International Tables for Crystallography (2006). Vol. B, Chapter 1.3, pp. 25–98.

1.3. Fourier transforms in crystallography: theory, algorithms and applications

BY G. BRICOGNE

1.3.1. General introduction

Since the publication of Volume II of *International Tables*, most aspects of the theory, computation and applications of Fourier transforms have undergone considerable development, often to the point of being hardly recognizable.

The mathematical analysis of the Fourier transformation has been extensively reformulated within the framework of distribution theory, following Schwartz's work in the early 1950s.

The computation of Fourier transforms has been revolutionized by the advent of digital computers and of the Cooley–Tukey algorithm, and progress has been made at an ever-accelerating pace in the design of new types of algorithms and in optimizing their interplay with machine architecture.

These advances have transformed both theory and practice in several fields which rely heavily on Fourier methods; much of electrical engineering, for instance, has become digital signal processing.

By contrast, crystallography has remained relatively unaffected by these developments. From the conceptual point of view, old-fashioned Fourier series are still adequate for the quantitative description of X-ray diffraction, as this rarely entails consideration of molecular transforms between reciprocal-lattice points. From the practical point of view, three-dimensional Fourier transforms have mostly been used as a tool for visualizing electron-density maps, so that only moderate urgency was given to trying to achieve ultimate efficiency in these relatively infrequent calculations.

Recent advances in phasing and refinement methods, however, have placed renewed emphasis on concepts and techniques long used in digital signal processing, *e.g.* flexible sampling, Shannon interpolation, linear filtering, and interchange between convolution and multiplication. These methods are iterative in nature, and thus generate a strong incentive to design new crystallographic Fourier transform algorithms making the fullest possible use of all available symmetry to save both storage and computation.

As a result, need has arisen for a modern and coherent account of Fourier transform methods in crystallography which would provide:

(i) a simple and foolproof means of switching between the three different guises in which the Fourier transformation is encountered (Fourier transforms, Fourier series and discrete Fourier transforms), both formally and computationally;

(ii) an up-to-date presentation of the most important algorithms for the efficient numerical calculation of discrete Fourier transforms;

(iii) a systematic study of the incorporation of symmetry into the calculation of crystallographic discrete Fourier transforms;

(iv) a survey of the main types of crystallographic computations based on the Fourier transformation.

The rapid pace of progress in these fields implies that such an account would be struck by quasi-immediate obsolescence if it were written solely for the purpose of compiling a catalogue of results and formulae 'customized' for crystallographic use. Instead, the emphasis has been placed on a mode of presentation in which most results and formulae are *derived* rather than listed. This does entail a substantial mathematical overhead, but has the advantage of preserving in its 'native' form the context within which these results are obtained. It is this context, rather than any particular set of results, which constitutes the most fertile source of new ideas and new applications, and as such can have any hope at all of remaining useful in the long run.

These conditions have led to the following choices:

(i) the mathematical theory of the Fourier transformation has been cast in the language of Schwartz's theory of distributions which has long been adopted in several applied fields, in particular electrical engineering, with considerable success; the extra work involved handsomely pays for itself by allowing the three different types of Fourier transformations to be treated together, and by making all properties of the Fourier transform consequences of a single property (the convolution theorem). This is particularly useful in all questions related to the sampling theorem;

(ii) the various numerical algorithms have been presented as the consequences of basic algebraic phenomena involving Abelian groups, rings and finite fields; this degree of formalization greatly helps the subsequent incorporation of symmetry;

(iii) the algebraic nature of space groups has been re-emphasized so as to build up a framework which can accommodate both the phenomena used to factor the discrete Fourier transform and those which underlie the existence (and lead to the classification) of space groups; this common ground is found in the notion of *module over a group ring* (*i.e.* integral representation theory), which is then applied to the formulation of a large number of algorithms, many of which are new;

(iv) the survey of the main types of crystallographic computations has tried to highlight the roles played by various properties of the Fourier transformation, and the ways in which a better exploitation of these properties has been the driving force behind the discovery of more powerful methods.

In keeping with this philosophy, the theory is presented first, followed by the crystallographic applications. There are 'forward references' from mathematical results to the applications which later invoke them (thus giving 'real-life' examples rather than artificial ones), and 'backward references' as usual. In this way, the internal logic of the mathematical developments – the surest guide to future innovations – can be preserved, whereas the alternative solution of relegating these to appendices tends on the contrary to obscure that logic by subordinating it to that of the applications.

It is hoped that this attempt at an overall presentation of the main features of Fourier transforms and of their ubiquitous role in crystallography will be found useful by scientists both within and outside the field.

1.3.2. The mathematical theory of the Fourier transformation

1.3.2.1. *Introduction*

The Fourier transformation and the practical applications to which it gives rise occur in three different forms which, although they display a similar range of phenomena, normally require distinct formulations and different proof techniques:

(i) *Fourier transforms*, in which both function and transform depend on continuous variables;

(ii) *Fourier series*, which relate a periodic function to a discrete set of coefficients indexed by n-tuples of integers;

(iii) *discrete Fourier transforms*, which relate finite-dimensional vectors by linear operations representable by matrices.

At the same time, the most useful property of the Fourier transformation – the exchange between multiplication and convolution – is mathematically the most elusive and the one which requires the greatest caution in order to avoid writing down meaningless expressions.

It is the unique merit of Schwartz's theory of distributions (Schwartz, 1966) that it affords complete control over all the troublesome phenomena which had previously forced mathematicians to settle for a piecemeal, fragmented theory of the Fourier transformation. By its ability to handle rigorously highly 'singular'

objects (especially δ-functions, their derivatives, their tensor products, their products with smooth functions, their translates and lattices of these translates), distribution theory can deal with all the major properties of the Fourier transformation as particular instances of a single basic result (the exchange between multiplication and convolution), *and* can at the same time accommodate the three previously distinct types of Fourier theories within a unique framework. This brings great simplification to matters of central importance in crystallography, such as the relations between

(*a*) periodization, and sampling or decimation;
(*b*) Shannon interpolation, and masking by an indicator function;
(*c*) section, and projection;
(*d*) differentiation, and multiplication by a monomial;
(*e*) translation, and phase shift.

All these properties become subsumed under the same theorem.

This striking synthesis comes at a slight price, which is the relative complexity of the notion of distribution. It is first necessary to establish the notion of topological vector space and to gain sufficient control (or, at least, understanding) over convergence behaviour in certain of these spaces. The key notion of *metrizability* cannot be circumvented, as it underlies most of the constructs and many of the proof techniques used in distribution theory. Most of Section 1.3.2.2 builds up to the fundamental result at the end of Section 1.3.2.2.6.2, which is basic to the definition of a distribution in Section 1.3.2.3.4 and to all subsequent developments.

The reader mostly interested in applications will probably want to reach this section by starting with his or her favourite topic in Section 1.3.4, and following the backward references to the relevant properties of the Fourier transformation, then to the proof of these properties, and finally to the definitions of the objects involved. Hopefully, he or she will then feel inclined to follow the forward references and thus explore the subject from the abstract to the practical. The books by Dieudonné (1969) and Lang (1965) are particularly recommended as general references for all aspects of analysis and algebra.

1.3.2.2. *Preliminary notions and notation*

Throughout this text, $\mathbb{R}$ will denote the set of real numbers, $\mathbb{Z}$ the set of rational (signed) integers and $\mathbb{N}$ the set of natural (unsigned) integers. The symbol $\mathbb{R}^n$ will denote the Cartesian product of n copies of $\mathbb{R}$:

$$\mathbb{R}^n = \mathbb{R} \times \ldots \times \mathbb{R} \quad (n \text{ times}, n \geq 1),$$

so that an element $\mathbf{x}$ of $\mathbb{R}^n$ is an n-tuple of real numbers:

$$\mathbf{x} = (x_1, \ldots, x_n).$$

Similar meanings will be attached to $\mathbb{Z}^n$ and $\mathbb{N}^n$.

The symbol $\mathbb{C}$ will denote the set of complex numbers. If $z \in \mathbb{C}$, its modulus will be denoted by $|z|$, its conjugate by $\bar{z}$ (not z^*), and its real and imaginary parts by $\mathscr{Re}\,(z)$ and $\mathscr{Im}\,(z)$:

$$\mathscr{Re}\,(z) = \tfrac{1}{2}(z + \bar{z}), \qquad \mathscr{Im}\,(z) = \frac{1}{2i}(z - \bar{z}).$$

If X is a finite set, then $|X|$ will denote the number of its elements. If mapping f sends an element x of set X to the element $f(x)$ of set Y, the notation

$$f : x \longmapsto f(x)$$

will be used; the plain arrow $\rightarrow$ will be reserved for denoting limits, as in

$$\lim_{\rho \to \infty} \left(1 + \frac{x}{p}\right)^p = e^x.$$

If X is any set and S is a subset of X, the *indicator function* χ_S of S is the real-valued function on X defined by

$$\chi_S(x) = 1 \quad \text{if } x \in S$$
$$= 0 \quad \text{if } x \notin S.$$

1.3.2.2.1. *Metric and topological notions in* $\mathbb{R}^n$

The set $\mathbb{R}^n$ can be endowed with the structure of a vector space of dimension n over $\mathbb{R}$, and can be made into a Euclidean space by treating its standard basis as an orthonormal basis and defining the Euclidean norm:

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}.$$

By misuse of notation, $\mathbf{x}$ will sometimes also designate the column vector of coordinates of $\mathbf{x} \in \mathbb{R}^n$; if these coordinates are referred to an orthonormal basis of $\mathbb{R}^n$, then

$$\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2},$$

where $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$.

The distance between two points $\mathbf{x}$ and $\mathbf{y}$ defined by $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ allows the topological structure of $\mathbb{R}$ to be transferred to $\mathbb{R}^n$, making it a *metric space*. The basic notions in a metric space are those of neighbourhoods, of open and closed sets, of limit, of continuity, and of convergence (see Section 1.3.2.2.6.1).

A subset S of $\mathbb{R}^n$ is *bounded* if $\sup \|\mathbf{x} - \mathbf{y}\| < \infty$ as $\mathbf{x}$ and $\mathbf{y}$ run through S; it is *closed* if it contains the limits of all convergent sequences with elements in S. A subset K of $\mathbb{R}^n$ which is both bounded and closed has the property of being *compact*, *i.e.* that whenever K has been covered by a family of open sets, a finite subfamily can be found which suffices to cover K. Compactness is a very useful topological property for the purpose of proof, since it allows one to reduce the task of examining infinitely many local situations to that of examining only finitely many of them.

1.3.2.2.2. *Functions over* $\mathbb{R}^n$

Let φ be a complex-valued function over $\mathbb{R}^n$. The *support* of φ, denoted Supp φ, is the smallest closed subset of $\mathbb{R}^n$ outside which φ vanishes identically. If Supp φ is compact, φ is said to have compact support.

If $\mathbf{t} \in \mathbb{R}^n$, the *translate* of φ by $\mathbf{t}$, denoted $\tau_{\mathbf{t}}\varphi$, is defined by

$$(\tau_{\mathbf{t}}\varphi)(\mathbf{x}) = \varphi(\mathbf{x} - \mathbf{t}).$$

Its support is the geometric translate of that of φ:

$$\text{Supp } \tau_{\mathbf{t}}\varphi = \{\mathbf{x} + \mathbf{t} | \mathbf{x} \in \text{Supp } \varphi\}.$$

If A is a non-singular linear transformation in $\mathbb{R}^n$, the *image* of φ by A, denoted $A^{\#}\varphi$, is defined by

$$(A^{\#}\varphi)(\mathbf{x}) = \varphi[A^{-1}(\mathbf{x})].$$

Its support is the geometric image of Supp φ under A:

$$\text{Supp } A^{\#}\varphi = \{A(\mathbf{x}) | \mathbf{x} \in \text{Supp } \varphi\}.$$

If S is a non-singular affine transformation in $\mathbb{R}^n$ of the form

$$S(\mathbf{x}) = A(\mathbf{x}) + \mathbf{b}$$

with A linear, the image of φ by S is $S^{\#}\varphi = \tau_{\mathbf{b}}(A^{\#}\varphi)$, *i.e.*

$$(S^{\#}\varphi)(\mathbf{x}) = \varphi[A^{-1}(\mathbf{x} - \mathbf{b})].$$

Its support is the geometric image of Supp φ under S:

$$\text{Supp } S^{\#}\varphi = \{S(\mathbf{x}) | \mathbf{x} \in \text{Supp } \varphi\}.$$

It may be helpful to visualize the process of forming the image of a function by a geometric operation as consisting of applying that operation to the *graph* of that function, which is equivalent to

applying the *inverse* transformation to the coordinates $\mathbf{x}$. This use of the inverse later affords the 'left-representation property' [see Section 1.3.4.2.2.2(*e*)] when the geometric operations form a group, which is of fundamental importance in the treatment of crystallographic symmetry (Sections 1.3.4.2.2.4, 1.3.4.2.2.5).

1.3.2.2.3. *Multi-index notation*

When dealing with functions in n variables and their derivatives, considerable abbreviation of notation can be obtained through the use of multi-indices.

A *multi-index* $\mathbf{p} \in \mathbb{N}^n$ is an n-tuple of natural integers: $\mathbf{p} = (p_1, \ldots, p_n)$. The *length* of $\mathbf{p}$ is defined as

$$|\mathbf{p}| = \sum_{i=1}^{n} p_i,$$

and the following abbreviations will be used:

(i) $\mathbf{x}^{\mathbf{p}} = x_1^{p_1} \ldots x_n^{p_n}$

(ii) $D_i f = \dfrac{\partial f}{\partial x_i} = \partial_i f$

(iii) $D^{\mathbf{p}} f = D_1^{p_1} \ldots D_n^{p_n} f = \dfrac{\partial^{|\mathbf{p}|} f}{\partial x_1^{p_1} \ldots \partial x_n^{p_n}}$

(iv) $\mathbf{q} \leq \mathbf{p}$ if and only if $q_i \leq p_i$ for all $i = 1, \ldots, n$

(v) $\mathbf{p} - \mathbf{q} = (p_1 - q_1, \ldots, p_n - q_n)$

(vi) $\mathbf{p}! = p_1! \times \ldots \times p_n!$

(vii) $\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \times \ldots \times \begin{pmatrix} p_n \\ q_n \end{pmatrix}.$

Leibniz's formula for the repeated differentiation of products then assumes the concise form

$$D^{\mathbf{p}}(fg) = \sum_{\mathbf{q} \leq \mathbf{p}} \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} D^{\mathbf{p}-\mathbf{q}} f D^{\mathbf{q}} g,$$

while the Taylor expansion of f to order m about $\mathbf{x} = \mathbf{a}$ reads

$$f(\mathbf{x}) = \sum_{|\mathbf{p}| \leq m} \frac{1}{\mathbf{p}!} [D^{\mathbf{p}} f(\mathbf{a})](\mathbf{x} - \mathbf{a})^{\mathbf{p}} + o(\|\mathbf{x} - \mathbf{a}\|^m).$$

In certain sections the notation ∇f will be used for the gradient vector of f, and the notation $(\nabla\nabla^T)f$ for the Hessian matrix of its mixed second-order partial derivatives:

$$\nabla = \begin{pmatrix} \dfrac{\partial}{\partial x_1} \\ \vdots \\ \dfrac{\partial}{\partial x_n} \end{pmatrix}, \quad \nabla f = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix},$$

$$(\nabla\nabla^T)f = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

1.3.2.2.4. *Integration, L^p spaces*

The Riemann integral used in elementary calculus suffers from the drawback that vector spaces of Riemann-integrable functions over $\mathbb{R}^n$ are not *complete* for the topology of convergence in the mean: a Cauchy sequence of integrable functions may converge to a non-integrable function.

To obtain the property of completeness, which is fundamental in functional analysis, it was necessary to extend the notion of integral. This was accomplished by Lebesgue [see Berberian (1962), Dieudonné (1970), or Chapter 1 of Dym & McKean (1972) and the references therein, or Chapter 9 of Sprecher (1970)], and entailed identifying functions which differed only on a subset of zero measure in $\mathbb{R}^n$ (such functions are said to be equal 'almost everywhere'). The vector spaces $L^p(\mathbb{R}^n)$ consisting of function classes f modulo this identification for which

$$\|\mathbf{f}\|_p = \left(\int_{\mathbb{R}^n} |f(\mathbf{x})|^p \, \mathrm{d}^n\mathbf{x} \right)^{1/p} < \infty$$

are then complete for the topology induced by the norm $\|.\|_p$: the limit of every Cauchy sequence of functions in L^p is itself a function in L^p (Riesz–Fischer theorem).

The space $L^1(\mathbb{R}^n)$ consists of those function classes f such that

$$\|f\|_1 = \int_{\mathbb{R}^n} |f(\mathbf{x})| \, \mathrm{d}^n\mathbf{x} < \infty$$

which are called *summable* or *absolutely integrable*. The convolution product:

$$\begin{aligned}(f * g)(\mathbf{x}) &= \int_{\mathbb{R}^n} f(\mathbf{y}) g(\mathbf{x} - \mathbf{y}) \, \mathrm{d}^n\mathbf{y} \\ &= \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y}) g(\mathbf{y}) \, \mathrm{d}^n\mathbf{y} = (g * f)(\mathbf{x})\end{aligned}$$

is well defined; combined with the vector space structure of L^1, it makes L^1 into a (commutative) *convolution algebra*. However, this algebra has no unit element: there is no $f \in L^1$ such that $f * g = g$ for all $g \in L^1$; it has only approximate units, *i.e.* sequences (f_ν) such that $f_\nu * g$ tends to g in the L^1 topology as $\nu \to \infty$. This is one of the starting points of distribution theory.

The space $L^2(\mathbb{R}^n)$ of *square-integrable* functions can be endowed with a scalar product

$$(f, g) = \int_{\mathbb{R}^n} \overline{f(\mathbf{x})} g(\mathbf{x}) \, \mathrm{d}^n\mathbf{x}$$

which makes it into a *Hilbert space*. The Cauchy–Schwarz inequality

$$|(f, g)| \leq [(f, f)(g, g)]^{1/2}$$

generalizes the fact that the absolute value of the cosine of an angle is less than or equal to 1.

The space $L^\infty(\mathbb{R}^n)$ is defined as the space of functions f such that

$$\|f\|_\infty = \lim_{p \to \infty} \|f\|_p = \lim_{p \to \infty} \left(\int_{\mathbb{R}^n} |f(\mathbf{x})|^p \, \mathrm{d}^n\mathbf{x} \right)^{1/p} < \infty.$$

The quantity $\|f\|_\infty$ is called the 'essential sup norm' of f, as it is the smallest positive number which $|f(\mathbf{x})|$ exceeds only on a subset of zero measure in $\mathbb{R}^n$. A function $f \in L^\infty$ is called *essentially bounded*.

1.3.2.2.5. *Tensor products. Fubini's theorem*

Let $f \in L^1(\mathbb{R}^m)$, $g \in L^1(\mathbb{R}^n)$. Then the function

$$f \otimes g : (\mathbf{x}, \mathbf{y}) \longmapsto f(\mathbf{x}) g(\mathbf{y})$$

is called the *tensor product* of f and g, and belongs to $L^1(\mathbb{R}^m \times \mathbb{R}^n)$. The finite linear combinations of functions of the form $f \otimes g$ span a subspace of $L^1(\mathbb{R}^m \times \mathbb{R}^n)$ called the tensor product of $L^1(\mathbb{R}^m)$ and $L^1(\mathbb{R}^n)$ and denoted $L^1(\mathbb{R}^m) \otimes L^1(\mathbb{R}^n)$.

The integration of a general function over $\mathbb{R}^m \times \mathbb{R}^n$ may be accomplished in two steps according to *Fubini's theorem*. Given $F \in L^1(\mathbb{R}^m \times \mathbb{R}^n)$, the functions

$$F_1 : \mathbf{x} \longmapsto \int_{\mathbb{R}^n} F(\mathbf{x}, \mathbf{y})\, \mathrm{d}^n\mathbf{y}$$

$$F_2 : \mathbf{y} \longmapsto \int_{\mathbb{R}^m} F(\mathbf{x}, \mathbf{y})\, \mathrm{d}^m\mathbf{x}$$

exist for almost all $\mathbf{x} \in \mathbb{R}^m$ and almost all $\mathbf{y} \in \mathbb{R}^n$, respectively, are integrable, and

$$\int_{\mathbb{R}^m \times \mathbb{R}^n} F(\mathbf{x}, \mathbf{y})\, \mathrm{d}^m\mathbf{x}\, \mathrm{d}^n\mathbf{y} = \int_{\mathbb{R}^m} F_1(\mathbf{x})\, \mathrm{d}^m\mathbf{x} = \int_{\mathbb{R}^n} F_2(\mathbf{y})\, \mathrm{d}^n\mathbf{y}.$$

Conversely, if any one of the integrals

(i) $$\int_{\mathbb{R}^m \times \mathbb{R}^n} |F(\mathbf{x}, \mathbf{y})|\, \mathrm{d}^m\mathbf{x}\, \mathrm{d}^n\mathbf{y}$$

(ii) $$\int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} |F(\mathbf{x}, \mathbf{y})|\, \mathrm{d}^n\mathbf{y} \right) \mathrm{d}^m\mathbf{x}$$

(iii) $$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} |F(\mathbf{x}, \mathbf{y})|\, \mathrm{d}^m\mathbf{x} \right) \mathrm{d}^n\mathbf{y}$$

is finite, then so are the other two, and the identity above holds. It is then (and only then) permissible to change the order of integrations.

Fubini's theorem is of fundamental importance in the study of tensor products and convolutions of distributions.

1.3.2.2.6. *Topology in function spaces*

Geometric intuition, which often makes 'obvious' the topological properties of the real line and of ordinary space, cannot be relied upon in the study of function spaces: the latter are infinite-dimensional, and several inequivalent notions of convergence may exist. A careful analysis of topological concepts and of their interrelationship is thus a necessary prerequisite to the study of these spaces. The reader may consult Dieudonné (1969, 1970), Friedman (1970), Trèves (1967) and Yosida (1965) for detailed expositions.

1.3.2.2.6.1. *General topology*

Most topological notions are first encountered in the setting of *metric spaces*. A metric space E is a set equipped with a *distance function* d from $E \times E$ to the non-negative reals which satisfies:

(i)	$d(x,y) = d(y,x)$	$\forall x, y \in E$	(symmetry);
(ii)	$d(x,y) = 0$	iff $x = y$	(separation);
(iii)	$d(x,z) \leq d(x,y) + d(y,z)$	$\forall x, y, z \in E$	(triangular inequality).

By means of d, the following notions can be defined: open balls, neighbourhoods; open and closed sets, interior and closure; convergence of sequences, continuity of mappings; Cauchy sequences and completeness; compactness; connectedness. They suffice for the investigation of a great number of questions in analysis and geometry (see *e.g.* Dieudonné, 1969).

Many of these notions turn out to depend only on the properties of the collection $\mathcal{O}(E)$ of open subsets of E: two distance functions leading to the same $\mathcal{O}(E)$ lead to identical topological properties. An axiomatic reformulation of topological notions is thus possible: a *topology* in E is a collection $\mathcal{O}(E)$ of subsets of E which satisfy suitable axioms and are deemed open irrespective of the way they are obtained. From the practical standpoint, however, a topology which can be obtained from a distance function (called a *metrizable* topology) has the very useful property that *the notions of closure, limit and continuity may be defined by means of sequences*. For non-metrizable topologies, these notions are much more difficult to handle, requiring the use of 'filters' instead of sequences.

In some spaces E, a topology may be most naturally defined by a family of *pseudo-distances* $(d_\alpha)_{\alpha \in A}$, where each d_α satisfies (i) and (iii) but not (ii). Such spaces are called *uniformizable*. If for every pair $(x, y) \in E \times E$ there exists $\alpha \in A$ such that $d_\alpha(x, y) \neq 0$, then the separation property can be recovered. If furthermore a *countable* subfamily of the d_α suffices to define the topology of E, the latter can be shown to be *metrizable*, so that limiting processes in E may be studied by means of sequences.

1.3.2.2.6.2. *Topological vector spaces*

The function spaces E of interest in Fourier analysis have an underlying vector space structure over the field $\mathbb{C}$ of complex numbers. A topology on E is said to be *compatible* with a vector space structure on E if vector addition [*i.e.* the map $(\mathbf{x}, \mathbf{y}) \longmapsto \mathbf{x} + \mathbf{y}$] and scalar multiplication [*i.e.* the map $(\lambda, \mathbf{x}) \longmapsto \lambda\mathbf{x}$] are both *continuous*; E is then called a *topological vector space*. Such a topology may be defined by specifying a 'fundamental system S of neighbourhoods of $\mathbf{0}$', which can then be translated by vector addition to construct neighbourhoods of other points $\mathbf{x} \neq \mathbf{0}$.

A *norm* ν on a vector space E is a non-negative real-valued function on $E \times E$ such that

(i′)	$\nu(\lambda\mathbf{x}) = \lvert\lambda\rvert\nu(\mathbf{x})$	for all $\lambda \in \mathbb{C}$ and $\mathbf{x} \in E$;
(ii′)	$\nu(\mathbf{x}) = 0$	if and only if $\mathbf{x} = \mathbf{0}$;
(iii′)	$\nu(\mathbf{x} + \mathbf{y}) \leq \nu(\mathbf{x}) + \nu(\mathbf{y})$	for all $\mathbf{x}, \mathbf{y} \in E$.

