vernoslova, E. A., Chirgadze, Yu. N., Nevskaya, N. A. & Fomenkova, N. P. (1985). *Acta Cryst.* A**41**, 166–171.

Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.

Main, P. (1996). Microsymposium 02.06, XVII IUCr Congress, Seattle, WA, USA, 8–17 August 1996.

Podjarny, A. D., Rees, B. & Urzhumtsev, A. G. (1996). *Methods Mol. Biol.* **56**, 205–226.

Pouligny, B., Gabriel, G., Muzy, J. F., Arneodo, A. & Argoul, F. (1991). *J. Appl. Cryst.* **24**, 526–530.

Sevcik, J., Dodson, E. J. & Dodson, G. G. (1991). *Acta Cryst.* B**47**, 240–253.

Shiono, M. & Woolfson, M. M. (1992). *Acta Cryst.* A**48**, 451–456.

Somoza, J. R., Szöke, H. & Szöke, A. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier. *NATO ASI Series*, Vol. 507, pp. 499–502.

Urzhumtsev, A. G. (1997). *Acta Cryst.* D**53**, 540–543.

Urzhumtsev, A. G., Lunin, V. Yu. & Luzyanina, T. B. (1989). *Acta Cryst.* A**45**, 34–39.

Vellieux, F. M. D. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier. *NATO ASI Series*, Vol. 507, pp. 503–512.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.