Subsets of E defined by conditions of the form $\nu(\mathbf{x}) \leq r$ with $r > 0$ form a fundamental system of neighbourhoods of $\mathbf{0}$. The corresponding topology makes E a *normed space*. This topology is *metrizable*, since it is equivalent to that derived from the translation-invariant distance $d(\mathbf{x}, \mathbf{y}) = \nu(\mathbf{x} - \mathbf{y})$. Normed spaces which are *complete*, *i.e.* in which all Cauchy sequences converge, are called *Banach spaces*; they constitute the natural setting for the study of differential calculus.

A *semi-norm* σ on a vector space E is a positive real-valued function on $E \times E$ which satisfies (i′) and (iii′) but not (ii′). Given a set Σ of semi-norms on E such that any pair $(\mathbf{x}, \mathbf{y})$ in $E \times E$ is separated by at least one $\sigma \in \Sigma$, let B be the set of those subsets $\Gamma_{\sigma, r}$ of E defined by a condition of the form $\sigma(\mathbf{x}) \leq r$ with $\sigma \in \Sigma$ and $r > 0$; and let S be the set of finite intersections of elements of B. Then there exists a unique topology on E for which S is a fundamental system of neighbourhoods of $\mathbf{0}$. This topology is *uniformizable* since it is equivalent to that derived from the family of translation-invariant pseudo-distances $(\mathbf{x}, \mathbf{y}) \longmapsto \sigma(\mathbf{x} - \mathbf{y})$. It is *metrizable* if and only if it can be constructed by the above procedure with Σ a *countable* set of semi-norms. If furthermore E is complete, E is called a *Fréchet space*.

If E is a topological vector space over $\mathbb{C}$, its *dual* E^* is the set of all linear mappings from E to $\mathbb{C}$ (which are also called *linear forms*, or *linear functionals*, over E). The subspace of E^* consisting of all linear forms which are *continuous* for the topology of E is called the *topological dual* of E and is denoted E'. If the topology on E is metrizable, then the continuity of a linear form $T \in E'$ at $f \in E$ can be ascertained by means of sequences, *i.e.* by checking that the sequence $[T(f_j)]$ of complex numbers converges to $T(f)$ in $\mathbb{C}$ whenever the sequence (f_j) converges to f in E.

1.3.2.3. *Elements of the theory of distributions*

1.3.2.3.1. *Origins*

At the end of the 19th century, Heaviside proposed under the name of 'operational calculus' a set of rules for solving a class of

differential, partial differential and integral equations encountered in electrical engineering (today's 'signal processing'). These rules worked remarkably well but were devoid of mathematical justification (see Whittaker, 1928). In 1926, Dirac introduced his famous δ-function [see Dirac (1958), pp. 58–61], which was found to be related to Heaviside's constructs. Other singular objects, together with procedures to handle them, had already appeared in several branches of analysis [Cauchy's 'principal values'; Hadamard's 'finite parts' (Hadamard, 1932, 1952); Riesz's regularization methods for certain divergent integrals (Riesz, 1938, 1949)] as well as in the theories of Fourier series and integrals (see *e.g.* Bochner, 1932, 1959). Their very definition often verged on violating the rigorous rules governing limiting processes in analysis, so that subsequent recourse to limiting processes could lead to erroneous results; *ad hoc* precautions thus had to be observed to avoid mistakes in handling these objects.

In 1945–1950, Laurent Schwartz proposed his theory of distributions (see Schwartz, 1966), which provided a unified and definitive treatment of all these questions, with a striking combination of rigour and simplicity. Schwartz's treatment of Dirac's δ-function illustrates his approach in a most direct fashion. Dirac's original definition reads:

(i) $\delta(\mathbf{x}) = 0$ for $\mathbf{x} \neq \mathbf{0}$,

(ii) $\int_{\mathbb{R}^n} \delta(\mathbf{x})\ \mathrm{d}^n\mathbf{x} = 1$.

These two conditions are irreconcilable with Lebesgue's theory of integration: by (i), δ vanishes almost everywhere, so that its integral in (ii) must be 0, not 1.

A better definition consists in specifying that

(iii) $\int_{\mathbb{R}^n} \delta(\mathbf{x})\varphi(\mathbf{x})\ \mathrm{d}^n\mathbf{x} = \varphi(\mathbf{0})$

for any function φ sufficiently well behaved near $\mathbf{x} = \mathbf{0}$. This is related to the problem of finding a unit for convolution (Section 1.3.2.2.4). As will now be seen, this definition is still unsatisfactory. Let the sequence (f_ν) in $L^1(\mathbb{R}^n)$ be an approximate convolution unit, *e.g.*

$$f_\nu(\mathbf{x}) = \left(\frac{\nu}{2\pi}\right)^{1/2} \exp(-\tfrac{1}{2}\nu^2\|\mathbf{x}\|^2).$$

Then for any well behaved function φ the integrals

$$\int_{\mathbb{R}^n} f_\nu(\mathbf{x})\varphi(\mathbf{x})\ \mathrm{d}^n\mathbf{x}$$

exist, and the sequence of their numerical values tends to $\varphi(\mathbf{0})$. It is tempting to combine this with (iii) to conclude that δ is the limit of the sequence (f_ν) as $\nu \to \infty$. However,

$$\lim f_\nu(\mathbf{x}) = 0 \quad \text{as } \nu \to \infty$$

almost everywhere in $\mathbb{R}^n$ and the crux of the problem is that

$$\varphi(\mathbf{0}) = \lim_{\nu\to\infty} \int_{\mathbb{R}^n} f_\nu(\mathbf{x})\varphi(\mathbf{x})\ \mathrm{d}^n\mathbf{x}$$

$$\neq \int_{\mathbb{R}^n} \left[\lim_{\nu\to\infty} f_\nu(\mathbf{x})\right]\varphi(\mathbf{x})\ \mathrm{d}^n\mathbf{x} = 0$$

because the sequence (f_ν) does not satisfy the hypotheses of Lebesgue's dominated convergence theorem.

Schwartz's solution to this problem is deceptively simple: the regular behaviour one is trying to capture is an attribute not of the sequence of functions (f_ν), but of the sequence of continuous linear functionals

$$T_\nu : \varphi \longmapsto \int_{\mathbb{R}^n} f_\nu(\mathbf{x})\varphi(\mathbf{x})\ \mathrm{d}^n\mathbf{x}$$

which has as a limit the continuous functional

$$T : \varphi \longmapsto \varphi(\mathbf{0}).$$

It is the latter functional which constitutes the proper definition of δ. The previous paradoxes arose because one insisted on writing down the simple linear operation T in terms of an integral.

The essence of Schwartz's theory of distributions is thus that, rather than try to define and handle 'generalized functions' *via* sequences such as (f_ν) [an approach adopted *e.g.* by Lighthill (1958) and Erdélyi (1962)], one should instead look at them as continuous linear functionals over spaces of well behaved functions.

There are many books on distribution theory and its applications. The reader may consult in particular Schwartz (1965, 1966), Gel'fand & Shilov (1964), Bremermann (1965), Trèves (1967), Challifour (1972), Friedlander (1982), and the relevant chapters of Hörmander (1963) and Yosida (1965). Schwartz (1965) is especially recommended as an introduction.

1.3.2.3.2. *Rationale*

The guiding principle which leads to requiring that the functions φ above (traditionally called 'test functions') should be well behaved is that correspondingly 'wilder' behaviour can then be accommodated in the limiting behaviour of the f_ν while still keeping the integrals $\int_{\mathbb{R}^n} f_\nu\varphi\ \mathrm{d}^n\mathbf{x}$ under control. Thus

(i) to minimize restrictions on the limiting behaviour of the f_ν *at infinity*, the φ's will be chosen to have *compact support*;

(ii) to minimize restrictions on the *local* behaviour of the f_ν, the φ's will be chosen *infinitely differentiable*.

To ensure further the *continuity* of functionals such as T_ν with respect to the test function φ as the f_ν go increasingly wild, very strong control will have to be exercised in the way in which a sequence (φ_j) of test functions will be said to converge towards a limiting φ: conditions will have to be imposed not only on the values of the *functions* φ_j, but also on those of all their *derivatives*. Hence, defining a strong enough topology on the space of test functions φ is an *essential* prerequisite to the development of a satisfactory theory of distributions.

1.3.2.3.3. *Test-function spaces*

With this rationale in mind, the following function spaces will be defined for any open subset Ω of $\mathbb{R}^n$ (which may be the whole of $\mathbb{R}^n$):

(*a*) $\mathscr{E}(\Omega)$ is the space of complex-valued functions over Ω which are indefinitely differentiable;

(*b*) $\mathscr{D}(\Omega)$ is the subspace of $\mathscr{E}(\Omega)$ consisting of functions with (unspecified) compact support contained in $\mathbb{R}^n$;

(*c*) $\mathscr{D}_K(\Omega)$ is the subspace of $\mathscr{D}(\Omega)$ consisting of functions whose (compact) support is contained within a fixed compact subset K of Ω.

When Ω is unambiguously defined by the context, we will simply write $\mathscr{E}, \mathscr{D}, \mathscr{D}_K$.

It sometimes suffices to require the existence of continuous derivatives only up to finite order m inclusive. The corresponding spaces are then denoted $\mathscr{E}^{(m)}, \mathscr{D}^{(m)}, \mathscr{D}_K^{(m)}$ with the convention that if $m = 0$, only continuity is required.

The *topologies* on these spaces constitute the most important ingredients of distribution theory, and will be outlined in some detail.

1.3.2.3.3.1. *Topology on* $\mathscr{E}(\Omega)$

It is defined by the family of semi-norms

$$\varphi \in \mathscr{E}(\Omega) \longmapsto \sigma_{\mathbf{p},K}(\varphi) = \sup_{\mathbf{x}\in K} |D^{\mathbf{p}}\varphi(\mathbf{x})|,$$

where $\mathbf{p}$ is a multi-index and K a compact subset of Ω. A

fundamental system S of neighbourhoods of the origin in $\mathscr{E}(\Omega)$ is given by subsets of $\mathscr{E}(\Omega)$ of the form

$$V(m,\varepsilon,K)=\{\varphi\in\mathscr{E}(\Omega)||\mathbf{p}|\leq m\Rightarrow\sigma_{\mathbf{p},K}(\varphi)\ <\ \varepsilon\}$$

for all natural integers m, positive real ε, and compact subset K of Ω. Since a *countable* family of compact subsets K suffices to cover Ω, and since restricted values of ε of the form $\varepsilon=1/N$ lead to the same topology, S is equivalent to a *countable* system of neighbourhoods and hence $\mathscr{E}(\Omega)$ *is metrizable*.

Convergence in $\mathscr{E}$ may thus be defined by means of sequences. A sequence (φ_ν) in $\mathscr{E}$ will be said to converge to 0 if for any given $V(m,\varepsilon,K)$ there exists ν_0 such that $\varphi_\nu\in V(m,\varepsilon,K)$ whenever $\nu>\nu_0$; in other words, if the φ_ν and all their derivatives $D^{\mathbf{p}}\varphi_\nu$ converge to 0 uniformly on any given compact K in Ω.

1.3.2.3.3.2. *Topology on* $\mathscr{D}_k(\Omega)$

It is defined by the family of semi-norms

$$\varphi\in\mathscr{D}_K(\Omega)\longmapsto\sigma_{\mathbf{p}}(\varphi)=\sup_{\mathbf{x}\in K}|D^{\mathbf{p}}\varphi(\mathbf{x})|,$$

where K is now fixed. The fundamental system S of neighbourhoods of the origin in $\mathscr{D}_K$ is given by sets of the form

$$V(m,\varepsilon)=\{\varphi\in\mathscr{D}_K(\Omega)||\mathbf{p}|\leq m\Rightarrow\sigma_{\mathbf{p}}(\varphi)\ <\ \varepsilon\}.$$

It is equivalent to the countable subsystem of the $V(m,1/N)$, hence $\mathscr{D}_K(\Omega)$ *is metrizable*.

Convergence in $\mathscr{D}_K$ may thus be defined by means of sequences. A sequence (φ_ν) in $\mathscr{D}_K$ will be said to converge to 0 if for any given $V(m,\varepsilon)$ there exists ν_0 such that $\varphi_\nu\in V(m,\varepsilon)$ whenever $\nu>\nu_0$; in other words, if the φ_ν and all their derivatives $D^{\mathbf{p}}\varphi_\nu$ converge to 0 uniformly in K.

1.3.2.3.3.3. *Topology on* $\mathscr{D}(\Omega)$

It is defined by the fundamental system of neighbourhoods of the origin consisting of sets of the form

$$V((m),(\varepsilon))=\left\{\varphi\in\mathscr{D}(\Omega)||\mathbf{p}|\leq m_\nu\Rightarrow\sup_{\|\mathbf{x}\|\leq\nu}|D^{\mathbf{p}}\varphi(\mathbf{x})|\ <\ \varepsilon_\nu\text{ for all }\nu\right\},$$

where (m) is an increasing sequence (m_ν) of integers tending to $+\infty$ and (ε) is a decreasing sequence (ε_ν) of positive reals tending to 0, as $\nu\to\infty$.

This topology is *not metrizable*, because the sets of sequences (m) and (ε) are essentially uncountable. It can, however, be shown to be the *inductive limit* of the topology of the subspaces $\mathscr{D}_K$, in the following sense: V is a neighbourhood of the origin in $\mathscr{D}$ if and only if its intersection with $\mathscr{D}_K$ is a neighbourhood of the origin in $\mathscr{D}_K$ for any given compact K in Ω.

A sequence (φ_ν) in $\mathscr{D}$ will thus be said to converge to 0 in $\mathscr{D}$ if all the φ_ν belong to some $\mathscr{D}_K$ (with K a compact subset of Ω independent of ν) and if (φ_ν) converges to 0 in $\mathscr{D}_K$.

As a result, a complex-valued functional T on $\mathscr{D}$ will be said to be continuous for the topology of $\mathscr{D}$ if and only if, for any given compact K in Ω, its restriction to $\mathscr{D}_K$ is continuous for the topology of $\mathscr{D}_K$, *i.e.* maps convergent sequences in $\mathscr{D}_K$ to convergent sequences in $\mathbb{C}$.

This property of $\mathscr{D}$, *i.e.* having a non-metrizable topology which is the inductive limit of metrizable topologies in its subspaces $\mathscr{D}_K$, conditions the whole structure of distribution theory and dictates that of many of its proofs.

1.3.2.3.3.4. *Topologies on* $\mathscr{E}^{(m)}$, $\mathscr{D}_k^{(m)}$, $\mathscr{D}^{(m)}$

These are defined similarly, but only involve conditions on derivatives up to order m.

1.3.2.3.4. *Definition of distributions*

A *distribution* T on Ω is a *linear form* over $\mathscr{D}(\Omega)$, *i.e.* a map

$$T:\varphi\longmapsto\langle T,\varphi\rangle$$

which associates linearly a complex number $\langle T,\varphi\rangle$ to any $\varphi\in\mathscr{D}(\Omega)$, and which is *continuous* for the topology of that space. In the terminology of Section 1.3.2.2.6.2, T is an element of $\mathscr{D}'(\Omega)$, the *topological dual* of $\mathscr{D}(\Omega)$.

Continuity over $\mathscr{D}$ is equivalent to continuity over $\mathscr{D}_K$ for all compact K contained in Ω, and hence to the condition that for any sequence (φ_ν) in $\mathscr{D}$ such that

(i) Supp φ_ν is contained in some compact K independent of ν,

(ii) the sequences $(|D^{\mathbf{p}}\varphi_\nu|)$ converge uniformly to 0 on K for all multi-indices $\mathbf{p}$;

then the sequence of complex numbers $\langle T,\varphi_\nu\rangle$ converges to 0 in $\mathbb{C}$.

If the continuity of a distribution T requires (ii) for $|\mathbf{p}|\leq m$ only, T may be defined over $\mathscr{D}^{(m)}$ and thus $T\in\mathscr{D}'^{(m)}$; T is said to be a *distribution of finite order* m. In particular, for $m=0$, $\mathscr{D}^{(0)}$ is the space of continuous functions with compact support, and a distribution $T\in\mathscr{D}'^{(0)}$ is a (Radon) *measure* as used in the theory of integration. Thus measures are particular cases of distributions.

Generally speaking, the *larger* a space of test functions, the *smaller* its topological dual:

$$m\ <\ n\Rightarrow\mathscr{D}^{(m)}\supset\mathscr{D}^{(n)}\Rightarrow\mathscr{D}'^{(n)}\supset\mathscr{D}'^{(m)}.$$

This clearly results from the observation that if the φ's are allowed to be less regular, then less wildness can be accommodated in T if the continuity of the map $\varphi\longmapsto\langle T,\varphi\rangle$ with respect to φ is to be preserved.

1.3.2.3.5. *First examples of distributions*

(i) The linear map $\varphi\longmapsto\langle\delta,\varphi\rangle=\varphi(\mathbf{0})$ is a measure (*i.e.* a zeroth-order distribution) called Dirac's measure or (improperly) Dirac's 'δ-function'.

(ii) The linear map $\varphi\longmapsto\langle\delta_{(\mathbf{a})},\varphi\rangle=\varphi(\mathbf{a})$ is called Dirac's measure at point $\mathbf{a}\in\mathbb{R}^n$.

(iii) The linear map $\varphi\longmapsto(-1)^{\mathbf{p}}D^{\mathbf{p}}\varphi(\mathbf{a})$ is a distribution of order $m=|\mathbf{p}|>0$, and hence is not a measure.

(iv) The linear map $\varphi\longmapsto\sum_{\nu>0}\varphi^{(\nu)}(\nu)$ is a distribution of infinite order on $\mathbb{R}$: the order of differentiation is bounded for each φ (because φ has compact support) but is not as φ varies.

(v) If $(\mathbf{p}_\nu)$ is a sequence of multi-indices $\mathbf{p}_\nu=(p_{1\nu},\ldots,p_{n\nu})$ such that $|\mathbf{p}_\nu|\to\infty$ as $\nu\to\infty$, then the linear map $\varphi\longmapsto\sum_{\nu>0}(D^{\mathbf{p}_\nu}\varphi)(\mathbf{p}_\nu)$ is a distribution of infinite order on $\mathbb{R}^n$.

1.3.2.3.6. *Distributions associated to locally integrable functions*

Let f be a complex-valued function over Ω such that $\int_K|f(\mathbf{x})|\,\mathrm{d}^n\mathbf{x}$ exists for any given compact K in Ω; f is then called *locally integrable*.

The linear mapping from $\mathscr{D}(\Omega)$ to $\mathbb{C}$ defined by

$$\varphi\longmapsto\int_\Omega f(\mathbf{x})\varphi(\mathbf{x})\,\mathrm{d}^n\mathbf{x}$$

may then be shown to be continuous over $\mathscr{D}(\Omega)$. It thus defines a *distribution* $T_f\in\mathscr{D}'(\Omega)$:

$$\langle T_f,\varphi\rangle=\int_\Omega f(\mathbf{x})\varphi(\mathbf{x})\,\mathrm{d}^n\mathbf{x}.$$

As the continuity of T_f only requires that $\varphi\in\mathscr{D}^{(0)}(\Omega)$, T_f is actually a Radon measure.

It can be shown that two locally integrable functions f and g define the same distribution, *i.e.*

$$\langle T_f, \varphi\rangle = \langle T_K, \varphi\rangle \quad \text{for all } \varphi \in \mathscr{D},$$

if and only if they are equal almost everywhere. The classes of locally integrable functions modulo this equivalence form a vector space denoted $L^1_{\text{loc}}(\Omega)$; each element of $L^1_{\text{loc}}(\Omega)$ may therefore be identified with the distribution T_f defined by any one of its representatives f.

1.3.2.3.7. *Support of a distribution*

A distribution $T \in \mathscr{D}'(\Omega)$ is said to *vanish* on an open subset ω of Ω if it vanishes on all functions in $\mathscr{D}(\omega)$, *i.e.* if $\langle T, \varphi\rangle = 0$ whenever $\varphi \in \mathscr{D}(\omega)$.

The *support* of a distribution T, denoted Supp T, is then defined as the complement of the set-theoretic union of those open subsets ω on which T vanishes; or equivalently as the smallest closed subset of Ω outside which T vanishes.

When $T = T_f$ for $f \in L^1_{\text{loc}}(\Omega)$, then Supp T = Supp f, so that the two notions coincide. Clearly, if Supp T and Supp φ are disjoint subsets of Ω, then $\langle T, \varphi\rangle = 0$.

It can be shown that any distribution $T \in \mathscr{D}'$ with compact support may be extended from $\mathscr{D}$ to $\mathscr{E}$ while remaining continuous, so that $T \in \mathscr{E}'$; and that conversely, if $S \in \mathscr{E}'$, then its restriction T to $\mathscr{D}$ is a distribution with compact support. Thus, *the topological dual $\mathscr{E}'$ of $\mathscr{E}$ consists of those distributions in $\mathscr{D}'$ which have compact support*. This is intuitively clear since, if the condition of having compact support is fulfilled by T, it needs no longer be required of φ, which may then roam through $\mathscr{E}$ rather than $\mathscr{D}$.

1.3.2.3.8. *Convergence of distributions*

A *sequence* (T_j) of distributions will be said to converge in $\mathscr{D}'$ to a distribution T as $j \to \infty$ if, for any given $\varphi \in \mathscr{D}$, the sequence of complex numbers $(\langle T_j, \varphi\rangle)$ converges in $\mathbb{C}$ to the complex number $\langle T, \varphi\rangle$.

A *series* $\sum_{j=0}^{\infty} T_j$ of distributions will be said to converge in $\mathscr{D}'$ and to have distribution S as its sum if the sequence of partial sums $S_k = \sum_{j=0}^{k}$ converges to S.

These definitions of convergence in $\mathscr{D}'$ assume that the limits T and S are known in advance, and are distributions. This raises the question of the *completeness* of $\mathscr{D}'$: if a sequence (T_j) in $\mathscr{D}'$ is such that the sequence $(\langle T_j, \varphi\rangle)$ has a limit in $\mathbb{C}$ for all $\varphi \in \mathscr{D}$, does the map

$$\varphi \longmapsto \lim_{j\to\infty} \langle T_j, \varphi\rangle$$

define a distribution $T \in \mathscr{D}'$? In other words, does the limiting process preserve continuity with respect to φ? It is a remarkable theorem that, because of the strong topology on $\mathscr{D}$, this is actually the case. An analogous statement holds for series. This notion of convergence does not coincide with any of the classical notions used for ordinary functions: for example, the sequence (φ_ν) with $\varphi_\nu(x) = \cos \nu x$ converges to 0 in $\mathscr{D}'(\mathbb{R})$, but fails to do so by any of the standard criteria.

An example of convergent sequences of distributions is provided by sequences which converge to δ. If (f_ν) is a sequence of locally summable functions on $\mathbb{R}^n$ such that

(i) $\int_{\|\mathbf{x}\|<b} f_\nu(\mathbf{x})\, \mathrm{d}^n\mathbf{x} \to 1$ as $\nu \to \infty$ for all $b > 0$;

(ii) $\int_{a\le\|\mathbf{x}\|\le 1/a} |f_\nu(\mathbf{x})|\, \mathrm{d}^n\mathbf{x} \to 0$ as $\nu \to \infty$ for all $0 < a < 1$;

(iii) there exists $d > 0$ and $M > 0$ such that $\int_{\|\mathbf{x}\|<d} |f_\nu(\mathbf{x})|\, \mathrm{d}^n\mathbf{x} < M$ for all ν;

then the sequence (T_{f_ν}) of distributions converges to δ in $\mathscr{D}'(\mathbb{R}^n)$.

1.3.2.3.9. *Operations on distributions*

As a general rule, the definitions are chosen so that the operations coincide with those on functions whenever a distribution is associated to a function.

Most definitions consist in transferring to a distribution T an operation which is well defined on $\varphi \in \mathscr{D}$ by 'transposing' it in the duality product $\langle T, \varphi\rangle$; this procedure will map T to a new distribution provided the original operation maps $\mathscr{D}$ *continuously* into itself.

1.3.2.3.9.1. *Differentiation*

(*a*) *Definition and elementary properties*

If T is a distribution on $\mathbb{R}^n$, its partial derivative $\partial_i T$ with respect to x_i is defined by

$$\langle \partial_i T, \varphi\rangle = -\langle T, \partial_i \varphi\rangle$$

for all $\varphi \in \mathscr{D}$. This does define a distribution, because the partial differentiations $\varphi \longmapsto \partial_i\varphi$ are continuous for the topology of $\mathscr{D}$.

Suppose that $T = T_f$ with f a locally integrable function such that $\partial_i f$ exists and is almost everywhere continuous. Then integration by parts along the x_i axis gives

$$\begin{aligned}\int_{\mathbb{R}^n} &\partial_i f(x_1, \ldots, x_i, \ldots, x_n)\varphi(x_1, \ldots, x_i, \ldots, x_n)\, \mathrm{d}x_i \\ &= (f\varphi)(x_1, \ldots, +\infty, \ldots, x_n) - (f\varphi)(x_1, \ldots, -\infty, \ldots, x_n) \\ &\quad - \int_{\mathbb{R}^n} f(x_1, \ldots, x_i, \ldots, x_n)\partial_i\varphi(x_1, \ldots, x_i, \ldots, x_n)\, \mathrm{d}x_i;\end{aligned}$$

the integrated term vanishes, since φ has compact support, showing that $\partial_i T_f = T_{\partial_i f}$.

The test functions $\varphi \in \mathscr{D}$ are infinitely differentiable. Therefore, transpositions like that used to define $\partial_i T$ may be repeated, so that *any distribution is infinitely differentiable*. For instance,

$$\langle \partial^2_{ij} T, \varphi\rangle = -\langle \partial_j T, \partial_i\varphi\rangle = \langle T, \partial^2_{ij}\varphi\rangle,$$

$$\langle D^{\mathbf{p}} T, \varphi\rangle = (-1)^{|\mathbf{p}|}\langle T, D^{\mathbf{p}}\varphi\rangle,$$

$$\langle \Delta T, \varphi\rangle = \langle T, \Delta\varphi\rangle,$$

where Δ is the Laplacian operator. The derivatives of Dirac's δ distribution are

$$\langle D^{\mathbf{p}}\delta, \varphi\rangle = (-1)^{|\mathbf{p}|}\langle \delta, D^{\mathbf{p}}\varphi\rangle = (-1)^{|\mathbf{p}|} D^{\mathbf{p}}\varphi(\mathbf{0}).$$

It is remarkable that *differentiation is a continuous operation* for the topology on $\mathscr{D}'$: if a sequence (T_j) of distributions converges to distribution T, then the sequence $(D^{\mathbf{p}}T_j)$ of derivatives converges to $D^{\mathbf{p}}T$ for any multi-index $\mathbf{p}$, since as $j \to \infty$

$$\langle D^{\mathbf{p}}T_j, \varphi\rangle = (-1)^{|\mathbf{p}|}\langle T_j, D^{\mathbf{p}}\varphi\rangle \to (-1)^{|\mathbf{p}|}\langle T, D^{\mathbf{p}}\varphi\rangle = \langle D^{\mathbf{p}}T, \varphi\rangle.$$

An analogous statement holds for series: any convergent series of distributions may be differentiated termwise to all orders. This illustrates how 'robust' the constructs of distribution theory are in comparison with those of ordinary function theory, where similar statements are notoriously untrue.

(*b*) *Differentiation under the duality bracket*

Limiting processes and differentiation may also be carried out under the duality bracket $\langle , \rangle$ as under the integral sign with ordinary functions. Let the function $\varphi = \varphi(\mathbf{x}, \lambda)$ depend on a parameter $\lambda \in \Lambda$ and a vector $\mathbf{x} \in \mathbb{R}^n$ in such a way that all functions

$$\varphi_\lambda : \mathbf{x} \longmapsto \varphi(\mathbf{x}, \lambda)$$

be in $\mathscr{D}(\mathbb{R}^n)$ for all $\lambda \in \Lambda$. Let $T \in \mathscr{D}'(\mathbb{R}^n)$ be a distribution, let

$$I(\lambda) = \langle T, \varphi_\lambda\rangle$$

and let $\lambda_0 \in \Lambda$ be given parameter value. Suppose that, as λ runs through a small enough neighbourhood of λ_0,

(i) all the φ_λ have their supports in a fixed compact subset K of $\mathbb{R}^n$;

(ii) all the derivatives $D^{\mathbf{p}}\varphi_\lambda$ have a partial derivative with respect to λ which is continuous with respect to $\mathbf{x}$ and λ.

Under these hypotheses, $I(\lambda)$ is differentiable (in the usual sense) with respect to λ near λ_0, and its derivative may be obtained by 'differentiation under the $\langle , \rangle$ sign':

$$\frac{\mathrm{d}I}{\mathrm{d}\lambda} = \langle T, \partial_\lambda \varphi_\lambda \rangle.$$

(*c*) *Effect of discontinuities*

When a function f or its derivatives are no longer continuous, the derivatives $D^{\mathbf{p}}T_f$ of the associated distribution T_f may no longer coincide with the distributions associated to the functions $D^{\mathbf{p}}f$.

In dimension 1, the simplest example is Heaviside's unit step function Y [$Y(x) = 0$ for $x < 0, Y(x) = 1$ for $x \geq 0$]:

$$\langle (T_Y)', \varphi \rangle = -\langle (T_Y), \varphi' \rangle = -\int_0^{+\infty} \varphi'(x)\, \mathrm{d}x = \varphi(0) = \langle \delta, \varphi \rangle.$$

Hence $(T_Y)' = \delta$, a result long used 'heuristically' by electrical engineers [see also Dirac (1958)].

Let f be infinitely differentiable for $x < 0$ and $x > 0$ but have discontinuous derivatives $f^{(m)}$ at $x = 0$ [$f^{(0)}$ being f itself] with jumps $\sigma_m = f^{(m)}(0+) - f^{(m)}(0-)$. Consider the functions:

$$\begin{aligned} g_0 &= f - \sigma_0 Y \\ g_1 &= g_0' - \sigma_1 Y \\ &\text{- - - - - - - - - -} \\ g_k &= g_{k-1}' - \sigma_k Y. \end{aligned}$$

The g_k are continuous, their derivatives g_k' are continuous almost everywhere [which implies that $(T_{g_k})' = T_{g_k'}$ and $g_k' = f^{(k+1)}$ almost everywhere]. This yields immediately:

$$\begin{aligned} (T_f)' &= T_{f'} + \sigma_0 \delta \\ (T_f)'' &= T_{f''} + \sigma_0 \delta' + \sigma_1 \delta \\ &\text{- - - - - - - - - - - - - - - - - - -} \\ (T_f)^{(m)} &= T_{f^{(m)}} + \sigma_0 \delta^{(m-1)} + \ldots + \sigma_{m-1}\delta. \\ &\text{- - - - - - - - - - - - - - - - - - -} \end{aligned}$$

Thus the 'distributional derivatives' $(T_f)^{(m)}$ differ from the usual functional derivatives $T_{f^{(m)}}$ by singular terms associated with discontinuities.

In dimension n, let f be infinitely differentiable everywhere except on a smooth hypersurface S, across which its partial derivatives show discontinuities. Let σ_0 and σ_ν denote the discontinuities of f and its normal derivative $\partial_\nu \varphi$ across S (both σ_0 and σ_ν are functions of position on S), and let $\delta_{(S)}$ and $\partial_\nu \delta_{(S)}$ be defined by

$$\begin{aligned} \langle \delta_{(S)}, \varphi \rangle &= \int_S \varphi\, \mathrm{d}^{n-1}S \\ \langle \partial_\nu \delta_{(S)}, \varphi \rangle &= -\int_S \partial_\nu \varphi\, \mathrm{d}^{n-1}S. \end{aligned}$$

Integration by parts shows that

$$\partial_i T_f = T_{\partial_i f} + \sigma_0 \cos\theta_i \delta_{(S)},$$

where θ_i is the angle between the x_i axis and the normal to S along which the jump σ_0 occurs, and that the Laplacian of T_f is given by

$$\Delta(T_f) = T_{\Delta f} + \sigma_\nu \delta_{(S)} + \partial_\nu [\sigma_0 \delta_{(S)}].$$

The latter result is a statement of Green's theorem in terms of distributions. It will be used in Section 1.3.4.4.3.5 to calculate the Fourier transform of the indicator function of a molecular envelope.

1.3.2.3.9.2. *Integration of distributions in dimension 1*

The reverse operation from differentiation, namely calculating the 'indefinite integral' of a distribution S, consists in finding a distribution T such that $T' = S$.

For all $\chi \in \mathcal{D}$ such that $\chi = \psi'$ with $\psi \in \mathcal{D}$, we must have

$$\langle T, \chi \rangle = -\langle S, \psi \rangle.$$

This condition defines T in a 'hyperplane' $\mathcal{H}$ of $\mathcal{D}$, whose equation

$$\langle 1, \chi \rangle \equiv \langle 1, \psi' \rangle = 0$$

reflects the fact that ψ has compact support.

To specify T in the whole of $\mathcal{D}$, it suffices to specify the value of $\langle T, \varphi_0 \rangle$ where $\varphi_0 \in \mathcal{D}$ is such that $\langle 1, \varphi_0 \rangle = 1$: then any $\varphi \in \mathcal{D}$ may be written uniquely as

$$\varphi = \lambda \varphi_0 + \psi'$$

with

$$\lambda = \langle 1, \varphi \rangle, \qquad \chi = \varphi - \lambda \varphi_0, \qquad \psi(x) = \int_0^x \chi(t)\, \mathrm{d}t,$$

and T is defined by

$$\langle T, \varphi \rangle = \lambda \langle T, \varphi_0 \rangle - \langle S, \psi \rangle.$$

The freedom in the choice of φ_0 means that T is defined up to an additive constant.

1.3.2.3.9.3. *Multiplication of distributions by functions*

The product αT of a distribution T on $\mathbb{R}^n$ by a function α over $\mathbb{R}^n$ will be defined by transposition:

$$\langle \alpha T, \varphi \rangle = \langle T, \alpha \varphi \rangle \quad \text{for all } \varphi \in \mathcal{D}.$$

In order that αT be a distribution, the mapping $\varphi \longmapsto \alpha\varphi$ must send $\mathcal{D}(\mathbb{R}^n)$ continuously into itself; hence *the multipliers α must be infinitely differentiable*. The product of two general distributions cannot be defined. The need for a careful treatment of multipliers of distributions will become clear when it is later shown (Section 1.3.2.5.8) that the Fourier transformation turns convolutions into multiplications and *vice versa*.

If T is a distribution of order m, then α needs only have continuous derivatives up to order m. For instance, δ is a distribution of order zero, and $\alpha\delta = \alpha(\mathbf{0})\delta$ is a distribution provided α is continuous; this relation is of fundamental importance in the theory of sampling and of the properties of the Fourier transformation related to sampling (Sections 1.3.2.6.4, 1.3.2.6.6). More generally, $D^{\mathbf{p}}\delta$ is a distribution of order $|\mathbf{p}|$, and the following formula holds for all $\alpha \in \mathcal{D}^{(m)}$ with $m = |\mathbf{p}|$:

$$\alpha(D^{\mathbf{p}}\delta) = \sum_{\mathbf{q} \leq \mathbf{p}} (-1)^{|\mathbf{p}-\mathbf{q}|} \binom{\mathbf{p}}{\mathbf{q}} (D^{\mathbf{p}-\mathbf{q}}\alpha)(\mathbf{0}) D^{\mathbf{q}}\delta.$$

The derivative of a product is easily shown to be

$$\partial_i(\alpha T) = (\partial_i \alpha)T + \alpha(\partial_i T)$$

and generally for any multi-index $\mathbf{p}$

$$D^{\mathbf{p}}(\alpha T) = \sum_{\mathbf{q} \leq \mathbf{p}} \binom{\mathbf{p}}{\mathbf{q}} (D^{\mathbf{p}-\mathbf{q}}\alpha)(\mathbf{0}) D^{\mathbf{q}}T.$$

1.3.2.3.9.4. *Division of distributions by functions*

Given a distribution S on $\mathbb{R}^n$ and an infinitely differentiable multiplier function α, the division problem consists in finding a distribution T such that $\alpha T = S$.

If α never vanishes, $T = S/\alpha$ is the unique answer. If $n = 1$, and if α has only isolated zeros of finite order, it can be reduced to a collection of cases where the multiplier is x^m, for which the general solution can be shown to be of the form

$$T = U + \sum_{i=0}^{m-1} c_i \delta^{(i)},$$

where U is a particular solution of the division problem $x^m U = S$ and the c_i are arbitrary constants.

In dimension $n > 1$, the problem is much more difficult, but is of fundamental importance in the theory of linear partial differential equations, since the Fourier transformation turns the problem of solving these into a division problem for distributions [see Hörmander (1963)].

1.3.2.3.9.5. *Transformation of coordinates*

Let σ be a smooth non-singular change of variables in $\mathbb{R}^n$, *i.e.* an infinitely differentiable mapping from an open subset Ω of $\mathbb{R}^n$ to Ω' in $\mathbb{R}^n$, whose Jacobian

$$J(\sigma) = \det\left[\frac{\partial \sigma(\mathbf{x})}{\partial \mathbf{x}}\right]$$

vanishes nowhere in Ω. By the implicit function theorem, the inverse mapping σ^{-1} from Ω' to Ω is well defined.

If f is a locally summable function on Ω, then the function $\sigma^{\#} f$ defined by

$$(\sigma^{\#} f)(\mathbf{x}) = f[\sigma^{-1}(\mathbf{x})]$$

is a locally summable function on Ω', and for any $\varphi \in \mathscr{D}(\Omega')$ we may write:

$$\begin{aligned}\int_{\Omega'} (\sigma^{\#} f)(\mathbf{x})\varphi(\mathbf{x})\, \mathrm{d}^n\mathbf{x} &= \int_{\Omega'} f[\sigma^{-1}(\mathbf{x})]\varphi(\mathbf{x})\, \mathrm{d}^n\mathbf{x} \\ &= \int_{\Omega'} f(\mathbf{y})\varphi[\sigma(\mathbf{y})]|J(\sigma)|\, \mathrm{d}^n\mathbf{y} \quad \text{by } \mathbf{x} = \sigma(\mathbf{y}).\end{aligned}$$

In terms of the associated distributions

$$\langle T_{\sigma^{\#} f}, \varphi\rangle = \langle T_f, |J(\sigma)|(\sigma^{-1})^{\#}\varphi\rangle.$$

This operation can be extended to an arbitrary distribution T by defining its *image* $\sigma^{\#} T$ under coordinate transformation σ through

$$\langle \sigma^{\#} T, \varphi\rangle = \langle T, |J(\sigma)|(\sigma^{-1})^{\#}\varphi\rangle,$$

which is well defined provided that σ is *proper*, *i.e.* that $\sigma^{-1}(K)$ is compact whenever K is compact.

For instance, if $\sigma : \mathbf{x} \longmapsto \mathbf{x} + \mathbf{a}$ is a *translation* by a vector $\mathbf{a}$ in $\mathbb{R}^n$, then $|J(\sigma)| = 1$; $\sigma^{\#}$ is denoted by $\tau_{\mathbf{a}}$, and the translate $\tau_{\mathbf{a}} T$ of a distribution T is defined by

$$\langle \tau_{\mathbf{a}} T, \varphi\rangle = \langle T, \tau_{-\mathbf{a}}\varphi\rangle.$$

Let $A : \mathbf{x} \longmapsto \mathbf{A}\mathbf{x}$ be a linear transformation defined by a non-singular matrix $\mathbf{A}$. Then $J(A) = \det \mathbf{A}$, and

$$\langle A^{\#} T, \varphi\rangle = |\det \mathbf{A}|\langle T, (A^{-1})^{\#}\varphi\rangle.$$

This formula will be shown later (Sections 1.3.2.6.5, 1.3.4.2.1.1) to be the basis for the definition of the reciprocal lattice.

In particular, if $\mathbf{A} = -\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, A is an inversion through a centre of symmetry at the origin, and denoting $A^{\#}\varphi$ by $\breve{\varphi}$ we have:

$$\langle \breve{T}, \varphi\rangle = \langle T, \breve{\varphi}\rangle.$$

T is called an even distribution if $\breve{T} = T$, an odd distribution if $\breve{T} = -T$.

If $\mathbf{A} = \lambda\mathbf{I}$ with $\lambda > 0$, A is called a *dilation* and

$$\langle A^{\#} T, \varphi\rangle = \lambda^n \langle T, (A^{-1})^{\#}\varphi\rangle.$$

Writing symbolically δ as $\delta(\mathbf{x})$ and $A^{\#}\delta$ as $\delta(\mathbf{x}/\lambda)$, we have:

$$\delta(\mathbf{x}/\lambda) = \lambda^n \delta(\mathbf{x}).$$

If $n = 1$ and f is a function with isolated simple zeros x_j, then in the same symbolic notation

$$\delta[f(x)] = \sum_j \frac{1}{|f'(x_j)|}\delta(x_j),$$

where each $\lambda_j = 1/|f'(x_j)|$ is analogous to a 'Lorentz factor' at zero x_j.

1.3.2.3.9.6. *Tensor product of distributions*

The purpose of this construction is to extend Fubini's theorem to distributions. Following Section 1.3.2.2.5, we may define the tensor product $L^1_{\mathrm{loc}}(\mathbb{R}^m) \otimes L^1_{\mathrm{loc}}(\mathbb{R}^n)$ as the vector space of finite linear combinations of functions of the form

$$f \otimes g : (\mathbf{x}, \mathbf{y}) \longmapsto f(\mathbf{x})g(\mathbf{y}),$$

where $\mathbf{x} \in \mathbb{R}^m, \mathbf{y} \in \mathbb{R}^n, f \in L^1_{\mathrm{loc}}(\mathbb{R}^m)$ and $g \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$.

Let $S_{\mathbf{x}}$ and $T_{\mathbf{y}}$ denote the distributions associated to f and g, respectively, the subscripts $\mathbf{x}$ and $\mathbf{y}$ acting as mnemonics for $\mathbb{R}^m$ and $\mathbb{R}^n$. It follows from Fubini's theorem (Section 1.3.2.2.5) that $f \otimes g \in L^1_{\mathrm{loc}}(\mathbb{R}^m \times \mathbb{R}^n)$, and hence defines a distribution over $\mathbb{R}^m \times \mathbb{R}^n$; the rearrangement of integral signs gives

$$\langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle = \langle S_{\mathbf{x}}, \langle T_{\mathbf{y}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle\rangle = \langle T_{\mathbf{y}}, \langle S_{\mathbf{x}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle\rangle$$

for all $\varphi_{\mathbf{x},\mathbf{y}} \in \mathscr{D}(\mathbb{R}^m \times \mathbb{R}^n)$. In particular, if $\varphi(\mathbf{x}, \mathbf{y}) = u(\mathbf{x})v(\mathbf{y})$ with $u \in \mathscr{D}(\mathbb{R}^m), v \in \mathscr{D}(\mathbb{R}^n)$, then

$$\langle S \otimes T, u \otimes v\rangle = \langle S, u\rangle\langle T, v\rangle.$$

This construction can be extended to general distributions $S \in \mathscr{D}'(\mathbb{R}^m)$ and $T \in \mathscr{D}'(\mathbb{R}^n)$. Given any test function $\varphi \in \mathscr{D}(\mathbb{R}^m \times \mathbb{R}^n)$, let $\varphi_{\mathbf{x}}$ denote the map $\mathbf{y} \longmapsto \varphi(\mathbf{x}, \mathbf{y})$; let $\varphi_{\mathbf{y}}$ denote the map $\mathbf{x} \longmapsto \varphi(\mathbf{x}, \mathbf{y})$; and define the two functions $\theta(\mathbf{x}) = \langle T, \varphi_{\mathbf{x}}\rangle$ and $\omega(\mathbf{y}) = \langle S, \varphi_{\mathbf{y}}\rangle$. Then, by the lemma on differentiation under the $\langle , \rangle$ sign of Section 1.3.2.3.9.1, $\theta \in \mathscr{D}(\mathbb{R}^m), \omega \in \mathscr{D}(\mathbb{R}^n)$, and there exists a unique distribution $S \otimes T$ such that

$$\langle S \otimes T, \varphi\rangle = \langle S, \theta\rangle = \langle T, \omega\rangle.$$

$S \otimes T$ is called the *tensor product* of S and T.

With the mnemonic introduced above, this definition reads identically to that given above for distributions associated to locally integrable functions:

$$\langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle = \langle S_{\mathbf{x}}, \langle T_{\mathbf{y}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle\rangle = \langle T_{\mathbf{y}}, \langle S_{\mathbf{x}}, \varphi_{\mathbf{x},\mathbf{y}}\rangle\rangle.$$

The tensor product of distributions is associative:

$$(R \otimes S) \otimes T = R \otimes (S \otimes T).$$

Derivatives may be calculated by

$$D^{\mathbf{p}}_{\mathbf{x}} D^{\mathbf{q}}_{\mathbf{y}}(S_{\mathbf{x}} \otimes T_{\mathbf{y}}) = (D^{\mathbf{p}}_{\mathbf{x}} S_{\mathbf{x}}) \otimes (D^{\mathbf{q}}_{\mathbf{y}} T_{\mathbf{y}}).$$

The support of a tensor product is the Cartesian product of the supports of the two factors.

1.3.2.3.9.7. *Convolution of distributions*

The convolution $f * g$ of two functions f and g on $\mathbb{R}^n$ is defined by

$$(f * g)(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{y})g(\mathbf{x} - \mathbf{y})\, \mathrm{d}^n\mathbf{y} = \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y})g(\mathbf{y})\, \mathrm{d}^n\mathbf{y}$$

whenever the integral exists. This is the case when f and g are both in $L^1(\mathbb{R}^n)$; then $f * g$ is also in $L^1(\mathbb{R}^n)$. Let S, T and W denote the distributions associated to f, g and $f * g$, respectively: a change of variable immediately shows that for any $\varphi \in \mathscr{D}(\mathbb{R}^n)$,

$$\langle W, \varphi \rangle = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(\mathbf{x}) g(\mathbf{y}) \varphi(\mathbf{x} + \mathbf{y}) \, \mathrm{d}^n\mathbf{x} \, \mathrm{d}^n\mathbf{y}.$$

Introducing the map σ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n$ defined by $\sigma(\mathbf{x}, \mathbf{y}) = \mathbf{x} + \mathbf{y}$, the latter expression may be written:

$$\langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi \circ \sigma \rangle$$

(where $\circ$ denotes the composition of mappings) or by a slight abuse of notation:

$$\langle W, \varphi \rangle = \langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi(\mathbf{x} + \mathbf{y}) \rangle.$$

A difficulty arises in extending this definition to general distributions S and T because the mapping σ is not proper: if K is compact in $\mathbb{R}^n$, then $\sigma^{-1}(K)$ is a cylinder with base K and generator the 'second bisector' $\mathbf{x} + \mathbf{y} = \mathbf{0}$ in $\mathbb{R}^n \times \mathbb{R}^n$. However, $\langle S \otimes T, \varphi \circ \sigma \rangle$ is defined whenever the intersection between Supp $(S \otimes T) =$ (Supp S) $\times$ (Supp T) and σ^{-1}(Supp φ) is compact.

We may therefore define the *convolution* $S * T$ of two distributions S and T on $\mathbb{R}^n$ by

$$\langle S * T, \varphi \rangle = \langle S \otimes T, \varphi \circ \sigma \rangle = \langle S_{\mathbf{x}} \otimes T_{\mathbf{y}}, \varphi(\mathbf{x} + \mathbf{y}) \rangle$$

whenever the following *support condition* is fulfilled:

'the set $\{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \in A, \mathbf{y} \in B, \mathbf{x} + \mathbf{y} \in K\}$ is compact in $\mathbb{R}^n \times \mathbb{R}^n$ for all K compact in $\mathbb{R}^n$'.

The latter condition is met, in particular, if S or T has compact support. The support of $S * T$ is easily seen to be contained in the closure of the vector sum

$$A + B = \{\mathbf{x} + \mathbf{y} | \mathbf{x} \in A, \mathbf{y} \in B\}.$$

Convolution by a fixed distribution S is a *continuous* operation for the topology on $\mathscr{D}'$: it maps convergent sequences (T_j) to convergent sequences $(S * T_j)$. Convolution is commutative: $S * T = T * S$.

The convolution of p distributions $T_1, \ldots, T_p$ with supports $A_1, \ldots, A_p$ can be defined by

$$\langle T_1 * \ldots * T_p, \varphi \rangle = \langle (T_1)_{\mathbf{x}_1} \otimes \ldots \otimes (T_p)_{\mathbf{x}_p}, \varphi(\mathbf{x}_1 + \ldots + \mathbf{x}_p) \rangle$$

whenever the following generalized support condition:

'the set $\{(\mathbf{x}_1, \ldots, \mathbf{x}_p) | \mathbf{x}_1 \in A_1, \ldots, \mathbf{x}_p \in A_p, \mathbf{x}_1 + \ldots + \mathbf{x}_p \in K\}$ is compact in $(\mathbb{R}^n)^p$ for all K compact in $\mathbb{R}^n$'

is satisfied. It is then associative. Interesting examples of associativity failure, which can be traced back to violations of the support condition, may be found in Bracewell (1986, pp. 436–437).

It follows from previous definitions that, for all distributions $T \in \mathscr{D}'$, the following identities hold:

(i) $\delta * T = T$: δ is the unit convolution;

(ii) $\delta_{(\mathbf{a})} * T = \tau_{\mathbf{a}} T$: translation is a convolution with the corresponding translate of δ;

(iii) $(D^{\mathbf{p}}\delta) * T = D^{\mathbf{p}} T$: differentiation is a convolution with the corresponding derivative of δ;

(iv) translates or derivatives of a convolution may be obtained by translating or differentiating any one of the factors: convolution 'commutes' with translation and differentiation, a property used in Section 1.3.4.4.7.7 to speed up least-squares model refinement for macromolecules.

The latter property is frequently used for the purpose of *regularization*: if T is a distribution, α an infinitely differentiable function, and at least one of the two has compact support, then $T * \alpha$ is an infinitely differentiable ordinary function. Since sequences (α_ν) of such functions α can be constructed which have compact support and converge to δ, it follows that any distribution T can be obtained as the limit of infinitely differentiable functions $T * \alpha_\nu$. In topological jargon: $\mathscr{D}(\mathbb{R}^n)$ is 'everywhere dense' in $\mathscr{D}'(\mathbb{R}^n)$. A standard function in $\mathscr{D}$ which is often used for such proofs is defined as follows: put

$$\theta(x) = \frac{1}{A} \exp\left(-\frac{1}{1 - x^2}\right) \quad \text{for } |x| \leq 1,$$

$$= 0 \quad \text{for } |x| \geq 1,$$

with

$$A = \int_{-1}^{+1} \exp\left(-\frac{1}{1 - x^2}\right) \mathrm{d}x$$

(so that θ is in $\mathscr{D}$ and is normalized), and put

$$\theta_\varepsilon(x) = \frac{1}{\varepsilon} \theta\left(\frac{x}{\varepsilon}\right) \quad \text{in dimension 1,}$$

$$\theta_\varepsilon(\mathbf{x}) = \prod_{j=1}^{n} \theta_\varepsilon(x_j) \quad \text{in dimension } n.$$

Another related result, also proved by convolution, is the *structure theorem*: the restriction of a distribution $T \in \mathscr{D}'(\mathbb{R}^n)$ to a bounded open set Ω in $\mathbb{R}^n$ is a derivative of finite order of a continuous function.

Properties (i) to (iv) are the basis of the symbolic or operational calculus (see Carslaw & Jaeger, 1948; Van der Pol & Bremmer, 1955; Churchill, 1958; Erdélyi, 1962; Moore, 1971) for solving integro-differential equations with constant coefficients by turning them into convolution equations, then using factorization methods for convolution algebras (Schwartz, 1965).

1.3.2.4. *Fourier transforms of functions*

1.3.2.4.1. *Introduction*

Given a complex-valued function f on $\mathbb{R}^n$ subject to suitable regularity conditions, its Fourier transform $\mathscr{F}[f]$ and Fourier cotransform $\bar{\mathscr{F}}[f]$ are defined as follows:

$$\mathscr{F}[f](\xi) = \int_{\mathbb{R}^n} f(\mathbf{x}) \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{x}) \, \mathrm{d}^n\mathbf{x}$$

$$\bar{\mathscr{F}}[f](\xi) = \int_{\mathbb{R}^n} f(\mathbf{x}) \exp(+2\pi i \boldsymbol{\xi} \cdot \mathbf{x}) \, \mathrm{d}^n\mathbf{x},$$

where $\boldsymbol{\xi} \cdot \mathbf{x} = \sum_{i=1}^{n} \xi_i x_i$ is the ordinary scalar product. The terminology and sign conventions given above are the standard ones in mathematics; those used in crystallography are slightly different (see Section 1.3.4.2.1.1). These transforms enjoy a number of remarkable properties, whose natural settings entail different regularity assumptions on f: for instance, properties relating to convolution are best treated in $L^1(\mathbb{R}^n)$, while Parseval's theorem requires the Hilbert space structure of $L^2(\mathbb{R}^n)$. After a brief review of these classical properties, the Fourier transformation will be examined in a space $\mathscr{S}(\mathbb{R}^n)$ particularly well suited to accommodating the full range of its properties, which will later serve as a space of test functions to extend the Fourier transformation to distributions.

There exists an abundant literature on the 'Fourier integral'. The books by Carslaw (1930), Wiener (1933), Titchmarsh (1948), Katznelson (1968), Sneddon (1951, 1972), and Dym & McKean (1972) are particularly recommended.

1.3.2.4.2. *Fourier transforms in L^1*

1.3.2.4.2.1. *Linearity*

Both transformations $\mathscr{F}$ and $\bar{\mathscr{F}}$ are obviously linear maps from L^1 to L^∞ when these spaces are viewed as vector spaces over the field $\mathbb{C}$ of complex numbers.

1.3.2.4.2.2. *Effect of affine coordinate transformations*

$\mathscr{F}$ and $\bar{\mathscr{F}}$ turn translations into phase shifts:

$$\mathscr{F}[\tau_{\mathbf{a}} f](\boldsymbol{\xi}) = \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{a})\mathscr{F}[f](\boldsymbol{\xi})$$
$$\bar{\mathscr{F}}[\tau_{\mathbf{a}} f](\boldsymbol{\xi}) = \exp(+2\pi i \boldsymbol{\xi}\cdot\mathbf{a})\bar{\mathscr{F}}[f](\boldsymbol{\xi}).$$

Under a general linear change of variable $\mathbf{x} \longmapsto \mathbf{A}\mathbf{x}$ with non-singular matrix $\mathbf{A}$, the transform of $A^{\#}f$ is

$$\begin{aligned}\mathscr{F}[A^{\#}f](\boldsymbol{\xi}) &= \int_{\mathbb{R}^n} f(\mathbf{A}^{-1}\mathbf{x})\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{x})\,\mathrm{d}^n\mathbf{x}\\ &= \int_{\mathbb{R}^n} f(\mathbf{y})\exp(-2\pi i(A^T\boldsymbol{\xi})\cdot\mathbf{y})|\det\mathbf{A}|\,\mathrm{d}^n\mathbf{y}\\ &\qquad\qquad\qquad\qquad\text{by } \mathbf{x}=\mathbf{A}\mathbf{y}\\ &= |\det\mathbf{A}|\mathscr{F}[f](\mathbf{A}^T\boldsymbol{\xi})\end{aligned}$$

i.e.

$$\mathscr{F}[A^{\#}f] = |\det\mathbf{A}|[(\mathbf{A}^{-1})^T]^{\#}\mathscr{F}[f]$$

and similarly for $\bar{\mathscr{F}}$. The matrix $(\mathbf{A}^{-1})^T$ is called the *contragredient* of matrix $\mathbf{A}$.

Under an affine change of coordinates $\mathbf{x} \longmapsto S(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ with non-singular matrix $\mathbf{A}$, the transform of $S^{\#}f$ is given by

$$\begin{aligned}\mathscr{F}[S^{\#}f](\boldsymbol{\xi}) &= \mathscr{F}[\tau_{\mathbf{b}}(A^{\#}f)](\boldsymbol{\xi})\\ &= \exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{b})\mathscr{F}[A^{\#}f](\boldsymbol{\xi})\\ &= \exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{b})|\det\mathbf{A}|\mathscr{F}[f](\mathbf{A}^T\boldsymbol{\xi})\end{aligned}$$

with a similar result for $\bar{\mathscr{F}}$, replacing $-i$ by $+i$.

1.3.2.4.2.3. *Conjugate symmetry*

The kernels of the Fourier transformations $\mathscr{F}$ and $\bar{\mathscr{F}}$ satisfy the following identities:

$$\exp(\pm 2\pi i\boldsymbol{\xi}\cdot\mathbf{x}) = \exp\overline{[\pm 2\pi i\boldsymbol{\xi}\cdot(-\mathbf{x})]} = \exp\overline{[\pm 2\pi i(-\boldsymbol{\xi})\cdot\mathbf{x}]}.$$

As a result the transformations $\mathscr{F}$ and $\bar{\mathscr{F}}$ themselves have the following 'conjugate symmetry' properties [where the notation $\breve{f}(\mathbf{x}) = f(-\mathbf{x})$ of Section 1.3.2.2.2 will be used]:

$$\mathscr{F}[f](\boldsymbol{\xi}) = \overline{\mathscr{F}[\bar{f}](-\boldsymbol{\xi})} = \overline{\mathscr{F}[\breve{\bar{f}}](\boldsymbol{\xi})}$$
$$\mathscr{F}[f](\boldsymbol{\xi}) = \overline{\mathscr{F}[\breve{f}](\boldsymbol{\xi})}.$$

Therefore,

(i) f real $\Leftrightarrow f = \bar{f} \Leftrightarrow \mathscr{F}[f] = \overline{\mathscr{F}[f]} \Leftrightarrow \mathscr{F}[f](\boldsymbol{\xi}) = \overline{\mathscr{F}[f](-\boldsymbol{\xi})}$: $\mathscr{F}[f]$ is said to possess *Hermitian symmetry*;

(ii) f centrosymmetric $\Leftrightarrow f = \breve{f} \Leftrightarrow \mathscr{F}[f] = \overline{\mathscr{F}[\bar{f}]}$;

(iii) f real centrosymmetric $\Leftrightarrow f = \bar{f} = \breve{f} \Leftrightarrow \mathscr{F}[f] = \overline{\mathscr{F}[f]} = \overline{\mathscr{F}[\breve{f}]} \Leftrightarrow \mathscr{F}[f]$ real centrosymmetric.

Conjugate symmetry is the basis of Friedel's law (Section 1.3.4.2.1.4) in crystallography.

1.3.2.4.2.4. *Tensor product property*

Another elementary property of $\mathscr{F}$ is its naturality with respect to tensor products. Let $u \in L^1(\mathbb{R}^m)$ and $v \in L^1(\mathbb{R}^n)$, and let $\mathscr{F}_{\mathbf{x}}, \mathscr{F}_{\mathbf{y}}, \mathscr{F}_{\mathbf{x},\mathbf{y}}$ denote the Fourier transformations in $L^1(\mathbb{R}^m), L^1(\mathbb{R}^n)$ and $L^1(\mathbb{R}^m\times\mathbb{R}^n)$, respectively. Then

$$\mathscr{F}_{\mathbf{x},\mathbf{y}}[u\otimes v] = \mathscr{F}_{\mathbf{x}}[u]\otimes\mathscr{F}_{\mathbf{y}}[v].$$

Furthermore, if $f \in L^1(\mathbb{R}^m\times\mathbb{R}^n)$, then $\mathscr{F}_{\mathbf{y}}[f] \in L^1(\mathbb{R}^m)$ as a function of $\mathbf{x}$ and $\mathscr{F}_{\mathbf{x}}[f] \in L^1(\mathbb{R}^n)$ as a function of $\mathbf{y}$, and

$$\mathscr{F}_{\mathbf{x},\mathbf{y}}[f] = \mathscr{F}_{\mathbf{x}}[\mathscr{F}_{\mathbf{y}}[f]] = \mathscr{F}_{\mathbf{y}}[\mathscr{F}_{\mathbf{x}}[f]].$$

This is easily proved by using Fubini's theorem and the fact that $(\boldsymbol{\xi},\boldsymbol{\eta})\cdot(\mathbf{x},\mathbf{y}) = \boldsymbol{\xi}\cdot\mathbf{x} + \boldsymbol{\eta}\cdot\mathbf{y}$, where $\mathbf{x},\boldsymbol{\xi}\in\mathbb{R}^m, \mathbf{y},\boldsymbol{\eta}\in\mathbb{R}^n$. This property may be written:

$$\mathscr{F}_{\mathbf{x},\mathbf{y}} = \mathscr{F}_{\mathbf{x}}\otimes\mathscr{F}_{\mathbf{y}}.$$

1.3.2.4.2.5. *Convolution property*

If f and g are summable, their convolution $f * g$ exists and is summable, and

$$\mathscr{F}[f*g](\boldsymbol{\xi}) = \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^n} f(\mathbf{y})g(\mathbf{x}-\mathbf{y})\,\mathrm{d}^n\mathbf{y}\right]\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{x})\,\mathrm{d}^n\mathbf{x}.$$

With $\mathbf{x} = \mathbf{y} + \mathbf{z}$, so that

$$\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{x}) = \exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{y})\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{z}),$$

and with Fubini's theorem, rearrangement of the double integral gives:

$$\mathscr{F}[f*g] = \mathscr{F}[f]\times\mathscr{F}[g]$$

and similarly

$$\bar{\mathscr{F}}[f*g] = \bar{\mathscr{F}}[f]\times\bar{\mathscr{F}}[g].$$

Thus *the Fourier transform and cotransform turn convolution into multiplication*.

1.3.2.4.2.6. *Reciprocity property*

In general, $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are not summable, and hence cannot be further transformed; however, as they are essentially bounded, their products with the Gaussians $G_t(\xi) = \exp(-2\pi^2\|\xi\|^2 t)$ are summable for all $t > 0$, and it can be shown that

$$f = \lim_{t\to 0}\bar{\mathscr{F}}[G_t\mathscr{F}[f]] = \lim_{t\to 0}\mathscr{F}[G_t\bar{\mathscr{F}}[f]],$$

where the limit is taken in the topology of the L^1 norm $\|.\|_1$. Thus $\mathscr{F}$ and $\bar{\mathscr{F}}$ are (in a sense) mutually inverse, which justifies the common practice of calling $\bar{\mathscr{F}}$ the 'inverse Fourier transformation'.

1.3.2.4.2.7. *Riemann–Lebesgue lemma*

If $f \in L^1(\mathbb{R}^n)$, *i.e.* is summable, then $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ exist and are continuous and essentially bounded:

$$\|\mathscr{F}[f]\|_\infty = \|\bar{\mathscr{F}}[f]\|_\infty \le \|f\|_1.$$

In fact one has the much stronger property, whose statement constitutes the *Riemann–Lebesgue lemma*, that $\mathscr{F}[f](\boldsymbol{\xi})$ and $\bar{\mathscr{F}}[f](\boldsymbol{\xi})$ both tend to zero as $\|\boldsymbol{\xi}\|\to\infty$.

1.3.2.4.2.8. *Differentiation*

Let us now suppose that $n = 1$ and that $f \in L^1(\mathbb{R})$ is differentiable with $f' \in L^1(\mathbb{R})$. Integration by parts yields

$$\begin{aligned}\mathscr{F}[f'](\xi) &= \int_{-\infty}^{+\infty} f'(x)\exp(-2\pi i\xi\cdot x)\,\mathrm{d}x\\ &= [f(x)\exp(-2\pi i\xi\cdot x)]_{-\infty}^{+\infty}\\ &\quad + 2\pi i\xi\int_{-\infty}^{+\infty} f(x)\exp(-2\pi i\xi\cdot x)\,\mathrm{d}x.\end{aligned}$$

Since f' is summable, f has a limit when $x\to\pm\infty$, and this limit must be 0 since f is summable. Therefore

$$\mathscr{F}[f'](\xi) = (2\pi i\xi)\mathscr{F}[f](\xi)$$

with the bound

$$\|2\pi\xi\mathscr{F}[f]\|_\infty \leq \|f'\|_1$$

so that $|\mathscr{F}[f](\xi)|$ decreases faster than $1/|\xi| \to \infty$.

This result can be easily extended to several dimensions and to any multi-index $\mathbf{m}$: if f is summable and has continuous summable partial derivatives up to order $|\mathbf{m}|$, then

$$\mathscr{F}[D^{\mathbf{m}}f](\boldsymbol{\xi}) = (2\pi i\boldsymbol{\xi})^{\mathbf{m}}\mathscr{F}[f](\boldsymbol{\xi})$$

and

$$\|(2\pi\boldsymbol{\xi})^{\mathbf{m}}\mathscr{F}[f]\|_\infty \leq \|D^{\mathbf{m}}f\|_1.$$

Similar results hold for $\bar{\mathscr{F}}$, with $2\pi i\boldsymbol{\xi}$ replaced by $-2\pi i\boldsymbol{\xi}$. Thus, the more differentiable f is, with summable derivatives, the faster $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ decrease at infinity.

The property of turning differentiation into multiplication by a monomial has many important applications in crystallography, for instance differential syntheses (Sections 1.3.4.2.1.9, 1.3.4.4.7.2, 1.3.4.4.7.5) and moment-generating functions [Section 1.3.4.5.2.1(*c*)].

1.3.2.4.2.9. *Decrease at infinity*

Conversely, assume that f is summable on $\mathbb{R}^n$ and that f decreases fast enough at infinity for $\mathbf{x}^{\mathbf{m}}f$ also to be summable, for some multi-index $\mathbf{m}$. Then the integral defining $\mathscr{F}[f]$ may be subjected to the differential operator $D^{\mathbf{m}}$, still yielding a convergent integral: therefore $D^{\mathbf{m}}\mathscr{F}[f]$ exists, and

$$D^{\mathbf{m}}(\mathscr{F}[f])(\boldsymbol{\xi}) = \mathscr{F}[(-2\pi i\mathbf{x})^{\mathbf{m}}f](\boldsymbol{\xi})$$

with the bound

$$\|D^{\mathbf{m}}(\mathscr{F}[f])\|_\infty = \|(2\pi\mathbf{x})^{\mathbf{m}}f\|_1.$$

Similar results hold for $\bar{\mathscr{F}}$, with $-2\pi i\mathbf{x}$ replaced by $2\pi i\mathbf{x}$. Thus, the faster f decreases at infinity, the more $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are differentiable, with bounded derivatives. This property is the converse of that described in Section 1.3.2.4.2.8, and their combination is fundamental in the definition of the function space $\mathscr{S}$ in Section 1.3.2.4.4.1, of tempered distributions in Section 1.3.2.5, and in the extension of the Fourier transformation to them.

1.3.2.4.2.10. *The Paley–Wiener theorem*

An extreme case of the last instance occurs when f has *compact support*: then $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are so regular that they may be analytically continued from $\mathbb{R}^n$ to $\mathbb{C}^n$ where they are *entire* functions, *i.e.* have no singularities at finite distance (Paley & Wiener, 1934). This is easily seen for $\mathscr{F}[f]$: giving vector $\boldsymbol{\xi} \in \mathbb{R}^n$ a vector $\boldsymbol{\eta} \in \mathbb{R}^n$ of imaginary parts leads to

$$\begin{aligned}\mathscr{F}[f](\boldsymbol{\xi} + i\boldsymbol{\eta}) &= \int_{\mathbb{R}^n} f(\mathbf{x})\exp[-2\pi i(\boldsymbol{\xi} + i\boldsymbol{\eta})\cdot\mathbf{x}]\,\mathrm{d}^n\mathbf{x} \\ &= \mathscr{F}[\exp(2\pi\boldsymbol{\eta}\cdot\mathbf{x})f](\boldsymbol{\xi}),\end{aligned}$$

where the latter transform always exists since $\exp(2\pi\boldsymbol{\eta}\cdot\mathbf{x})f$ is summable with respect to $\mathbf{x}$ for all values of $\boldsymbol{\eta}$. This analytic continuation forms the basis of the saddlepoint method in probability theory [Section 1.3.4.5.2.1(f)] and leads to the use of maximum-entropy distributions in the statistical theory of direct phase determination [Section 1.3.4.5.2.2(*e*)].

By Liouville's theorem, an entire function in $\mathbb{C}^n$ cannot vanish identically on the complement of a compact subset of $\mathbb{R}^n$ without vanishing everywhere: therefore $\mathscr{F}[f]$ cannot have compact support if f has, and hence $\mathscr{D}(\mathbb{R}^n)$ *is not stable by Fourier transformation*.

1.3.2.4.3. *Fourier transforms in* L^2

Let f belong to $L^2(\mathbb{R}^n)$, *i.e.* be such that

$$\|f\|_2 = \left(\int_{\mathbb{R}^n} |f(\mathbf{x})|^2\,\mathrm{d}^n\mathbf{x}\right)^{1/2} < \infty.$$

1.3.2.4.3.1. *Invariance of* L^2

$\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ exist and are functions in L^2, *i.e.* $\mathscr{F}L^2 = L^2$, $\bar{\mathscr{F}}L^2 = L^2$.

1.3.2.4.3.2. *Reciprocity*

$\mathscr{F}[\bar{\mathscr{F}}[f]] = f$ and $\bar{\mathscr{F}}[\mathscr{F}[f]] = f$, equality being taken as 'almost everywhere' equality. This again leads to calling $\bar{\mathscr{F}}$ the 'inverse Fourier transformation' rather than the Fourier cotransformation.

1.3.2.4.3.3. *Isometry*

$\mathscr{F}$ and $\bar{\mathscr{F}}$ preserve the L^2 norm:

$\|\mathscr{F}[f]\|_2 = \|\bar{\mathscr{F}}[f]\|_2 = \|f\|_2$ (Parseval's/Plancherel's theorem).

This property, which may be written in terms of the inner product $(,)$ in $L^2(\mathbb{R}^n)$ as

$$(\mathscr{F}[f], \mathscr{F}[g]) = (\bar{\mathscr{F}}[f], \bar{\mathscr{F}}[g]) = (f, g),$$

implies that $\mathscr{F}$ and $\bar{\mathscr{F}}$ are *unitary* transformations of $L^2(\mathbb{R}^n)$ into itself, *i.e.* infinite-dimensional 'rotations'.

1.3.2.4.3.4. *Eigenspace decomposition of* L^2

Some light can be shed on the geometric structure of these rotations by the following simple considerations. Note that

$$\begin{aligned}\mathscr{F}^2[f](\mathbf{x}) &= \int_{\mathbb{R}^n} \mathscr{F}[f](\boldsymbol{\xi})\exp(-2\pi i\mathbf{x}\cdot\boldsymbol{\xi})\,\mathrm{d}^n\boldsymbol{\xi} \\ &= \bar{\mathscr{F}}[\mathscr{F}[f]](-\mathbf{x}) = f(-\mathbf{x})\end{aligned}$$

so that $\mathscr{F}^4$ (and similarly $\bar{\mathscr{F}}^4$) is the identity map. Any eigenvalue of $\mathscr{F}$ or $\bar{\mathscr{F}}$ is therefore a fourth root of unity, *i.e.* ± 1 or $\pm i$, and $L^2(\mathbb{R}^n)$ splits into an orthogonal direct sum

$$\mathbf{H}_0 \otimes \mathbf{H}_1 \otimes \mathbf{H}_2 \otimes \mathbf{H}_3,$$

where $\mathscr{F}$ (respectively $\bar{\mathscr{F}}$) acts in each subspace $\mathbf{H}_k (k = 0, 1, 2, 3)$ by multiplication by $(-i)^k$. Orthonormal bases for these subspaces can be constructed from Hermite functions (*cf.* Section 1.3.2.4.4.2) This method was used by Wiener (1933, pp. 51–71).

1.3.2.4.3.5. *The convolution theorem and the isometry property*

In L^2, the convolution theorem (when applicable) and the Parseval/Plancherel theorem are not independent. Suppose that f, g, $f \times g$ and $f * g$ are all in L^2 (without questioning whether these properties are independent). Then $f * g$ may be written in terms of the inner product in L^2 as follows:

$$(f * g)(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y})g(\mathbf{y})\,\mathrm{d}^n\mathbf{y} = \int_{\mathbb{R}^n} \overline{\breve{f}(\mathbf{y}-\mathbf{x})}g(\mathbf{y})\,\mathrm{d}^n\mathbf{y},$$

i.e.

$$(f * g)(\mathbf{x}) = (\tau_{\mathbf{x}}\breve{f}, g).$$

Invoking the isometry property, we may rewrite the right-hand side as

$$\begin{aligned}(\mathscr{F}[\tau_{\mathbf{x}}\check{\bar{f}}],\mathscr{F}[g]) &= (\exp(-2\pi i\mathbf{x}\cdot\boldsymbol{\xi})\overline{\mathscr{F}[f]}_{\boldsymbol{\xi}},\mathscr{F}[g]_{\boldsymbol{\xi}})\\ &= \int_{\mathbb{R}^n}(\mathscr{F}[f]\times\mathscr{F}[g])(\mathbf{x})\\ &\quad\times\exp(+2\pi i\mathbf{x}\cdot\boldsymbol{\xi})\,\mathrm{d}^n\boldsymbol{\xi}\\ &= \bar{\mathscr{F}}[\mathscr{F}[f]\times\mathscr{F}[g]],\end{aligned}$$

so that the initial identity yields the convolution theorem.

To obtain the converse implication, note that

$$\begin{aligned}(f,g) &= \int_{\mathbb{R}^n}\overline{f(\mathbf{y})}g(\mathbf{y})\,\mathrm{d}^n\mathbf{y} = (\check{\bar{f}}*g)(\mathbf{0})\\ &= \bar{\mathscr{F}}[\mathscr{F}[\check{\bar{f}}]\times\mathscr{F}[g]](\mathbf{0})\\ &= \int_{\mathbb{R}^n}\overline{\mathscr{F}[f](\boldsymbol{\xi})}\mathscr{F}[g](\boldsymbol{\xi})\,\mathrm{d}^n\boldsymbol{\xi} = (\mathscr{F}[f],\mathscr{F}[g]),\end{aligned}$$

where conjugate symmetry (Section 1.3.2.4.2.2) has been used.

These relations have an important application in the calculation by Fourier transform methods of the derivatives used in the refinement of macromolecular structures (Section 1.3.4.4.7).

1.3.2.4.4. *Fourier transforms in $\mathscr{S}$*

1.3.2.4.4.1. *Definition and properties of $\mathscr{S}$*

The duality established in Sections 1.3.2.4.2.8 and 1.3.2.4.2.9 between the local differentiability of a function and the rate of decrease at infinity of its Fourier transform prompts one to consider the space $\mathscr{S}(\mathbb{R}^n)$ of functions f on $\mathbb{R}^n$ which are infinitely differentiable *and* all of whose derivatives are rapidly decreasing, so that for all multi-indices $\mathbf{k}$ and $\mathbf{p}$

$$(\mathbf{x}^{\mathbf{k}}D^{\mathbf{p}}f)(\mathbf{x})\to 0\quad\text{as }\|\mathbf{x}\|\to\infty.$$

The product of $f\in\mathscr{S}$ by any polynomial over $\mathbb{R}^n$ is still in $\mathscr{S}$ ($\mathscr{S}$ is an algebra over the ring of polynomials). Furthermore, $\mathscr{S}$ is invariant under translations and differentiation.

If $f\in\mathscr{S}$, then its transforms $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are

(i) infinitely differentiable because f is rapidly decreasing;

(ii) rapidly decreasing because f is infinitely differentiable;

hence $\mathscr{F}[f]$ and $\bar{\mathscr{F}}[f]$ are in $\mathscr{S}$: *$\mathscr{S}$ is invariant* under $\mathscr{F}$ and $\bar{\mathscr{F}}$.

Since $L^1\supset\mathscr{S}$ and $L^2\supset\mathscr{S}$, all properties of $\mathscr{F}$ and $\bar{\mathscr{F}}$ already encountered above are enjoyed by functions of $\mathscr{S}$, with all restrictions on differentiability and/or integrability lifted. For instance, given two functions f and g in $\mathscr{S}$, then both fg and $f*g$ are in $\mathscr{S}$ (which was not the case with L^1 nor with L^2) so that the reciprocity theorem inherited from L^2

$$\mathscr{F}[\bar{\mathscr{F}}[f]]=f\quad\text{and}\quad\bar{\mathscr{F}}[\mathscr{F}[f]]=f$$

allows one to state the reverse of the convolution theorem first established in L^1:

$$\begin{aligned}\mathscr{F}[fg] &= \mathscr{F}[f]*\mathscr{F}[g]\\ \bar{\mathscr{F}}[fg] &= \bar{\mathscr{F}}[f]*\bar{\mathscr{F}}[g].\end{aligned}$$

1.3.2.4.4.2. *Gaussian functions and Hermite functions*

Gaussian functions are particularly important elements of $\mathscr{S}$. In dimension 1, a well known contour integration (Schwartz, 1965, p. 184) yields

$$\mathscr{F}[\exp(-\pi x^2)](\xi)=\bar{\mathscr{F}}[\exp(-\pi x^2)](\xi)=\exp(-\pi\xi^2),$$

which shows that the 'standard Gaussian' $\exp(-\pi x^2)$ is invariant under $\mathscr{F}$ and $\bar{\mathscr{F}}$. By a tensor product construction, it follows that the same is true of the standard Gaussian

$$G(\mathbf{x})=\exp(-\pi\|\mathbf{x}\|^2)$$

in dimension n:

$$\mathscr{F}[G](\boldsymbol{\xi})=\bar{\mathscr{F}}[G](\boldsymbol{\xi})=G(\boldsymbol{\xi}).$$

In other words, G is an eigenfunction of $\mathscr{F}$ and $\bar{\mathscr{F}}$ for eigenvalue 1 (Section 1.3.2.4.3.4).

A complete system of eigenfunctions may be constructed as follows. In dimension 1, consider the family of functions

$$H_m=\frac{D^mG^2}{G}\quad(m\geq 0),$$

where D denotes the differentiation operator. The first two members of the family

$$H_0=G,\qquad H_1=2DG,$$

are such that $\mathscr{F}[H_0]=H_0$, as shown above, and

$$DG(x)=-2\pi xG(x)=i(2\pi ix)G(x)=i\mathscr{F}[DG](x),$$

hence

$$\mathscr{F}[H_1]=(-i)H_1.$$

We may thus take as an induction hypothesis that

$$\mathscr{F}[H_m]=(-i)^mH_m.$$

The identity

$$D\left(\frac{D^mG^2}{G}\right)=\frac{D^{m+1}G^2}{G}-\frac{DG}{G}\frac{D^mG^2}{G}$$

may be written

$$H_{m+1}(x)=(DH_m)(x)-2\pi xH_m(x),$$

and the two differentiation theorems give:

$$\begin{aligned}\mathscr{F}[DH_m](\xi) &= (2\pi i\xi)\mathscr{F}[H_m](\xi)\\ \mathscr{F}[-2\pi xH_m](\xi) &= -iD(\mathscr{F}[H_m])(\xi).\end{aligned}$$

Combination of this with the induction hypothesis yields

$$\begin{aligned}\mathscr{F}[H_{m+1}](\xi) &= (-i)^{m+1}[(DH_m)(\xi)-2\pi\xi H_m(\xi)]\\ &= (-i)^{m+1}H_{m+1}(\xi),\end{aligned}$$

thus proving that H_m is an eigenfunction of $\mathscr{F}$ for eigenvalue $(-i)^m$ for all $m\geq 0$. The same proof holds for $\bar{\mathscr{F}}$, with eigenvalue i^m. If these eigenfunctions are normalized as

$$\mathscr{H}_m(x)=\frac{(-1)^m2^{1/4}}{\sqrt{m!}2^m\pi^{m/2}}H_m(x),$$

then it can be shown that the collection of *Hermite functions* $\{\mathscr{H}_m(x)\}_{m\geq 0}$ constitutes an orthonormal basis of $L^2(\mathbb{R})$ such that $\mathscr{H}_m$ is an eigenfunction of $\mathscr{F}$ (respectively $\bar{\mathscr{F}}$) for eigenvalue $(-i)^m$ (respectively i^m).

In dimension n, the same construction can be extended by tensor product to yield the multivariate Hermite functions

$$\mathscr{H}_{\mathbf{m}}(\mathbf{x})=\mathscr{H}_{m_1}(x_1)\times\mathscr{H}_{m_2}(x_2)\times\ldots\times\mathscr{H}_{m_n}(x_n)$$

(where $\mathbf{m}\geq\mathbf{0}$ is a multi-index). These constitute an orthonormal basis of $L^2(\mathbb{R}^n)$, with $\mathscr{H}_{\mathbf{m}}$ an eigenfunction of $\mathscr{F}$ (respectively $\bar{\mathscr{F}}$) for eigenvalue $(-i)^{|\mathbf{m}|}$ (respectively $i^{|\mathbf{m}|}$). Thus the subspaces $\mathbf{H}_k$ of Section 1.3.2.4.3.4 are spanned by those $\mathscr{H}_{\mathbf{m}}$ with $|\mathbf{m}|\equiv k \bmod 4$ $(k=0,1,2,3)$.

General multivariate Gaussians are usually encountered in the non-standard form

$$G_{\mathbf{A}}(\mathbf{x})=\exp(-\tfrac{1}{2}\mathbf{x}^T\cdot\mathbf{A}\mathbf{x}),$$

where $\mathbf{A}$ is a symmetric positive-definite matrix. Diagonalizing $\mathbf{A}$ as $\mathbf{E}\Lambda\mathbf{E}^T$ with $\mathbf{E}\mathbf{E}^T$ the identity matrix, and putting $\mathbf{A}^{1/2} = \mathbf{E}\Lambda^{1/2}\mathbf{E}^T$, we may write

$$G_{\mathbf{A}}(\mathbf{x}) = G\left[\left(\frac{\mathbf{A}}{2\pi}\right)^{1/2}\mathbf{x}\right]$$

i.e.

$$G_{\mathbf{A}} = [(2\pi\mathbf{A}^{-1})^{1/2}]^{\#}G;$$

hence (by Section 1.3.2.4.2.3)

$$\mathscr{F}[G_{\mathbf{A}}] = |\det\,(2\pi\mathbf{A}^{-1})|^{1/2}\left[\left(\frac{\mathbf{A}}{2\pi}\right)^{1/2}\right]^{\#}G,$$

i.e.

$$\mathscr{F}[G_{\mathbf{A}}](\boldsymbol{\xi}) = |\det\,(2\pi\mathbf{A}^{-1})|^{1/2}G[(2\pi\mathbf{A}^{-1})^{1/2}\boldsymbol{\xi}],$$

i.e. finally

$$\mathscr{F}[G_{\mathbf{A}}] = |\det\,(2\pi\mathbf{A}^{-1})|^{1/2}G_{4\pi^2\mathbf{A}^{-1}}.$$

This result is widely used in crystallography, *e.g.* to calculate form factors for anisotropic atoms (Section 1.3.4.2.2.6) and to obtain transforms of derivatives of Gaussian atomic densities (Section 1.3.4.4.7.10).

1.3.2.4.4.3. *Heisenberg's inequality, Hardy's theorem*

The result just obtained, which also holds for $\bar{\mathscr{F}}$, shows that the 'peakier' $G_{\mathbf{A}}$, the 'broader' $\mathscr{F}[G_{\mathbf{A}}]$. This is a general property of the Fourier transformation, expressed in dimension 1 by the *Heisenberg inequality* (Weyl, 1931):

$$\left(\int x^2|f(x)|^2\,\mathrm{d}x\right)\left(\int \xi^2|\mathscr{F}[f](\xi)|^2\,\mathrm{d}\xi\right) \geq \frac{1}{16\pi^2}\left(\int |f(x)|^2\,\mathrm{d}x\right)^2,$$

where, by a beautiful theorem of Hardy (1933), equality can only be attained for f Gaussian. Hardy's theorem is even stronger: if both f and $\mathscr{F}[f]$ behave at infinity as constant multiples of G, then each of them is *everywhere* a constant multiple of G; if both f and $\mathscr{F}[f]$ behave at infinity as constant multiples of $G \times$ monomial, then each of them is a finite linear combination of Hermite functions. Hardy's theorem is invoked in Section 1.3.4.4.5 to derive the optimal procedure for spreading atoms on a sampling grid in order to obtain the most accurate structure factors.

The search for optimal compromises between the confinement of f to a compact domain in x-space and of $\mathscr{F}[f]$ to a compact domain in ξ-space leads to consideration of prolate spheroidal wavefunctions (Pollack & Slepian, 1961; Landau & Pollack, 1961, 1962).

1.3.2.4.4.4. *Symmetry property*

A final formal property of the Fourier transform, best established in $\mathscr{S}$, is its *symmetry*: if f and g are in $\mathscr{S}$, then by Fubini's theorem

$$\begin{aligned}\langle\mathscr{F}[f], g\rangle &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} f(\mathbf{x})\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{x})\,\mathrm{d}^n\mathbf{x}\right)g(\boldsymbol{\xi})\,\mathrm{d}^n\boldsymbol{\xi}\\ &= \int_{\mathbb{R}^n} f(\mathbf{x})\left(\int_{\mathbb{R}^n} g(\boldsymbol{\xi})\exp(-2\pi i\boldsymbol{\xi}\cdot\mathbf{x})\,\mathrm{d}^n\boldsymbol{\xi}\right)\mathrm{d}^n\mathbf{x}\\ &= \langle f, \mathscr{F}[g]\rangle.\end{aligned}$$

This possibility of 'transposing' $\mathscr{F}$ (and $\bar{\mathscr{F}}$) from the left to the right of the duality bracket will be used in Section 1.3.2.5.4 to extend the Fourier transformation to distributions.

1.3.2.4.5. *Various writings of Fourier transforms*

Other ways of writing Fourier transforms in $\mathbb{R}^n$ exist besides the one used here. All have the form

$$\mathscr{F}_{h,\omega}[f](\boldsymbol{\xi}) = \frac{1}{h^n}\int_{\mathbb{R}^n} f(\mathbf{x})\exp(-i\omega\boldsymbol{\xi}\cdot\mathbf{x})\,\mathrm{d}^n\mathbf{x},$$

where h is real positive and ω real non-zero, with the reciprocity formula written:

$$f(\mathbf{x}) = \frac{1}{k^n}\int_{\mathbb{R}^n}\mathscr{F}_{h,\omega}[f](\boldsymbol{\xi})\exp(+i\omega\boldsymbol{\xi}\cdot\mathbf{x})\,\mathrm{d}^n\mathbf{x}$$

with k real positive. The consistency condition between h, k and ω is

$$hk = \frac{2\pi}{|\omega|}.$$

The usual choices are:

(i) $\omega = \pm 2\pi, h = k = 1$ (as here);

(ii) $\omega = \pm 1, h = 1, k = 2\pi$ (in probability theory and in solid-state physics);

(iii) $\omega = \pm 1, h = k = \sqrt{2\pi}$ (in much of classical analysis).

It should be noted that conventions (ii) and (iii) introduce numerical factors of 2π in convolution and Parseval formulae, while (ii) breaks the symmetry between $\mathscr{F}$ and $\bar{\mathscr{F}}$.

1.3.2.4.6. *Tables of Fourier transforms*

The books by Campbell & Foster (1948), Erdélyi (1954), and Magnus *et al.* (1966) contain extensive tables listing pairs of functions and their Fourier transforms. Bracewell (1986) lists those pairs particularly relevant to electrical engineering applications.

1.3.2.5. *Fourier transforms of tempered distributions*

1.3.2.5.1. *Introduction*

It was found in Section 1.3.2.4.2 that the usual space of test functions $\mathscr{D}$ is not invariant under $\mathscr{F}$ and $\bar{\mathscr{F}}$. By contrast, the space $\mathscr{S}$ of infinitely differentiable rapidly decreasing functions *is* invariant under $\mathscr{F}$ and $\bar{\mathscr{F}}$, and furthermore transposition formulae such as

$$\langle\mathscr{F}[f], g\rangle = \langle f, \mathscr{F}[g]\rangle$$

hold for all $f, g \in \mathscr{S}$. It is precisely this type of transposition which was used successfully in Sections 1.3.2.3.9.1 and 1.3.2.3.9.3 to define the derivatives of distributions and their products with smooth functions.

This suggests using $\mathscr{S}$ instead of $\mathscr{D}$ as a space of test functions φ, and defining the Fourier transform $\mathscr{F}[T]$ of a distribution T by

$$\langle\mathscr{F}[T], \varphi\rangle = \langle T, \mathscr{F}[\varphi]\rangle$$

whenever T is capable of being extended from $\mathscr{D}$ to $\mathscr{S}$ while remaining continuous. It is this latter proviso which will be subsumed under the adjective 'tempered'. As was the case with the construction of $\mathscr{D}'$, it is the definition of a sufficiently strong topology (*i.e.* notion of convergence) in $\mathscr{S}$ which will play a key role in transferring to the elements of its topological dual $\mathscr{S}'$ (called tempered distributions) all the properties of the Fourier transformation.

Besides the general references to distribution theory mentioned in Section 1.3.2.3.1 the reader may consult the books by Zemanian (1965, 1968). Lavoine (1963) contains tables of Fourier transforms of distributions.

1.3.2.5.2. $\mathscr{S}$ *as a test-function space*

A notion of convergence has to be introduced in $\mathscr{S}(\mathbb{R}^n)$ in order to be able to define and test the continuity of linear functionals on it.

A sequence (φ_j) of functions in $\mathscr{S}$ will be said to converge to 0 if, for any given multi-indices $\mathbf{k}$ and $\mathbf{p}$, the sequence $(\mathbf{x}^{\mathbf{k}} D^{\mathbf{p}} \varphi_j)$ tends to 0 uniformly on $\mathbb{R}^n$.

It can be shown that $\mathscr{D}(\mathbb{R}^n)$ is dense in $\mathscr{S}(\mathbb{R}^n)$. Translation is continuous for this topology. For any linear differential operator $P(D) = \sum_{\mathbf{p}} a_{\mathbf{p}} D^{\mathbf{p}}$ and any polynomial $Q(\mathbf{x})$ over $\mathbb{R}^n$, $(\varphi_j) \to 0$ implies $[Q(\mathbf{x}) \times P(D)\varphi_j] \to 0$ in the topology of $\mathscr{S}$. Therefore, differentiation and multiplication by polynomials are continuous for the topology on $\mathscr{S}$.

The Fourier transformations $\mathscr{F}$ and $\bar{\mathscr{F}}$ are also continuous for the topology of $\mathscr{S}$. Indeed, let (φ_j) converge to 0 for the topology on $\mathscr{S}$. Then, by Section 1.3.2.4.2,

$$\|(2\pi\boldsymbol{\xi})^{\mathbf{m}} D^{\mathbf{p}}(\mathscr{F}[\varphi_j])\|_\infty \leq \|D^{\mathbf{m}}[(2\pi\mathbf{x})^{\mathbf{p}}\varphi_j]\|_1.$$

The right-hand side tends to 0 as $j \to \infty$ by definition of convergence in $\mathscr{S}$, hence $\|\boldsymbol{\xi}\|^{\mathbf{m}} D^{\mathbf{p}}(\mathscr{F}[\varphi_j]) \to 0$ uniformly, so that $(\mathscr{F}[\varphi_j]) \to 0$ in $\mathscr{S}$ as $j \to \infty$. The same proof applies to $\bar{\mathscr{F}}$.

1.3.2.5.3. *Definition and examples of tempered distributions*

A distribution $T \in \mathscr{D}'(\mathbb{R}^n)$ is said to be *tempered* if it can be extended into a continuous linear functional on $\mathscr{S}$.

If $\mathscr{S}'(\mathbb{R}^n)$ is the topological dual of $\mathscr{S}(\mathbb{R}^n)$, and if $S \in \mathscr{S}'(\mathbb{R}^n)$, then its restriction to $\mathscr{D}$ is a tempered distribution; conversely, if $T \in \mathscr{D}'$ is tempered, then its extension to $\mathscr{S}$ is unique (because $\mathscr{D}$ is dense in $\mathscr{S}$), hence it defines an element S of $\mathscr{S}'$. We may therefore identify $\mathscr{S}'$ and the space of tempered distributions.

A distribution with compact support is tempered, *i.e.* $\mathscr{S}' \supset \mathscr{E}'$. By transposition of the corresponding properties of $\mathscr{S}$, it is readily established that the derivative, translate or product by a polynomial of a tempered distribution is still a tempered distribution.

These inclusion relations may be summarized as follows: since $\mathscr{S}$ contains $\mathscr{D}$ but is contained in $\mathscr{E}$, the reverse inclusions hold for the topological duals, and hence $\mathscr{S}'$ contains $\mathscr{E}'$ but is contained in $\mathscr{D}'$.

A locally summable function f on $\mathbb{R}^n$ will be said to be *of polynomial growth* if $|f(\mathbf{x})|$ can be majorized by a polynomial in $\|\mathbf{x}\|$ as $\|\mathbf{x}\| \to \infty$. It is easily shown that such a function f defines a tempered distribution T_f *via*

$$\langle T_f, \varphi \rangle = \int_{\mathbb{R}^n} f(\mathbf{x})\varphi(\mathbf{x})\, \mathrm{d}^n\mathbf{x}.$$

In particular, polynomials over $\mathbb{R}^n$ define tempered distributions, and so do functions in $\mathscr{S}$. The latter remark, together with the transposition identity (Section 1.3.2.4.4), invites the extension of $\mathscr{F}$ and $\bar{\mathscr{F}}$ from $\mathscr{S}$ to $\mathscr{S}'$.

1.3.2.5.4. *Fourier transforms of tempered distributions*

The Fourier transform $\mathscr{F}[T]$ and cotransform $\bar{\mathscr{F}}[T]$ of a tempered distribution T are defined by

$$\langle \mathscr{F}[T], \varphi \rangle = \langle T, \mathscr{F}[\varphi] \rangle$$
$$\langle \bar{\mathscr{F}}[T], \varphi \rangle = \langle T, \bar{\mathscr{F}}[\varphi] \rangle$$

for all test functions $\varphi \in \mathscr{S}$. Both $\mathscr{F}[T]$ and $\bar{\mathscr{F}}[T]$ are themselves tempered distributions, since the maps $\varphi \longmapsto \mathscr{F}[\varphi]$ and $\varphi \longmapsto \bar{\mathscr{F}}[\varphi]$ are both linear and continuous for the topology of $\mathscr{S}$. In the same way that $\mathbf{x}$ and $\boldsymbol{\xi}$ have been used consistently as arguments for φ and $\mathscr{F}[\varphi]$, respectively, the notation $T_{\mathbf{x}}$ and $\mathscr{F}[T]_{\boldsymbol{\xi}}$ will be used to indicate which variables are involved.

When T is a distribution with compact support, its Fourier transform may be written

$$\mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}} = \langle T_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{x}) \rangle$$

since the function $\mathbf{x} \longmapsto \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{x})$ is in $\mathscr{E}$ while $T_{\mathbf{x}} \in \mathscr{E}'$. It can be shown, as in Section 1.3.2.4.2, to be analytically continuable into an entire function over $\mathbb{C}^n$.

1.3.2.5.5. *Transposition of basic properties*

The duality between differentiation and multiplication by a monomial extends from $\mathscr{S}$ to $\mathscr{S}'$ by transposition:

$$\mathscr{F}[D^{\mathbf{p}}_{\mathbf{x}} T_{\mathbf{x}}]_{\boldsymbol{\xi}} = (2\pi i \boldsymbol{\xi})^{\mathbf{p}} \mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}}$$
$$D^{\mathbf{p}}_{\boldsymbol{\xi}}(\mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}}) = \mathscr{F}[(-2\pi i \mathbf{x})^{\mathbf{p}} T_{\mathbf{x}}]_{\boldsymbol{\xi}}.$$

Analogous formulae hold for $\bar{\mathscr{F}}$, with i replaced by $-i$.

The formulae expressing the duality between translation and phase shift, *e.g.*

$$\mathscr{F}[\tau_{\mathbf{a}} T_{\mathbf{x}}]_{\boldsymbol{\xi}} = \exp(-2\pi i \mathbf{a} \cdot \boldsymbol{\xi}) \mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}}$$
$$\tau_{\boldsymbol{\alpha}}(\mathscr{F}[T_{\mathbf{x}}]_{\boldsymbol{\xi}}) = \mathscr{F}[\exp(2\pi i \boldsymbol{\alpha} \cdot \mathbf{x}) T_{\mathbf{x}}]_{\boldsymbol{\xi}};$$

between a linear change of variable and its contragredient, *e.g.*

$$\mathscr{F}[A^{\#} T] = |\det \mathbf{A}| [(\mathbf{A}^{-1})^T]^{\#} \mathscr{F}[T];$$

are obtained similarly by transposition from the corresponding identities in $\mathscr{S}$. They give a transposition formula for an affine change of variables $\mathbf{x} \longmapsto S(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ with non-singular matrix $\mathbf{A}$:

$$\begin{aligned}\mathscr{F}[S^{\#} T] &= \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{b}) \mathscr{F}[A^{\#} T] \\ &= \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{b}) |\det \mathbf{A}| [(\mathbf{A}^{-1})^T]^{\#} \mathscr{F}[T],\end{aligned}$$

with a similar result for $\bar{\mathscr{F}}$, replacing $-i$ by $+i$.

Conjugate symmetry is obtained similarly:

$$\mathscr{F}[\bar{T}] = \overset{\smile}{\overline{\mathscr{F}[T]}}, \mathscr{F}[\breve{\bar{T}}] = \overline{\mathscr{F}[T]},$$

with the same identities for $\bar{\mathscr{F}}$.

The tensor product property also transposes to tempered distributions: if $U \in \mathscr{S}'(\mathbb{R}^m), V \in \mathscr{S}'(\mathbb{R}^n)$,

$$\mathscr{F}[U_{\mathbf{x}} \otimes V_{\mathbf{y}}] = \mathscr{F}[U]_{\boldsymbol{\xi}} \otimes \mathscr{F}[V]_{\boldsymbol{\eta}}$$
$$\bar{\mathscr{F}}[U_{\mathbf{x}} \otimes V_{\mathbf{y}}] = \bar{\mathscr{F}}[U]_{\boldsymbol{\xi}} \otimes \bar{\mathscr{F}}[V]_{\boldsymbol{\eta}}.$$

1.3.2.5.6. *Transforms of δ-functions*

Since δ has compact support,

$$\mathscr{F}[\delta_{\mathbf{x}}]_{\boldsymbol{\xi}} = \langle \delta_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{x}) \rangle = 1_{\boldsymbol{\xi}}, \quad \textit{i.e.}\ \mathscr{F}[\delta] = 1.$$

It is instructive to show that conversely $\mathscr{F}[1] = \delta$ without invoking the reciprocity theorem. Since $\partial_j 1 = 0$ for all $j = 1, \ldots, n$, it follows from Section 1.3.2.3.9.4 that $\mathscr{F}[1] = c\delta$; the constant c can be determined by using the invariance of the standard Gaussian G established in Section 1.3.2.4.3:

$$\langle \mathscr{F}[1]_{\mathbf{x}}, G_{\mathbf{x}} \rangle = \langle 1_{\boldsymbol{\xi}}, G_{\boldsymbol{\xi}} \rangle = 1;$$

hence $c = 1$. Thus, $\mathscr{F}[1] = \delta$.

The basic properties above then read (using multi-indices to denote differentiation):

$$\mathscr{F}[\delta_{\mathbf{x}}^{(\mathbf{m})}]_{\boldsymbol{\xi}} = (2\pi i\boldsymbol{\xi})^{\mathbf{m}}, \quad \mathscr{F}[\mathbf{x}^{\mathbf{m}}]_{\boldsymbol{\xi}} = (-2\pi i)^{-|\mathbf{m}|}\delta_{\boldsymbol{\xi}}^{(\mathbf{m})};$$
$$\mathscr{F}[\delta_{\mathbf{a}}]_{\boldsymbol{\xi}} = \exp(-2\pi i\mathbf{a}\cdot\boldsymbol{\xi}), \quad \mathscr{F}[\exp(2\pi i\boldsymbol{\alpha}\cdot\mathbf{x})]_{\boldsymbol{\xi}} = \delta_{\boldsymbol{\alpha}},$$

with analogous relations for $\bar{\mathscr{F}}$, i becoming $-i$. Thus derivatives of δ are mapped to monomials (and *vice versa*), while translates of δ are mapped to 'phase factors' (and *vice versa*).

1.3.2.5.7. *Reciprocity theorem*

The previous results now allow a self-contained and rigorous proof of the reciprocity theorem between $\mathscr{F}$ and $\bar{\mathscr{F}}$ to be given, whereas in traditional settings (*i.e.* in L^1 and L^2) the implicit handling of δ through a limiting process is always the sticking point.

Reciprocity is first established in $\mathscr{S}$ as follows:

$$\begin{aligned}\bar{\mathscr{F}}[\mathscr{F}[\varphi]](\mathbf{x}) &= \int_{\mathbb{R}^n} \mathscr{F}[\varphi](\boldsymbol{\xi})\exp(2\pi i\boldsymbol{\xi}\cdot\mathbf{x})\ \mathrm{d}^n\boldsymbol{\xi}\\ &= \int_{\mathbb{R}^n} \mathscr{F}[\tau_{-\mathbf{x}}\varphi](\boldsymbol{\xi})\ \mathrm{d}^n\boldsymbol{\xi}\\ &= \langle 1, \mathscr{F}[\tau_{-\mathbf{x}}\varphi]\rangle\\ &= \langle \mathscr{F}[1], \tau_{-\mathbf{x}}\varphi\rangle\\ &= \langle \tau_{\mathbf{x}}\delta, \varphi\rangle\\ &= \varphi(\mathbf{x})\end{aligned}$$

and similarly

$$\mathscr{F}[\bar{\mathscr{F}}[\varphi]](\mathbf{x}) = \varphi(\mathbf{x}).$$

The reciprocity theorem is then proved in $\mathscr{S}'$ by transposition:

$$\bar{\mathscr{F}}[\mathscr{F}[T]] = \mathscr{F}[\bar{\mathscr{F}}[T]] = T \quad \text{for all } T \in \mathscr{S}'.$$

Thus the Fourier cotransformation $\bar{\mathscr{F}}$ in $\mathscr{S}'$ may legitimately be called the 'inverse Fourier transformation'.

The method of Section 1.3.2.4.3 may then be used to show that $\mathscr{F}$ and $\bar{\mathscr{F}}$ both have period 4 in $\mathscr{S}'$.

1.3.2.5.8. *Multiplication and convolution*

Multiplier functions $\alpha(\mathbf{x})$ for tempered distributions must be infinitely differentiable, as for ordinary distributions; furthermore, they must grow sufficiently slowly as $\|x\| \to \infty$ to ensure that $\alpha\varphi \in \mathscr{S}$ for all $\varphi \in \mathscr{S}$ and that the map $\varphi \longmapsto \alpha\varphi$ is continuous for the topology of $\mathscr{S}$. This leads to choosing for multipliers the subspace $\mathscr{O}_M$ consisting of functions $\alpha \in \mathscr{E}$ *of polynomial growth*. It can be shown that if f is in $\mathscr{O}_M$, then the associated distribution T_f is in $\mathscr{S}'$ (*i.e.* is a tempered distribution); and that conversely if T is in $\mathscr{S}'$, $\mu * T$ is in $\mathscr{O}_M$ for all $\mu \in \mathscr{D}$.

Corresponding restrictions must be imposed to define the space $\mathscr{O}'_C$ of those distributions T whose convolution $S * T$ with a tempered distribution S is still a tempered distribution: T must be such that, for all $\varphi \in \mathscr{S}, \theta(\mathbf{x}) = \langle T_{\mathbf{y}}, \varphi(\mathbf{x}+\mathbf{y})\rangle$ is in $\mathscr{S}$; and such that the map $\varphi \longmapsto \theta$ be continuous for the topology of $\mathscr{S}$. This implies that S is 'rapidly decreasing'. It can be shown that if f is in $\mathscr{S}$, then the associated distribution T_f is in $\mathscr{O}'_C$; and that conversely if T is in $\mathscr{O}'_C$, $\mu * T$ is in $\mathscr{S}$ for all $\mu \in \mathscr{D}$.

The two spaces $\mathscr{O}_M$ and $\mathscr{O}'_C$ are mapped into each other by the Fourier transformation

$$\mathscr{F}(\mathscr{O}_M) = \bar{\mathscr{F}}(\mathscr{O}_M) = \mathscr{O}'_C$$
$$\mathscr{F}(\mathscr{O}'_C) = \bar{\mathscr{F}}(\mathscr{O}'_C) = \mathscr{O}_M$$

and the convolution theorem takes the form

$$\mathscr{F}[\alpha S] = \mathscr{F}[\alpha] * \mathscr{F}[S] \quad S \in \mathscr{S}', \alpha \in \mathscr{O}_M, \mathscr{F}[\alpha] \in \mathscr{O}'_C;$$
$$\mathscr{F}[S * T] = \mathscr{F}[S] \times \mathscr{F}[T] \quad S \in \mathscr{S}', T \in \mathscr{O}'_C, \mathscr{F}[T] \in \mathscr{O}_M.$$

The same identities hold for $\bar{\mathscr{F}}$. Taken together with the reciprocity theorem, these show that $\mathscr{F}$ and $\bar{\mathscr{F}}$ establish mutually inverse isomorphisms between $\mathscr{O}_M$ and $\mathscr{O}'_C$, and exchange multiplication for convolution in $\mathscr{S}'$.

It may be noticed that most of the basic properties of $\mathscr{F}$ and $\bar{\mathscr{F}}$ may be deduced from this theorem and from the properties of δ. Differentiation operators $D^{\mathbf{m}}$ and translation operators $\tau_{\mathbf{a}}$ are convolutions with $D^{\mathbf{m}}\delta$ and $\tau_{\mathbf{a}}\delta$; they are turned, respectively, into multiplication by monomials $(\pm 2\pi i\boldsymbol{\xi})^{\mathbf{m}}$ (the transforms of $D^{\mathbf{m}}\delta$) or by phase factors $\exp(\pm 2\pi i\boldsymbol{\xi}\cdot\boldsymbol{\alpha})$ (the transforms of $\tau_{\mathbf{a}}\delta$).

Another consequence of the convolution theorem is the duality established by the Fourier transformation between sections and projections of a function and its transform. For instance, in $\mathbb{R}^3$, the *projection* of $f(x,y,z)$ on the x, y plane along the z axis may be written

$$(\delta_x \otimes \delta_y \otimes 1_z) * f;$$

its Fourier transform is then

$$(1_\xi \otimes 1_\eta \otimes \delta_\zeta) \times \mathscr{F}[f],$$

which is the *section* of $\mathscr{F}[f]$ by the plane $\zeta = 0$, orthogonal to the z axis used for projection. There are numerous applications of this property in crystallography (Section 1.3.4.2.1.8) and in fibre diffraction (Section 1.3.4.5.1.3).

1.3.2.5.9. *L^2 aspects, Sobolev spaces*

The special properties of $\mathscr{F}$ in the space of square-integrable functions $L^2(\mathbb{R}^n)$, such as Parseval's identity, can be accommodated within distribution theory: if $u \in L^2(\mathbb{R}^n)$, then T_u is a tempered distribution in $\mathscr{S}'$ (the map $u \longmapsto T_u$ being continuous) and it can be shown that $S = \mathscr{F}[T_u]$ is of the form S_v, where $u = \mathscr{F}[u]$ is the Fourier transform of u in $L^2(\mathbb{R}^n)$. By Plancherel's theorem, $\|u\|_2 = \|v\|_2$.

This embedding of L^2 into $\mathscr{S}'$ can be used to derive the convolution theorem for L^2. If u and v are in $L^2(\mathbb{R}^n)$, then $u * v$ can be shown to be a bounded continuous function; thus $u * v$ is not in L^2, but it is in $\mathscr{S}'$, so that its Fourier transform is a distribution, and

$$\mathscr{F}[u * v] = \mathscr{F}[u] \times \mathscr{F}[v].$$

Spaces of tempered distributions related to $L^2(\mathbb{R}^n)$ can be defined as follows. For any real s, define the Sobolev space $H_s(\mathbb{R}^n)$ to consist of all tempered distributions $S \in \mathscr{S}'(\mathbb{R}^n)$ such that

$$(1 + |\boldsymbol{\xi}|^2)^{s/2}\mathscr{F}[S]_{\boldsymbol{\xi}} \in L^2(\mathbb{R}^n).$$

These spaces play a fundamental role in the theory of partial differential equations, and in the mathematical theory of tomographic reconstruction – a subject not unrelated to the crystallographic phase problem (Natterer, 1986).

1.3.2.6. *Periodic distributions and Fourier series*

1.3.2.6.1. *Terminology*

Let $\mathbb{Z}^n$ be the subset of $\mathbb{R}^n$ consisting of those points with (signed) integer coordinates; it is an n-dimensional *lattice*, *i.e.* a free Abelian group on n generators. A particularly simple set of n generators is given by the standard basis of $\mathbb{R}^n$, and hence $\mathbb{Z}^n$ will be called the *standard lattice in* $\mathbb{R}^n$. Any other 'non-standard' n-dimensional lattice Λ in $\mathbb{R}^n$ is the image of this standard lattice by a general linear transformation.

If we identify any two points in $\mathbb{R}^n$ whose coordinates are congruent modulo $\mathbb{Z}^n$, *i.e.* differ by a vector in $\mathbb{Z}^n$, we obtain the *standard n-torus* $\mathbb{R}^n/\mathbb{Z}^n$. The latter may be viewed as $(\mathbb{R}/\mathbb{Z})^n$, *i.e.* as the Cartesian product of n circles. The same identification may be carried out modulo a non-standard lattice Λ, yielding a *non-*

standard n-torus $\mathbb{R}^n/\Lambda$. The correspondence to crystallographic terminology is that 'standard' coordinates over the standard 3-torus $\mathbb{R}^3/\mathbb{Z}^3$ are called 'fractional' coordinates over the unit cell; while Cartesian coordinates, *e.g.* in ångströms, constitute a set of non-standard coordinates.

Finally, we will denote by I the unit cube $[0,1]^n$ and by C_ε the subset

$$C_\varepsilon = \{\mathbf{x} \in \mathbb{R}^n \| x_j| \ < \ \varepsilon \text{ for all } j = 1, \ldots, n\}.$$

1.3.2.6.2. $\mathbb{Z}^n$-periodic distributions in $\mathbb{R}^n$

A distribution $T \in \mathscr{D}'(\mathbb{R}^n)$ is called *periodic with period lattice* $\mathbb{Z}^n$ (or $\mathbb{Z}^n$-periodic) if $\tau_{\mathbf{m}} T = T$ for all $\mathbf{m} \in \mathbb{Z}^n$ (in crystallography the period lattice is the *direct* lattice).

Given a distribution with compact support $T^0 \in \mathscr{E}'(\mathbb{R}^n)$, then $T = \sum_{\mathbf{m}\in\mathbb{Z}^n} \tau_{\mathbf{m}} T^0$ is a $\mathbb{Z}^n$-periodic distribution. Note that we may write $T = r * T^0$, where $r = \sum_{\mathbf{m}\in\mathbb{Z}^n} \delta_{(\mathbf{m})}$ consists of Dirac δ's at all nodes of the period lattice $\mathbb{Z}^n$.

Conversely, any $\mathbb{Z}^n$-periodic distribution T may be written as $r * T^0$ for some $T^0 \in \mathscr{E}'$. To retrieve such a 'motif' T^0 from T, a function ψ will be constructed in such a way that $\psi \in \mathscr{D}$ (hence has compact support) and $r * \psi = 1$; then $T^0 = \psi T$. Indicator functions (Section 1.3.2.2) such as χ_1 or $\chi_{C_{1/2}}$ cannot be used directly, since they are discontinuous; but regularized versions of them may be constructed by convolution (see Section 1.3.2.3.9.7) as $\psi_0 = \chi_{C_\varepsilon} * \theta_\eta$, with ε and η such that $\psi_0(\mathbf{x}) = 1$ on $C_{1/2}$ and $\psi_0(\mathbf{x}) = 0$ outside $C_{3/4}$. Then the function

$$\psi = \frac{\psi_0}{\sum_{\mathbf{m}\in\mathbb{Z}^n} \tau_{\mathbf{m}} \psi_0}$$

has the desired property. The sum in the denominator contains at most 2^n non-zero terms at any given point $\mathbf{x}$ and acts as a smoothly varying 'multiplicity correction'.

1.3.2.6.3. *Identification with distributions over* $\mathbb{R}^n/\mathbb{Z}^n$

Throughout this section, 'periodic' will mean '$\mathbb{Z}^n$-periodic'.

Let $s \in \mathbb{R}$, and let $[s]$ denote the largest integer $\leq s$. For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, let $\tilde{\mathbf{x}}$ be the unique vector $(\tilde{x}_1, \ldots, \tilde{x}_n)$ with $\tilde{x}_j = x_j - [x_j]$. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then $\tilde{\mathbf{x}} = \tilde{\mathbf{y}}$ if and only if $\mathbf{x} - \mathbf{y} \in \mathbb{Z}^n$. The image of the map $\mathbf{x} \longmapsto \tilde{\mathbf{x}}$ is thus $\mathbb{R}^n$ modulo $\mathbb{Z}^n$, or $\mathbb{R}^n/\mathbb{Z}^n$.

If f is a periodic function over $\mathbb{R}^n$, then $\tilde{\mathbf{x}} = \tilde{\mathbf{y}}$ implies $f(\mathbf{x}) = f(\mathbf{y})$; we may thus define a function $\tilde{f}$ over $\mathbb{R}^n/\mathbb{Z}^n$ by putting $\tilde{f}(\tilde{\mathbf{x}}) = f(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} - \tilde{\mathbf{x}} \in \mathbb{Z}^n$. Conversely, if $\tilde{f}$ is a function over $\mathbb{R}^n/\mathbb{Z}^n$, then we may define a function f over $\mathbb{R}^n$ by putting $f(\mathbf{x}) = \tilde{f}(\tilde{\mathbf{x}})$, and f will be periodic. Periodic functions over $\mathbb{R}^n$ may thus be identified with functions over $\mathbb{R}^n/\mathbb{Z}^n$, and this identification preserves the notions of convergence, local summability and differentiability.

Given $\varphi^0 \in \mathscr{D}(\mathbb{R}^n)$, we may define

$$\varphi(\mathbf{x}) = \sum_{\mathbf{m}\in\mathbb{Z}^n} (\tau_{\mathbf{m}} \varphi^0)(\mathbf{x})$$

since the sum only contains finitely many non-zero terms; φ is periodic, and $\tilde{\varphi} \in \mathscr{D}(\mathbb{R}^n/\mathbb{Z}^n)$. Conversely, if $\tilde{\varphi} \in \mathscr{D}(\mathbb{R}^n/\mathbb{Z}^n)$ we may define $\varphi \in \mathscr{E}(\mathbb{R}^n)$ periodic by $\varphi(\mathbf{x}) = \tilde{\varphi}(\tilde{\mathbf{x}})$, and $\varphi^0 \in \mathscr{D}(\mathbb{R}^n)$ by putting $\varphi^0 = \psi\varphi$ with ψ constructed as above.

By transposition, a distribution $\tilde{T} \in \mathscr{D}'(\mathbb{R}^n/\mathbb{Z}^n)$ defines a unique periodic distribution $T \in \mathscr{D}'(\mathbb{R}^n)$ by $\langle T, \varphi^0 \rangle = \langle \tilde{T}, \tilde{\varphi} \rangle$; conversely, $T \in \mathscr{D}'(\mathbb{R}^n)$ periodic defines uniquely $\tilde{T} \in \mathscr{D}'(\mathbb{R}^n/\mathbb{Z}^n)$ by $\langle \tilde{T}, \tilde{\varphi} \rangle = \langle T, \varphi^0 \rangle$.

We may therefore identify $\mathbb{Z}^n$-periodic distributions over $\mathbb{R}^n$ with distributions over $\mathbb{R}^n/\mathbb{Z}^n$. We will, however, use mostly the former presentation, as it is more closely related to the crystallographer's perception of periodicity (see Section 1.3.4.1).

1.3.2.6.4. *Fourier transforms of periodic distributions*

The content of this section is perhaps the central result in the relation between Fourier theory and crystallography (Section 1.3.4.2.1.1).

Let $T = r * T^0$ with r defined as in Section 1.3.2.6.2. Then $r \in \mathscr{S}'$, $T^0 \in \mathscr{E}'$ hence $T^0 \in \mathscr{O}'_C$, so that $T \in \mathscr{S}'$: $\mathbb{Z}^n$-periodic distributions are tempered, hence have a Fourier transform. The convolution theorem (Section 1.3.2.5.8) is applicable, giving:

$$\mathscr{F}[T] = \mathscr{F}[r] \times \mathscr{F}[T^0]$$

and similarly for $\bar{\mathscr{F}}$.

Since $\mathscr{F}[\delta_{(\mathbf{m})}](\xi) = \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{m})$, formally

$$\mathscr{F}[r]_{\boldsymbol{\xi}} = \sum_{\mathbf{m}\in\mathbb{Z}^n} \exp(-2\pi i \boldsymbol{\xi} \cdot \mathbf{m}) = Q,$$

say.

It is readily shown that Q is tempered and periodic, so that $Q = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} \tau_{\boldsymbol{\mu}}(\psi Q)$, while the periodicity of r implies that

$$[\exp(-2\pi i \xi_j) - 1]\psi Q = 0, \quad j = 1, \ldots, n.$$

Since the first factors have single isolated zeros at $\xi_j = 0$ in $C_{3/4}$, $\psi Q = c\delta$ (see Section 1.3.2.3.9.4) and hence by periodicity $Q = cr$; convoluting with χ_{C_1} shows that $c = 1$. Thus we have the fundamental result:

$$\boxed{\mathscr{F}[r] = r}$$

so that

$$\mathscr{F}[T] = r \times \mathscr{F}[T^0];$$

i.e., according to Section 1.3.2.3.9.3,

$$\mathscr{F}[T]_{\boldsymbol{\xi}} = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} \mathscr{F}[T^0](\boldsymbol{\mu}) \times \delta_{(\boldsymbol{\mu})}.$$

The right-hand side is a *weighted* lattice distribution, whose nodes $\boldsymbol{\mu} \in \mathbb{Z}^n$ are weighted by the *sample values* $\mathscr{F}[T^0](\boldsymbol{\mu})$ of the transform of the motif T^0 at those nodes. Since $T^0 \in \mathscr{E}'$, the latter values may be written

$$\mathscr{F}[T^0](\boldsymbol{\mu}) = \langle T^0_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\mu} \cdot \mathbf{x}) \rangle.$$

By the structure theorem for distributions with compact support (Section 1.3.2.3.9.7), T^0 is a derivative of finite order of a continuous function; therefore, from Section 1.3.2.4.2.8 and Section 1.3.2.5.8, $\mathscr{F}[T^0](\boldsymbol{\mu})$ grows at most polynomially as $\|\boldsymbol{\mu}\| \to \infty$ (see also Section 1.3.2.6.10.3 about this property). Conversely, let $W = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} w_{\boldsymbol{\mu}} \delta_{(\boldsymbol{\mu})}$ be a weighted lattice distribution such that the weights $w_{\boldsymbol{\mu}}$ grow at most polynomially as $\|\boldsymbol{\mu}\| \to \infty$. Then W is a tempered distribution, whose Fourier cotransform $T_{\mathbf{x}} = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} w_{\boldsymbol{\mu}} \exp(+2\pi i \boldsymbol{\mu} \cdot \mathbf{x})$ is periodic. If T is now written as $r * T^0$ for some $T^0 \in \mathscr{E}'$, then by the reciprocity theorem

$$w_{\boldsymbol{\mu}} = \mathscr{F}[T^0](\boldsymbol{\mu}) = \langle T^0_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\mu} \cdot \mathbf{x}) \rangle.$$

Although the choice of T^0 is not unique, and need not yield back the same motif as may have been used to build T initially, different choices of T^0 will lead to the same coefficients $w_{\boldsymbol{\mu}}$ because of the periodicity of $\exp(-2\pi i \boldsymbol{\mu} \cdot \mathbf{x})$.

The Fourier transformation thus establishes a duality between periodic distributions and weighted lattice distributions. The pair of relations

(i) $$w_{\boldsymbol{\mu}} = \langle T^0_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\mu}\cdot\mathbf{x})\rangle$$

(ii) $$T_{\mathbf{x}} = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} w_{\boldsymbol{\mu}} \exp(+2\pi i \boldsymbol{\mu}\cdot\mathbf{x})$$

are referred to as the *Fourier analysis* and the *Fourier synthesis* of T, respectively (there is a discrepancy between this terminology and the crystallographic one, see Section 1.3.4.2.1.1). In other words, any periodic distribution $T \in \mathscr{S}'$ may be represented by a Fourier series (ii), whose coefficients are calculated by (i). The convergence of (ii) towards T in $\mathscr{S}'$ will be investigated later (Section 1.3.2.6.10).

1.3.2.6.5. *The case of non-standard period lattices*

Let Λ denote the non-standard lattice consisting of all vectors of the form $\sum_{j=1} m_j \mathbf{a}_j$, where the m_j are rational integers and $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are n linearly independent vectors in $\mathbb{R}^n$. Let R be the corresponding lattice distribution: $R = \sum_{x\in\Lambda}\delta_{(\mathbf{x})}$.

Let $\mathbf{A}$ be the non-singular $n\times n$ matrix whose successive columns are the coordinates of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ in the standard basis of $\mathbb{R}^n$; $\mathbf{A}$ will be called the *period matrix* of Λ, and the mapping $\mathbf{x} \longmapsto \mathbf{A}\mathbf{x}$ will be denoted by A. According to Section 1.3.2.3.9.5 we have

$$\langle R, \varphi\rangle = \sum_{\mathbf{m}\in\mathbb{Z}^n} \varphi(\mathbf{A}\mathbf{m}) = \langle r, (A^{-1})^{\#}\varphi\rangle = |\det \mathbf{A}|^{-1}\langle A^{\#} r, \varphi\rangle$$

for any $\varphi \in \mathscr{S}$, and hence $R = |\det \mathbf{A}|^{-1} A^{\#} r$. By Fourier transformation, according to Section 1.3.2.5.5,

$$\mathscr{F}[R] = |\det \mathbf{A}|^{-1}\mathscr{F}[A^{\#} r] = [(\mathbf{A}^{-1})^T]^{\#}\mathscr{F}[r] = [(\mathbf{A}^{-1})^T]^{\#} r,$$

which we write:

$$\mathscr{F}[R] = |\det \mathbf{A}|^{-1} R^*$$

with

$$R^* = |\det \mathbf{A}|[(\mathbf{A}^{-1})^T]^{\#} r.$$

R^* is a lattice distribution:

$$R^* = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} \delta_{[(\mathbf{A}^{-1})^T\boldsymbol{\mu}]} = \sum_{\boldsymbol{\xi}\in\Lambda^*} \delta_{(\boldsymbol{\xi})}$$

associated with the *reciprocal lattice* Λ^* whose basis vectors $\mathbf{a}_1^*, \ldots, \mathbf{a}_n^*$ are the columns of $(\mathbf{A}^{-1})^T$. Since the latter matrix is equal to the adjoint matrix (*i.e.* the matrix of co-factors) of $\mathbf{A}$ divided by $\det \mathbf{A}$, the components of the reciprocal basis vectors can be written down explicitly (see Section 1.3.4.2.1.1 for the crystallographic case $n = 3$).

A distribution T will be called Λ-*periodic* if $\tau_{\boldsymbol{\xi}} T = T$ for all $\boldsymbol{\xi} \in \Lambda$; as previously, T may be written $R * T^0$ for some motif distribution T^0 with compact support. By Fourier transformation,

$$\begin{aligned}\mathscr{F}[T] &= |\det \mathbf{A}|^{-1} R^* \cdot \mathscr{F}[T^0] \\ &= |\det \mathbf{A}|^{-1} \sum_{\boldsymbol{\xi}\in\Lambda^*} \mathscr{F}[T^0](\boldsymbol{\xi})\delta_{(\boldsymbol{\xi})} \\ &= |\det \mathbf{A}|^{-1} \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} \mathscr{F}[T^0][(\mathbf{A}^{-1})^T\boldsymbol{\mu}]\delta_{[(\mathbf{A}^{-1})^T\boldsymbol{\mu}]}\end{aligned}$$

so that $\mathscr{F}[T]$ is a weighted reciprocal-lattice distribution, the weight attached to node $\boldsymbol{\xi} \in \Lambda^*$ being $|\det \mathbf{A}|^{-1}$ times the value $\mathscr{F}[T^0](\boldsymbol{\xi})$ of the Fourier transform of the motif T^0.

This result may be further simplified if T and its motif T^0 are referred to the standard period lattice $\mathbb{Z}^n$ by defining t and t^0 so that $T = A^{\#} t$, $T^0 = A^{\#} t^0$, $t = r * t^0$. Then

$$\mathscr{F}[T^0](\boldsymbol{\xi}) = |\det \mathbf{A}|\mathscr{F}[t^0](\mathbf{A}^T\boldsymbol{\xi}),$$

hence

$$\mathscr{F}[T^0][(\mathbf{A}^{-1})^T\boldsymbol{\mu}] = |\det \mathbf{A}|\mathscr{F}[t^0](\boldsymbol{\mu}),$$

so that

$$\mathscr{F}[T] = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} \mathscr{F}[t^0](\boldsymbol{\mu})\delta_{[(\mathbf{A}^{-1})^T\boldsymbol{\mu}]}$$

in non-standard coordinates, while

$$\mathscr{F}[t] = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} \mathscr{F}[t^0](\boldsymbol{\mu})\delta_{(\boldsymbol{\mu})}$$

in standard coordinates.

The reciprocity theorem may then be written:

(iii) $$W_{\boldsymbol{\xi}} = |\det \mathbf{A}|^{-1}\langle T^0_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\xi}\cdot\mathbf{x})\rangle, \quad \boldsymbol{\xi} \in \boldsymbol{\Lambda}^*$$

(iv) $$T_{\mathbf{x}} = \sum_{\boldsymbol{\xi}\in\Lambda^*} W_{\boldsymbol{\xi}} \exp(+2\pi i \boldsymbol{\xi}\cdot\mathbf{x})$$

in non-standard coordinates, or equivalently:

(v) $$w_{\boldsymbol{\mu}} = \langle t^0_{\mathbf{x}}, \exp(-2\pi i \boldsymbol{\mu}\cdot\mathbf{x})\rangle, \quad \boldsymbol{\mu} \in \mathbb{Z}^n$$

(vi) $$t_{\mathbf{x}} = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} w_{\boldsymbol{\mu}} \exp(+2\pi i \boldsymbol{\mu}\cdot\mathbf{x})$$

in standard coordinates. It gives an n-dimensional Fourier series representation for any periodic distribution over $\mathbb{R}^n$. The convergence of such series in $\mathscr{S}'(\mathbb{R}^n)$ will be examined in Section 1.3.2.6.10.

1.3.2.6.6. *Duality between periodization and sampling*

Let T^0 be a distribution with compact support (the 'motif'). Its Fourier transform $\bar{\mathscr{F}}[T^0]$ is analytic (Section 1.3.2.5.4) and may thus be used as a multiplier.

We may rephrase the preceding results as follows:

(i) if T^0 is 'periodized by R' to give $R * T^0$, then $\bar{\mathscr{F}}[T^0]$ is 'sampled by R^*' to give $|\det \mathbf{A}|^{-1} R^* \cdot \bar{\mathscr{F}}[T^0]$;

(ii) if $\bar{\mathscr{F}}[T^0]$ is 'sampled by R^*' to give $R^* \cdot \bar{\mathscr{F}}[T^0]$, then T^0 is 'periodized by R' to give $|\det \mathbf{A}| R * T^0$.

Thus the Fourier transformation establishes a duality between the periodization of a distribution by a period lattice Λ and the sampling of its transform at the nodes of lattice Λ^* reciprocal to Λ. This is a particular instance of the convolution theorem of Section 1.3.2.5.8.

At this point it is traditional to break the symmetry between $\mathscr{F}$ and $\bar{\mathscr{F}}$ which distribution theory has enabled us to preserve even in the presence of periodicity, and to perform two distinct identifications:

(i) a Λ-periodic distribution T will be handled as a distribution $\tilde{T}$ on $\mathbb{R}^n/\Lambda$, was done in Section 1.3.2.6.3;

(ii) a weighted lattice distribution $W = \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n} W_{\boldsymbol{\mu}}\delta_{[(\mathbf{A}^{-1})^T\boldsymbol{\mu}]}$ will be identified with the collection $\{W_{\boldsymbol{\mu}} | \boldsymbol{\mu} \in \mathbb{Z}^n\}$ of its n-tuply indexed coefficients.

1.3.2.6.7. *The Poisson summation formula*

Let $\varphi \in \mathscr{S}$, so that $\mathscr{F}[\varphi] \in \mathscr{S}$. Let R be the lattice distribution associated to lattice Λ, with period matrix $\mathbf{A}$, and let R^* be associated to the reciprocal lattice Λ^*. Then we may write:

$$\begin{aligned}\langle R, \varphi\rangle &= \langle R, \bar{\mathscr{F}}[\mathscr{F}[\varphi]]\rangle \\ &= \langle \bar{\mathscr{F}}[R], \mathscr{F}[\varphi]\rangle \\ &= |\det \mathbf{A}|^{-1}\langle R^*, \mathscr{F}[\varphi]\rangle\end{aligned}$$

i.e.

$$\sum_{\mathbf{x}\in\Lambda}\varphi(\mathbf{x}) = |\det \mathbf{A}|^{-1}\sum_{\boldsymbol{\xi}\in\Lambda^*}\mathscr{F}[\varphi](\boldsymbol{\xi}).$$

This identity, which also holds for $\bar{\mathscr{F}}$, is called the *Poisson summation formula*. Its usefulness follows from the fact that the speed of decrease at infinity of φ and $\mathscr{F}[\varphi]$ are inversely related (Section 1.3.2.4.4.3), so that if one of the series (say, the left-hand side) is slowly convergent, the other (say, the right-hand side) will be rapidly convergent. This procedure has been used by Ewald (1921) [see also Bertaut (1952), Born & Huang (1954)] to evaluate lattice sums (Madelung constants) involved in the calculation of the internal electrostatic energy of crystals (see Chapter 3.4 in this volume on convergence acceleration techniques for crystallographic lattice sums).

When φ is a multivariate Gaussian

$$\varphi(\mathbf{x}) = G_{\mathbf{B}}(\mathbf{x}) = \exp(-\tfrac{1}{2}\mathbf{x}^T\mathbf{B}\mathbf{x}),$$

then

$$\mathscr{F}[\varphi](\boldsymbol{\xi}) = |\det (2\pi\mathbf{B}^{-1})|^{1/2}G_{\mathbf{B}^{-1}}(\boldsymbol{\xi}),$$

and Poisson's summation formula for a lattice with period matrix $\mathbf{A}$ reads:

$$\sum_{\mathbf{m}\in\mathbb{Z}^n}G_{\mathbf{B}}(\mathbf{A}\mathbf{m}) = |\det \mathbf{A}|^{-1}|\det (2\pi\mathbf{B}^{-1})|^{1/2} \times \sum_{\boldsymbol{\mu}\in\mathbb{Z}^n}G_{4\pi^2\mathbf{B}^{-1}}[(\mathbf{A}^{-1})^T\boldsymbol{\mu}]$$

or equivalently

$$\sum_{\mathbf{m}\in\mathbb{Z}^n}G_C(\mathbf{m}) = |\det (2\pi\mathbf{C}^{-1})|^{1/2}\sum_{\boldsymbol{\mu}\in\mathbb{Z}^n}G_{4\pi^2\mathbf{C}^{-1}}(\boldsymbol{\mu})$$

with $\mathbf{C} = \mathbf{A}^T\mathbf{B}\mathbf{A}$.

1.3.2.6.8. *Convolution of Fourier series*

Let $S = R * S^0$ and $T = R * T^0$ be two Λ-periodic distributions, the motifs S^0 and T^0 having compact support. The convolution $S * T$ does not exist, because S and T do not satisfy the support condition (Section 1.3.2.3.9.7). However, the three distributions R, S^0 and T^0 do satisfy the generalized support condition, so that their convolution is defined; then, by associativity and commutativity:

$$R * S^0 * T^0 = S * T^0 = S^0 * T.$$

By Fourier transformation and by the convolution theorem:

$$\begin{aligned}R^* \times \mathscr{F}[S^0 * T^0] &= (R^* \times \mathscr{F}[S^0]) \times \mathscr{F}[T^0] \\ &= \mathscr{F}[T^0] \times (R^* \times \mathscr{F}[S^0]).\end{aligned}$$

Let $\{U_{\boldsymbol{\xi}}\}_{\boldsymbol{\xi}\in\Lambda^*}$, $\{V_{\boldsymbol{\xi}}\}_{\boldsymbol{\xi}\in\Lambda^*}$ and $\{W_{\boldsymbol{\xi}}\}_{\boldsymbol{\xi}\in\Lambda^*}$ be the sets of Fourier coefficients associated to S, T and $S * T^0 (= S^0 * T)$, respectively. Identifying the coefficients of $\delta_{\boldsymbol{\xi}}$ for $\boldsymbol{\xi}\in\Lambda^*$ yields the forward version of the convolution theorem for Fourier series:

$$W_{\boldsymbol{\xi}} = |\det \mathbf{A}|U_{\boldsymbol{\xi}}V_{\boldsymbol{\xi}}.$$

The backward version of the theorem requires that T be infinitely differentiable. The distribution $S \times T$ is then well defined and its Fourier coefficients $\{Q_{\boldsymbol{\xi}}\}_{\boldsymbol{\xi}\in\Lambda^*}$ are given by

$$Q_{\boldsymbol{\xi}} = \sum_{\boldsymbol{\eta}\in\Lambda^*}U_{\boldsymbol{\eta}}V_{\boldsymbol{\xi}-\boldsymbol{\eta}}.$$

1.3.2.6.9. *Toeplitz forms, Szegö's theorem*

Toeplitz forms were first investigated by Toeplitz (1907, 1910, 1911*a*). They occur in connection with the 'trigonometric moment problem' (Shohat & Tamarkin, 1943; Akhiezer, 1965) and probability theory (Grenander, 1952) and play an important role in several direct approaches to the crystallographic phase problem [see Sections 1.3.4.2.1.10, 1.3.4.5.2.2(*e*)]. Many aspects of their theory and applications are presented in the book by Grenander & Szegö (1958).

1.3.2.6.9.1. *Toeplitz forms*

Let $f \in L^1(\mathbb{R}/\mathbb{Z})$ be real-valued, so that its Fourier coefficients satisfy the relations $c_{-m}(f) = \overline{c_m(f)}$. The Hermitian form in $n+1$ complex variables

$$T_n[f](\mathbf{u}) = \sum_{\mu=0}^{n}\sum_{\nu=0}^{n}\overline{u_\mu}c_{\mu-\nu}u_\nu$$

is called the nth *Toeplitz form* associated to f. It is a straightforward consequence of the convolution theorem and of Parseval's identity that $T_n[f]$ may be written:

$$T_n[f](\mathbf{u}) = \int_0^1\left|\sum_{\nu=0}^{n}u_\nu\exp(2\pi i\nu x)\right|^2 f(x)\,\mathrm{d}x.$$

1.3.2.6.9.2. *The Toeplitz–Carathéodory–Herglotz theorem*

It was shown independently by Toeplitz (1911*b*), Carathéodory (1911) and Herglotz (1911) that a function $f \in L^1$ is almost everywhere non-negative *if and only if* the Toeplitz forms $T_n[f]$ associated to f are positive semidefinite for all values of n.

This is equivalent to the infinite system of determinantal inequalities

$$D_n = \det\begin{pmatrix} c_0 & c_{-1} & \cdot & \cdot & c_{-n} \\ c_1 & c_0 & c_{-1} & \cdot & \cdot \\ \cdot & c_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & c_{-1} \\ c_n & \cdot & \cdot & c_1 & c_0 \end{pmatrix} \geq 0 \quad \text{for all } n.$$

The D_n are called *Toeplitz determinants*. Their application to the crystallographic phase problem is described in Section 1.3.4.2.1.10.

1.3.2.6.9.3. *Asymptotic distribution of eigenvalues of Toeplitz forms*

The eigenvalues of the Hermitian form $T_n[f]$ are defined as the $n+1$ real roots of the characteristic equation $\det\{T_n[f-\lambda]\} = 0$. They will be denoted by

$$\lambda_1^{(n)}, \lambda_2^{(n)}, \ldots, \lambda_{n+1}^{(n)}.$$

It is easily shown that if $m \leq f(x) \leq M$ for all x, then $m \leq \lambda_\nu^{(n)} \leq M$ for all n and all $\nu = 1, \ldots, n+1$. As $n \to \infty$ these bounds, and the distribution of the $\lambda^{(n)}$ within these bounds, can be made more precise by introducing two new notions.

(i) *Essential bounds*: define ess inf f as the largest m such that $f(x) \geq m$ except for values of x forming a set of measure 0; and define ess sup f similarly.

(ii) *Equal distribution*. For each n, consider two sets of $n+1$ real numbers:

$$a_1^{(n)}, a_2^{(n)}, \ldots, a_{n+1}^{(n)}, \quad \text{and} \quad b_1^{(n)}, b_2^{(n)}, \ldots, b_{n+1}^{(n)}.$$

Assume that for each ν and each n, $|a_\nu^{(n)}| < K$ and $|b_\nu^{(n)}| < K$ with K independent of ν and n. The sets $\{a_\nu^{(n)}\}$ and $\{b_\nu^{(n)}\}$ are said to be equally distributed in $[-K, +K]$ if, for any function F over $[-K, +K]$,

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{\nu=1}^{n+1}[F(a_\nu^{(n)}) - F(b_\nu^{(n)})] = 0.$$

We may now state an important theorem of Szegö (1915, 1920). Let $f \in L^1$, and put $m = \text{ess inf } f$, $M = \text{ess sup } f$. If m and M are finite, then for any continuous function $F(\lambda)$ defined in the interval $[m, M]$ we have

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{\nu=1}^{n+1} F(\lambda_\nu^{(n)}) = \int_0^1 F[f(x)]\, \mathrm{d}x.$$

In other words, the eigenvalues $\lambda_\nu^{(n)}$ of the T_n and the values $f[\nu/(n+2)]$ of f on a regular subdivision of]0, 1[are equally distributed.

Further investigations into the spectra of Toeplitz matrices may be found in papers by Hartman & Wintner (1950, 1954), Kac *et al.* (1953), Widom (1965), and in the notes by Hirschman & Hughes (1977).

1.3.2.6.9.4. *Consequences of Szegö's theorem*

(i) If the λ's are ordered in ascending order, then

$$\lim_{n\to\infty} \lambda_1^{(n)} = m = \text{ess inf } f, \quad \lim_{n\to\infty} \lambda_{n+1}^{(n)} = M = \text{ess sup } f.$$

Thus, when $f \geq 0$, the condition number $\lambda_{n+1}^{(n)}/\lambda_1^{(n)}$ of $T_n[f]$ tends towards the 'essential dynamic range' M/m of f.

(ii) Let $F(\lambda) = \lambda^s$ where s is a positive integer. Then

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{\nu=1}^{n+1} [\lambda_\nu^{(n)}]^s = \int_0^1 [f(x)]^s\, \mathrm{d}x.$$

(iii) Let $m > 0$, so that $\lambda_\nu^{(n)} > 0$, and let $D_n(f) = \det T_n(f)$. Then

$$D_n(f) = \prod_{\nu=1}^{n+1} \lambda_\nu^{(n)},$$

hence

$$\log D_n(f) = \sum_{\nu=1}^{n+1} \log \lambda_\nu^{(n)}.$$

Putting $F(\lambda) = \log \lambda$, it follows that

$$\lim_{n\to\infty} [D_n(f)]^{1/(n+1)} = \exp\left\{ \int_0^1 \log f(x)\, \mathrm{d}x \right\}.$$

Further terms in this limit were obtained by Szegö (1952) and interpreted in probabilistic terms by Kac (1954).

1.3.2.6.10. *Convergence of Fourier series*

The investigation of the convergence of Fourier series and of more general trigonometric series has been the subject of intense study for over 150 years [see *e.g.* Zygmund (1976)]. It has been a constant source of new mathematical ideas and theories, being directly responsible for the birth of such fields as set theory, topology and functional analysis.

This section will briefly survey those aspects of the classical results in dimension 1 which are relevant to the practical use of Fourier series in crystallography. The books by Zygmund (1959), Tolstov (1962) and Katznelson (1968) are standard references in the field, and Dym & McKean (1972) is recommended as a stimulant.

1.3.2.6.10.1. *Classical L^1 theory*

The space $L^1(\mathbb{R}/\mathbb{Z})$ consists of (equivalence classes of) complex-valued functions f on the circle which are summable, *i.e.* for which

$$\|f\|_1 \equiv \int_0^1 |f(x)|\, \mathrm{d}x < +\infty.$$

It is a convolution algebra: If f and g are in L^1, then $f * g$ is in L^1.

The mth Fourier coefficient $c_m(f)$ of f,

$$c_m(f) = \int_0^1 f(x) \exp(-2\pi i m x)\, \mathrm{d}x$$

is bounded: $|c_m(f)| \leq \|f\|_1$, and by the Riemann–Lebesgue lemma $c_m(f) \to 0$ as $m \to \infty$. By the convolution theorem, $c_m(f * g) = c_m(f) c_m(g)$.

The pth *partial sum* $S_p(f)$ of the Fourier series of f,

$$S_p(f)(x) = \sum_{|m|\leq p} c_m(f) \exp(2\pi i m x),$$

may be written, by virtue of the convolution theorem, as $S_p(f) = D_p * f$, where

$$D_p(x) = \sum_{|m|\leq p} \exp(2\pi i m x) = \frac{\sin[(2p+1)\pi x]}{\sin \pi x}$$

is the *Dirichlet kernel*. Because D_p comprises numerous slowly decaying oscillations, both positive and negative, $S_p(f)$ may not converge towards f in a strong sense as $p \to \infty$. Indeed, spectacular pathologies are known to exist where the partial sums, examined pointwise, diverge everywhere (Zygmund, 1959, Chapter VIII). When f is piecewise continuous, but presents isolated jumps, convergence near these jumps is marred by the *Gibbs phenomenon*: $S_p(f)$ always 'overshoots the mark' by about 9%, the area under the spurious peak tending to 0 as $p \to \infty$ but not its height [see Larmor (1934) for the history of this phenomenon].

By contrast, the *arithmetic mean* of the partial sums, also called the pth Cesàro sum,

$$C_p(f) = \frac{1}{p+1} [S_0(f) + \ldots + S_p(f)],$$

converges to f in the sense of the L^1 norm: $\|C_p(f) - f\|_1 \to 0$ as $p \to \infty$. If furthermore f is *continuous*, then the convergence is *uniform*, *i.e.* the error is bounded everywhere by a quantity which goes to 0 as $p \to \infty$. It may be shown that

$$C_p(f) = F_p * f,$$

where

$$F_p(x) = \sum_{|m|\leq p} \left(1 - \frac{|m|}{p+1}\right) \exp(2\pi i m x) = \frac{1}{p+1} \left[\frac{\sin(p+1)\pi x}{\sin \pi x}\right]^2$$

is the *Fejér kernel*. F_p has over D_p the advantage of being everywhere positive, so that the Cesàro sums $C_p(f)$ of a positive function f are always positive.

The de la Vallée Poussin kernel

$$V_p(x) = 2F_{2p+1}(x) - F_p(x)$$

has a trapezoidal distribution of coefficients and is such that $c_m(V_p) = 1$ if $|m| \leq p+1$; therefore $V_p * f$ is a trigonometric polynomial with the same Fourier coefficients as f over that range of values of m.

The Poisson kernel

$$P_r(x) = 1 + 2\sum_{m=1}^{\infty} r^m \cos 2\pi mx$$
$$= \frac{1 - r^2}{1 - 2r\cos 2\pi mx + r^2}$$

with $0 \leq r < 1$ gives rise to an Abel summation procedure [Tolstov (1962, p. 162); Whittaker & Watson (1927, p. 57)] since

$$(P_r * f)(x) = \sum_{m\in\mathbb{Z}} c_m(f) r^{|m|} \exp(2\pi imx).$$

Compared with the other kernels, P_r has the disadvantage of not being a trigonometric polynomial; however, P_r is the real part of the Cauchy kernel (Cartan, 1961; Ahlfors, 1966):

$$P_r(x) = \mathscr{R}e\left[\frac{1 + r\exp(2\pi ix)}{1 - r\exp(2\pi ix)}\right]$$

and hence provides a link between trigonometric series and analytic functions of a complex variable.

Other methods of summation involve forming a moving average of f by convolution with other sequences of functions $\alpha_p(\mathbf{x})$ besides D_p of F_p which 'tend towards δ' as $p \to \infty$. The convolution is performed by multiplying the Fourier coefficients of f by those of α_p, so that one forms the quantities

$$S'_p(f)(x) = \sum_{|m|\leq p} c_m(\alpha_p) c_m(f) \exp(2\pi imx).$$

For instance the 'sigma factors' of Lanczos (Lanczos, 1966, p. 65), defined by

$$\sigma_m = \frac{\sin[m\pi/p]}{m\pi/p},$$

lead to a summation procedure whose behaviour is intermediate between those using the Dirichlet and the Fejér kernels; it corresponds to forming a moving average of f by convolution with

$$\alpha_p = p\chi_{[-1/(2p),\,1/(2p)]} * D_p,$$

which is itself the convolution of a 'rectangular pulse' of width $1/p$ and of the Dirichlet kernel of order p.

A review of the summation problem in crystallography is given in Section 1.3.4.2.1.3.

1.3.2.6.10.2. *Classical L^2 theory*

The space $L^2(\mathbb{R}/\mathbb{Z})$ of (equivalence classes of) square-integrable complex-valued functions f on the circle is contained in $L^1(\mathbb{R}/\mathbb{Z})$, since by the Cauchy–Schwarz inequality

$$\|f\|_1^2 = \left(\int_0^1 |f(x)| \times 1\ \mathrm{d}x\right)^2$$
$$\leq \left(\int_0^1 |f(x)|^2\ \mathrm{d}x\right)\left(\int_0^1 1^2\ \mathrm{d}x\right) = \|f\|_2^2 \leq \infty.$$

Thus all the results derived for L^1 hold for L^2, a great simplification over the situation in $\mathbb{R}$ or $\mathbb{R}^n$ where neither L^1 nor L^2 was contained in the other.

However, more can be proved in L^2, because L^2 is a Hilbert space (Section 1.3.2.2.4) for the inner product

$$(f, g) = \int_0^1 \overline{f(x)} g(x)\ \mathrm{d}x,$$

and because the family of functions $\{\exp(2\pi imx)\}_{m\in\mathbb{Z}}$ constitutes an orthonormal Hilbert basis for L^2.

The sequence of Fourier coefficients $c_m(f)$ of $f \in L^2$ belongs to the space $\ell^2(\mathbb{Z})$ of square-summable sequences:

$$\sum_{m\in\mathbb{Z}} |c_m(f)|^2 < \infty.$$

Conversely, every element $c = (c_m)$ of ℓ^2 is the sequence of Fourier coefficients of a unique function in L^2. The inner product

$$(c, d) = \sum_{m\in\mathbb{Z}} \overline{c_m} d_m$$

makes ℓ^2 into a Hilbert space, and the map from L^2 to ℓ^2 established by the Fourier transformation is an isometry (Parseval/Plancherel):

$$\|f\|_{L^2} = \|c(f)\|_{\ell^2}$$

or equivalently:

$$(f, g) = (c(f), c(g)).$$

This is a useful property in applications, since (f, g) may be calculated either from f and g themselves, or from their Fourier coefficients $c(f)$ and $c(g)$ (see Section 1.3.4.4.6) for crystallographic applications).

By virtue of the orthogonality of the basis $\{\exp(2\pi imx)\}_{m\in\mathbb{Z}}$, the partial sum $S_p(f)$ is the best mean-square fit to f in the linear subspace of L^2 spanned by $\{\exp(2\pi imx)\}_{|m|\leq p}$, and hence (Bessel's inequality)

$$\sum_{|m|\leq p} |c_m(f)|^2 = \|f\|_2^2 - \sum_{|M|\geq p} |c_M(f)|^2 \leq \|f\|_2^2.$$

1.3.2.6.10.3. *The viewpoint of distribution theory*

The use of distributions enlarges considerably the range of behaviour which can be accommodated in a Fourier series, even in the case of general dimension n where classical theories meet with even more difficulties than in dimension 1.

Let $\{w_m\}_{m\in\mathbb{Z}}$ be a sequence of complex numbers with $|w_m|$ growing at most polynomially as $|m| \to \infty$, say $|w_m| \leq C|m|^K$. Then the sequence $\{w_m/(2\pi im)^{K+2}\}_{m\in\mathbb{Z}}$ is in ℓ^2 and even defines a *continuous* function $f \in L^2(\mathbb{R}/\mathbb{Z})$ and an associated tempered distribution $T_f \in \mathscr{D}'(\mathbb{R}/\mathbb{Z})$. Differentiation of T_f $(K+2)$ times then yields a tempered distribution whose Fourier transform leads to the original sequence of coefficients. Conversely, by the structure theorem for distributions with compact support (Section 1.3.2.3.9.7), the motif T^0 of a $\mathbb{Z}$-periodic distribution is a derivative of finite order of a continuous function; hence its Fourier coefficients will grow at most polynomially with $|m|$ as $|m| \to \infty$.

Thus distribution theory allows the manipulation of Fourier series whose coefficients exhibit polynomial growth as their order goes to infinity, while those derived from functions had to tend to 0 by virtue of the Riemann–Lebesgue lemma. The distribution-theoretic approach to Fourier series holds even in the case of general dimension n, where classical theories meet with even more difficulties (see Ash, 1976) than in dimension 1.

1.3.2.7. *The discrete Fourier transformation*

1.3.2.7.1. *Shannon's sampling theorem and interpolation formula*

Let $\varphi \in \mathscr{E}(\mathbb{R}^n)$ be such that $\Phi = \mathscr{F}[\varphi]$ has compact support K. Let φ be sampled at the nodes of a lattice Λ^*, yielding the lattice distribution $R^* \times \varphi$. The Fourier transform of this sampled version of φ is

$$\mathscr{F}[R^* \times \varphi] = |\det \mathbf{A}|(R * \Phi),$$

which is essentially Φ periodized by period lattice $\Lambda = (\Lambda^*)^*$, with period matrix $\mathbf{A}$.

Let us assume that Λ is such that the translates of K by different period vectors of Λ are disjoint. Then we may recover Φ from $R * \Phi$ by masking the contents of a 'unit cell' $\mathscr{V}$ of Λ (*i.e.* a fundamental domain for the action of Λ in $\mathbb{R}^n$) whose boundary does not meet K. If $\chi_{\mathscr{V}}$ is the indicator function of $\mathscr{V}$, then

$$\Phi = \chi_{\mathscr{V}} \times (R * \Phi).$$

Transforming both sides by $\bar{\mathscr{F}}$ yields

$$\varphi = \bar{\mathscr{F}}\left[\chi_{\mathscr{V}} \times \frac{1}{|\det \mathbf{A}|}\mathscr{F}[R^* \times \varphi]\right],$$

i.e.

$$\varphi = \left(\frac{1}{V}\bar{\mathscr{F}}[\chi_{\mathscr{V}}]\right) * (R^* \times \varphi)$$

since $|\det \mathbf{A}|$ is the volume V of $\mathscr{V}$.

This interpolation formula is traditionally credited to Shannon (1949), although it was discovered much earlier by Whittaker (1915). It shows that φ may be recovered from its sample values on Λ^* (*i.e.* from $R^* \times \varphi$) provided Λ^* is sufficiently fine that no overlap (or 'aliasing') occurs in the periodization of Φ by the dual lattice Λ. The interpolation kernel is the transform of the normalized indicator function of a unit cell of Λ containing the support K of Φ.

If K is contained in a sphere of radius $1/\Delta$ and if Λ and Λ^* are rectangular, the length of each basis vector of Λ must be greater than $2/\Delta$, and thus the sampling interval must be smaller than $\Delta/2$. This requirement constitutes the Shannon sampling criterion.

1.3.2.7.2. *Duality between subdivision and decimation of period lattices*

1.3.2.7.2.1. *Geometric description of sublattices*

Let $\Lambda_{\mathbf{A}}$ be a period lattice in $\mathbb{R}^n$ with matrix $\mathbf{A}$, and let $\Lambda^*_{\mathbf{A}}$ be the lattice reciprocal to $\Lambda_{\mathbf{A}}$, with period matrix $(A^{-1})^T$. Let $\Lambda_{\mathbf{B}}, \mathbf{B}, \Lambda^*_{\mathbf{B}}$ be defined similarly, and let us suppose that $\Lambda_{\mathbf{A}}$ *is a sublattice of* $\Lambda_{\mathbf{B}}$, *i.e.* that $\Lambda_{\mathbf{B}} \supset \Lambda_{\mathbf{A}}$ as a set.

The relation between $\Lambda_{\mathbf{A}}$ and $\Lambda_{\mathbf{B}}$ may be described in two different fashions: (i) multiplicatively, and (ii) additively.

(i) We may write $\mathbf{A} = \mathbf{BN}$ for some non-singular matrix $\mathbf{N}$ with integer entries. $\mathbf{N}$ may be viewed as the period matrix of the coarser lattice $\Lambda_{\mathbf{A}}$ with respect to the period basis of the finer lattice $\Lambda_{\mathbf{B}}$. It will be more convenient to write $\mathbf{A} = \mathbf{DB}$, where $\mathbf{D} = \mathbf{BNB}^{-1}$ is a rational matrix (with integer determinant since $\det \mathbf{D} = \det \mathbf{N}$) in terms of which the two lattices are related by

$$\Lambda_{\mathbf{A}} = \mathbf{D}\Lambda_{\mathbf{B}}.$$

(ii) Call two vectors in $\Lambda_{\mathbf{B}}$ *congruent modulo* $\Lambda_{\mathbf{A}}$ if their difference lies in $\Lambda_{\mathbf{A}}$. Denote the set of congruence classes (or 'cosets') by $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$, and the number of these classes by $[\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}]$. The 'coset decomposition'

$$\Lambda_{\mathbf{B}} = \bigcup_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} (\ell + \Lambda_{\mathbf{A}})$$

represents $\Lambda_{\mathbf{B}}$ as the disjoint union of $[\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}]$ translates of $\Lambda_{\mathbf{A}}$. $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ is a finite lattice with $[\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}]$ elements, called the *residual lattice* of $\Lambda_{\mathbf{B}}$ modulo $\Lambda_{\mathbf{A}}$.

The two descriptions are connected by the relation $[\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}] = \det \mathbf{D} = \det \mathbf{N}$, which follows from a volume calculation. We may also combine (i) and (ii) into

$$\text{(iii)} \qquad \Lambda_{\mathbf{B}} = \bigcup_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} (\ell + \mathbf{D}\Lambda_{\mathbf{B}})$$

which may be viewed as the n-dimensional equivalent of the Euclidean algorithm for integer division: ℓ is the 'remainder' of the division by $\Lambda_{\mathbf{A}}$ of a vector in $\Lambda_{\mathbf{B}}$, the quotient being the matrix $\mathbf{D}$.

1.3.2.7.2.2. *Sublattice relations for reciprocal lattices*

Let us now consider the two *reciprocal lattices* $\Lambda^*_{\mathbf{A}}$ and $\Lambda^*_{\mathbf{B}}$. Their period matrices $(\mathbf{A}^{-1})^T$ and $(\mathbf{B}^{-1})^T$ are related by: $(\mathbf{B}^{-1})^T = (\mathbf{A}^{-1})^T\mathbf{N}^T$, where $\mathbf{N}^T$ is an integer matrix; or equivalently by $(\mathbf{B}^{-1})^T = \mathbf{D}^T(\mathbf{A}^{-1})^T$. This shows that the roles are reversed in that $\Lambda^*_{\mathbf{B}}$ *is a sublattice of* $\Lambda^*_{\mathbf{A}}$, which we may write:

$$\text{(i)}^* \qquad \Lambda^*_{\mathbf{B}} = \mathbf{D}^T\Lambda^*_{\mathbf{A}}$$

$$\text{(ii)}^* \qquad \Lambda^*_{\mathbf{A}} = \bigcup_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} (\ell^* + \Lambda^*_{\mathbf{B}}).$$

The residual lattice $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$ is finite, with $[\Lambda^*_{\mathbf{A}} : \Lambda^*_{\mathbf{B}}] = \det \mathbf{D} = \det \mathbf{N} = [\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}]$, and we may again combine $(\text{i})^*$ and $(\text{ii})^*$ into

$$\text{(iii)}^* \qquad \Lambda^*_{\mathbf{A}} = \bigcup_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} (\ell^* + \mathbf{D}^T\Lambda^*_{\mathbf{A}}).$$

1.3.2.7.2.3. *Relation between lattice distributions*

The above relations between lattices may be rewritten in terms of the corresponding *lattice distributions* as follows:

$$\text{(i)} \qquad R_{\mathbf{A}} = \frac{1}{|\det \mathbf{D}|}\mathbf{D}^{\#}R^*_{\mathbf{B}}$$

$$\text{(ii)} \qquad R_{\mathbf{B}} = T_{\mathbf{B/A}} * R_{\mathbf{A}}$$

$$\text{(i)}^* \qquad R^*_{\mathbf{B}} = \frac{1}{|\det \mathbf{D}|}(\mathbf{D}^T)^{\#}R^*_{\mathbf{A}}$$

$$\text{(ii)}^* \qquad R^*_{\mathbf{A}} = T^*_{\mathbf{A/B}} * R^*_{\mathbf{B}}$$

where

$$T_{\mathbf{B/A}} = \sum_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} \delta_{(\ell)}$$

and

$$T^*_{\mathbf{A/B}} = \sum_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} \delta_{(\ell^*)}$$

are (finite) residual-lattice distributions. We may incorporate the factor $1/|\det \mathbf{D}|$ in (i) and $(\text{i})^*$ into these distributions and define

$$S_{\mathbf{B/A}} = \frac{1}{|\det \mathbf{D}|}T_{\mathbf{B/A}}, \quad S^*_{\mathbf{A/B}} = \frac{1}{|\det \mathbf{D}|}T^*_{\mathbf{A/B}}.$$

Since $|\det \mathbf{D}| = [\Lambda_{\mathbf{B}} : \Lambda_{\mathbf{A}}] = [\Lambda^*_{\mathbf{A}} : \Lambda^*_{\mathbf{B}}]$, convolution with $S_{\mathbf{B/A}}$ and $S^*_{\mathbf{A/B}}$ has the effect of *averaging* the translates of a distribution under the elements (or 'cosets') of the residual lattices $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ and $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$, respectively. This process will be called 'coset averaging'. Eliminating $R_{\mathbf{A}}$ and $R_{\mathbf{B}}$ between (i) and (ii), and $R^*_{\mathbf{A}}$ and $R^*_{\mathbf{B}}$ between $(\text{i})^*$ and $(\text{ii})^*$, we may write:

$$\text{(i}')\qquad R_{\mathbf{A}} = \mathbf{D}^{\#}(S_{\mathbf{B/A}} * R_{\mathbf{A}})$$

$$\text{(ii}')\qquad R_{\mathbf{B}} = S_{\mathbf{B/A}} * (\mathbf{D}^{\#}R_{\mathbf{B}})$$

$$\text{(i}')^*\qquad R^*_{\mathbf{B}} = (\mathbf{D}^T)^{\#}(S^*_{\mathbf{A/B}} * R^*_{\mathbf{B}})$$

$$\text{(ii}')^*\qquad R^*_{\mathbf{A}} = S^*_{\mathbf{A/B}} * [(\mathbf{D}^T)^{\#}R^*_{\mathbf{A}}].$$

These identities show that *period subdivision* by convolution with

$S_{\mathbf{B/A}}$ (respectively $S^*_{\mathbf{A/B}}$) on the one hand, and *period decimation* by 'dilation' by $\mathbf{D}^{\#}$ on the other hand, are mutually inverse operations on $R_{\mathbf{A}}$ and $R_{\mathbf{B}}$ (respectively $R^*_{\mathbf{A}}$ and $R^*_{\mathbf{B}}$).

1.3.2.7.2.4. *Relation between Fourier transforms*

Finally, let us consider the relations between the *Fourier transforms* of these lattice distributions. Recalling the basic relation of Section 1.3.2.6.5,

$$\begin{aligned}\mathscr{F}[R_{\mathbf{A}}] &= \frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{A}} \\ &= \frac{1}{|\det \mathbf{DB}|} T^*_{\mathbf{A/B}} * R^*_{\mathbf{B}} \qquad \text{by (ii)}^* \\ &= \left(\frac{1}{|\det \mathbf{D}|} T^*_{\mathbf{A/B}}\right) * \left(\frac{1}{|\det \mathbf{B}|} R^*_{\mathbf{B}}\right)\end{aligned}$$

i.e.

$$\text{(iv)} \qquad \mathscr{F}[R_{\mathbf{A}}] = S^*_{\mathbf{A/B}} * \mathscr{F}[R_{\mathbf{B}}]$$

and similarly:

$$\text{(v)} \qquad \mathscr{F}[R^*_{\mathbf{B}}] = S_{\mathbf{B/A}} * \mathscr{F}[R^*_{\mathbf{A}}].$$

Thus $R_{\mathbf{A}}$ (respectively $R^*_{\mathbf{B}}$), a *decimated* version of $R_{\mathbf{B}}$ (respectively $R^*_{\mathbf{A}}$), is transformed by $\mathscr{F}$ into a *subdivided* version of $\mathscr{F}[R_{\mathbf{B}}]$ (respectively $\mathscr{F}[R^*_{\mathbf{A}}]$).

The converse is also true:

$$\begin{aligned}\mathscr{F}[R_{\mathbf{B}}] &= \frac{1}{|\det \mathbf{B}|} R^*_{\mathbf{B}} \\ &= \frac{1}{|\det \mathbf{B}|}\frac{1}{|\det \mathbf{D}|} (\mathbf{D}^T)^{\#} R^*_{\mathbf{A}} \qquad \text{by (i)}^* \\ &= (\mathbf{D}^T)^{\#}\left(\frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{A}}\right)\end{aligned}$$

i.e.

$$\text{(iv}')\qquad \mathscr{F}[R_{\mathbf{B}}] = (\mathbf{D}^T)^{\#}\mathscr{F}[R_{\mathbf{A}}]$$

and similarly

$$\text{(v}')\qquad \mathscr{F}[R^*_{\mathbf{A}}] = \mathbf{D}^{\#}\mathscr{F}[R^*_{\mathbf{B}}].$$

Thus $R_{\mathbf{B}}$ (respectively $R^*_{\mathbf{A}}$), a *subdivided* version of $R_{\mathbf{A}}$ (respectively $R^*_{\mathbf{B}}$) is transformed by $\mathscr{F}$ into a *decimated* version of $\mathscr{F}[R_{\mathbf{A}}]$ (respectively $\mathscr{F}[R^*_{\mathbf{B}}]$). Therefore, *the Fourier transform exchanges subdivision and decimation of period lattices for lattice distributions*.

Further insight into this phenomenon is provided by applying $\bar{\mathscr{F}}$ to both sides of (iv) and (v) and invoking the convolution theorem:

$$\text{(iv}'')\qquad R_{\mathbf{A}} = \bar{\mathscr{F}}[S^*_{\mathbf{A/B}}] \times R_{\mathbf{B}}$$

$$\text{(v}'')\qquad R^*_{\mathbf{B}} = \bar{\mathscr{F}}[S_{\mathbf{B/A}}] \times R^*_{\mathbf{A}}.$$

These identities show that multiplication by the transform of the period-subdividing distribution $S^*_{\mathbf{A/B}}$ (respectively $S_{\mathbf{B/A}}$) has the effect of decimating $R_{\mathbf{B}}$ to $R_{\mathbf{A}}$ (respectively $R^*_{\mathbf{A}}$ to $R^*_{\mathbf{B}}$). They clearly imply that, if $\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ and $\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$, then

$$\begin{aligned}\bar{\mathscr{F}}[S^*_{\mathbf{A/B}}](\ell) &= 1 \text{ if } \ell = \mathbf{0} \quad (\textit{i.e.} \text{ if } \ell \text{ belongs} \\ &\qquad\qquad \text{to the class of } \Lambda_{\mathbf{A}}), \\ &= 0 \text{ if } \ell \neq \mathbf{0}; \\ \bar{\mathscr{F}}[S_{\mathbf{B/A}}](\ell^*) &= 1 \text{ if } \ell^* = \mathbf{0} \quad (\textit{i.e.} \text{ if } \ell^* \text{ belongs} \\ &\qquad\qquad \text{to the class of } \Lambda^*_{\mathbf{B}}), \\ &= 0 \text{ if } \ell^* \neq \mathbf{0}.\end{aligned}$$

Therefore, the duality between subdivision and decimation may be viewed as another aspect of that between convolution and multiplication.

There is clearly a strong analogy between the sampling/periodization duality of Section 1.3.2.6.6 and the decimation/subdivision duality, which is viewed most naturally in terms of subgroup relationships: both sampling and decimation involve restricting a function to a *discrete additive subgroup* of the domain over which it is initially given.

1.3.2.7.2.5. *Sublattice relations in terms of periodic distributions*

The usual presentation of this duality is not in terms of lattice distributions, but of periodic distributions obtained by convolving them with a motif.

Given $T^0 \in \mathscr{E}'(\mathbb{R}^n)$, let us form $R_{\mathbf{A}} * T^0$, then *decimate* its transform $(1/|\det \mathbf{A}|)R^*_{\mathbf{A}} \times \bar{\mathscr{F}}[T^0]$ by keeping only its values at the points of the coarser lattice $\Lambda^*_{\mathbf{B}} = \mathbf{D}^T\Lambda^*_{\mathbf{A}}$; as a result, $R^*_{\mathbf{A}}$ is replaced by $(1/|\det \mathbf{D}|)R^*_{\mathbf{B}}$, and the reverse transform then yields

$$\frac{1}{|\det \mathbf{D}|} R_{\mathbf{B}} * T^0 = S_{\mathbf{B/A}} * (R_{\mathbf{A}} * T^0) \qquad \text{by (ii)},$$

which is the *coset-averaged* version of the original $R_{\mathbf{A}} * T^0$. The converse situation is analogous to that of Shannon's sampling theorem. Let a function $\varphi \in \mathscr{E}(\mathbb{R}^n)$ whose transform $\Phi = \mathscr{F}[\varphi]$ has compact support be sampled as $R_{\mathbf{B}} \times \varphi$ at the nodes of $\Lambda_{\mathbf{B}}$. Then

$$\mathscr{F}[R_{\mathbf{B}} \times \varphi] = \frac{1}{|\det \mathbf{B}|}(R^*_{\mathbf{B}} * \Phi)$$

is periodic with period lattice $\Lambda^*_{\mathbf{B}}$. If the sampling lattice $\Lambda_{\mathbf{B}}$ is decimated to $\Lambda_{\mathbf{A}} = \mathbf{D}\Lambda_{\mathbf{B}}$, the inverse transform becomes

$$\begin{aligned}\mathscr{F}[R_{\mathbf{A}} \times \varphi] &= \frac{1}{|\det \mathbf{D}|}(R^*_{\mathbf{A}} * \Phi) \\ &= S^*_{\mathbf{A/B}} * (R^*_{\mathbf{B}} * \Phi) \qquad \text{by (ii)}^*,\end{aligned}$$

hence becomes periodized more finely by averaging over the cosets of $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$. With this finer periodization, the various copies of Supp Φ may start to overlap (a phenomenon called 'aliasing'), indicating that decimation has produced too coarse a sampling of φ.

1.3.2.7.3. *Discretization of the Fourier transformation*

Let $\varphi^0 \in \mathscr{E}(\mathbb{R}^n)$ be such that $\Phi^0 = \mathscr{F}[\varphi^0]$ has compact support (φ^0 is said to be *band-limited*). Then $\varphi = R_{\mathbf{A}} * \varphi^0$ is $\Lambda_{\mathbf{A}}$-periodic, and $\Phi = \mathscr{F}[\varphi] = (1/|\det \mathbf{A}|)R^*_{\mathbf{A}} \times \Phi^0$ is such that only a finite number of points $\lambda^*_{\mathbf{A}}$ of $\Lambda^*_{\mathbf{A}}$ have a non-zero Fourier coefficient $\Phi^0(\lambda^*_{\mathbf{A}})$ attached to them. We may therefore find a *decimation* $\Lambda^*_{\mathbf{B}} = \mathbf{D}^T\Lambda^*_{\mathbf{A}}$ of $\Lambda^*_{\mathbf{A}}$ such that the distinct translates of Supp Φ^0 by vectors of $\Lambda^*_{\mathbf{B}}$ do not intersect.

The distribution Φ can be uniquely recovered from $R^*_{\mathbf{B}} * \Phi$ by the procedure of Section 1.3.2.7.1, and we may write:

$$\begin{aligned}R^*_{\mathbf{B}} * \Phi &= \frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{B}} * (R^*_{\mathbf{A}} \times \Phi^0) \\ &= \frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{A}} \times (R^*_{\mathbf{B}} * \Phi^0) \\ &= \frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{B}} * [T^*_{\mathbf{A/B}} \times (R^*_{\mathbf{B}} * \Phi^0)];\end{aligned}$$

these rearrangements being legitimate because Φ^0 and $T^*_{\mathbf{A/B}}$ have compact supports which are intersection-free under the action of $\Lambda^*_{\mathbf{B}}$. By virtue of its $\Lambda^*_{\mathbf{B}}$-periodicity, this distribution is entirely characterized by its 'motif' $\tilde{\Phi}$ with respect to $\Lambda^*_{\mathbf{B}}$:

$$\tilde{\Phi} = \frac{1}{|\det \mathbf{A}|} T^*_{\mathbf{A/B}} \times (R^*_{\mathbf{B}} * \Phi^0).$$

Similarly, φ may be uniquely recovered by Shannon interpolation from the distribution sampling its values at the nodes of $\Lambda_{\mathbf{B}} = \mathbf{D}^{-1}\Lambda_{\mathbf{A}}$ ($\Lambda_{\mathbf{B}}$ is a *subdivision* of $\Lambda_{\mathbf{B}}$). By virtue of its $\Lambda_{\mathbf{A}}$-periodicity, this distribution is completely characterized by its motif:

$$\tilde{\varphi} = T_{\mathbf{B/A}} \times \varphi = T_{\mathbf{B/A}} \times (R^*_{\mathbf{A}} * \varphi^0).$$

Let $\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ and $\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$, and define the two sets of coefficients

$$\begin{aligned}
(1)\quad \tilde{\varphi}(\ell) &= \varphi(\ell + \boldsymbol{\lambda}_{\mathbf{A}}) && \text{for any } \boldsymbol{\lambda}_{\mathbf{A}} \in \Lambda_{\mathbf{A}} \\
&&& \text{(all choices of } \boldsymbol{\lambda}_{\mathbf{A}} \text{ give the same } \tilde{\varphi}), \\
(2)\quad \tilde{\Phi}(\ell^*) &= \Phi^0(\ell^* + \boldsymbol{\lambda}^*_{\mathbf{B}}) && \text{for the unique } \boldsymbol{\lambda}^*_{\mathbf{B}} \text{ (if it exists)} \\
&&& \text{such that } \ell^* + \boldsymbol{\lambda}^*_{\mathbf{B}} \in \operatorname{Supp} \Phi^0, \\
&= 0 && \text{if no such } \boldsymbol{\lambda}^*_{\mathbf{B}} \text{ exists.}
\end{aligned}$$

Define the two distributions

$$\omega = \sum_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} \tilde{\varphi}(\ell)\delta_{(\ell)}$$

and

$$\Omega = \sum_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} \tilde{\Phi}(\ell^*)\delta_{(\ell^*)}.$$

The relation between ω and Ω has two equivalent forms:

$$\text{(i)} \qquad R_{\mathbf{A}} * \omega = \mathscr{F}[R^*_{\mathbf{B}} * \Omega]$$

$$\text{(ii)} \qquad \bar{\mathscr{F}}[R_{\mathbf{A}} * \omega] = R^*_{\mathbf{B}} * \Omega.$$

By (i), $R_{\mathbf{A}} * \omega = |\det \mathbf{B}| R_{\mathbf{B}} \times \mathscr{F}[\Omega]$. Both sides are weighted lattice distributions concentrated at the nodes of $\Lambda_{\mathbf{B}}$, and equating the weights at $\boldsymbol{\lambda}_{\mathbf{B}} = \ell + \boldsymbol{\lambda}_{\mathbf{A}}$ gives

$$\tilde{\varphi}(\ell) = \frac{1}{|\det \mathbf{D}|} \sum_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} \tilde{\Phi}(\ell^*) \exp[-2\pi i \ell^* \cdot (\ell + \boldsymbol{\lambda}_{\mathbf{A}})].$$

Since $\ell^* \in \Lambda^*_{\mathbf{A}}$, $\ell^* \cdot \boldsymbol{\lambda}_{\mathbf{A}}$ is an integer, hence

$$\tilde{\varphi}(\ell) = \frac{1}{|\det \mathbf{D}|} \sum_{\ell^* \in \Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}} \tilde{\Phi}(\ell^*) \exp(-2\pi i \ell^* \cdot \ell).$$

By (ii), we have

$$\frac{1}{|\det \mathbf{A}|} R^*_{\mathbf{B}} * [T^*_{\mathbf{A/B}} \times (R^*_{\mathbf{B}} * \Phi^0)] = \frac{1}{|\det \mathbf{A}|} \bar{\mathscr{F}}[R_{\mathbf{A}} * \omega].$$

Both sides are weighted lattice distributions concentrated at the nodes of $\Lambda^*_{\mathbf{B}}$, and equating the weights at $\boldsymbol{\lambda}^*_{\mathbf{A}} = \ell^* + \boldsymbol{\lambda}^*_{\mathbf{B}}$ gives

$$\tilde{\Phi}(\ell^*) = \sum_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} \tilde{\varphi}(\ell) \exp[+2\pi i \ell \cdot (\ell^* + \boldsymbol{\lambda}^*_{\mathbf{B}})].$$

Since $\ell \in \Lambda_{\mathbf{B}}$, $\ell \cdot \boldsymbol{\lambda}^*_{\mathbf{B}}$ is an integer, hence

$$\tilde{\Phi}(\ell^*) = \sum_{\ell \in \Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}} \tilde{\varphi}(\ell) \exp(+2\pi i \ell \cdot \ell^*).$$

Now the decimation/subdivision relations between $\Lambda_{\mathbf{A}}$ and $\Lambda_{\mathbf{B}}$ may be written:

$$\mathbf{A} = \mathbf{DB} = \mathbf{BN},$$

so that

$$\begin{aligned}
\ell &= \mathbf{B}\mathscr{k} && \text{for } \mathscr{k} \in \mathbb{Z}^n \\
\ell^* &= (\mathbf{A}^{-1})^T \mathscr{k}^* && \text{for } \mathscr{k}^* \in \mathbb{Z}^n
\end{aligned}$$

with $(\mathbf{A}^{-1})^T = (\mathbf{B}^{-1})^T (\mathbf{N}^{-1})^T$, hence finally

$$\ell^* \cdot \ell = \ell \cdot \ell^* = \mathscr{k}^* \cdot (\mathbf{N}^{-1}\mathscr{k}).$$

Denoting $\tilde{\varphi}(\mathbf{B}\mathscr{k})$ by $\psi(\mathscr{k})$ and $\tilde{\Phi}[(\mathbf{A}^{-1})^T \mathscr{k}^*]$ by $\Psi(\mathscr{k}^*)$, the relation between ω and Ω may be written in the equivalent form

$$\text{(i)} \qquad \psi(\mathscr{k}) = \frac{1}{|\det \mathbf{N}|} \sum_{\mathscr{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n} \Psi(\mathscr{k}^*) \exp[-2\pi i \mathscr{k}^* \cdot (\mathbf{N}^{-1}\mathscr{k})]$$

$$\text{(ii)} \qquad \Psi(\mathscr{k}^*) = \sum_{\mathscr{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n} \psi(\mathscr{k}) \exp[+2\pi i \mathscr{k}^* \cdot (\mathbf{N}^{-1}\mathscr{k})],$$

where the summations are now over *finite* residual lattices in standard form.

Equations (i) and (ii) describe two mutually inverse linear transformations $\mathscr{F}(\mathbf{N})$ and $\bar{\mathscr{F}}(\mathbf{N})$ between two vector spaces $W_{\mathbf{N}}$ and $W^*_{\mathbf{N}}$ of dimension $|\det \mathbf{N}|$. $\mathscr{F}(\mathbf{N})$ [respectively $\bar{\mathscr{F}}(\mathbf{N})$] is the *discrete* Fourier (respectively inverse Fourier) transform associated to matrix $\mathbf{N}$.

The vector spaces $W_{\mathbf{N}}$ and $W^*_{\mathbf{N}}$ may be viewed from two different standpoints:

(1) as vector spaces of *weighted residual-lattice distributions*, of the form $\alpha(\mathbf{x})T_{\mathbf{B/A}}$ and $\beta(\mathbf{x})T^*_{\mathbf{A/B}}$; the canonical basis of $W_{\mathbf{N}}$ (respectively $W^*_{\mathbf{N}}$) then consists of the $\delta_{(\mathscr{k})}$ for $\mathscr{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ [respectively $\delta_{(\mathscr{k}^*)}$ for $\mathscr{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$];

(2) as vector spaces of *weight vectors* for the $|\det \mathbf{N}|$ δ-functions involved in the expression for $T_{\mathbf{B/A}}$ (respectively $T^*_{\mathbf{A/B}}$); the canonical basis of $W_{\mathbf{N}}$ (respectively $W^*_{\mathbf{N}}$) consists of weight vectors $\mathbf{u}_{\mathscr{k}}$ (respectively $\mathbf{v}_{\mathscr{k}^*}$) giving weight 1 to element $\mathscr{k}$ (respectively $\mathscr{k}^*$) and 0 to the others.

These two spaces are said to be 'isomorphic' (a relation denoted $\cong$), the isomorphism being given by the one-to-one correspondence:

$$\begin{aligned}
\omega = \sum_{\mathscr{k}} \psi(\mathscr{k})\delta_{(\mathscr{k})} \qquad &\leftrightarrow \qquad \psi = \sum_{\mathscr{k}} \psi(\mathscr{k})\mathbf{u}_{\mathscr{k}} \\
\Omega = \sum_{\mathscr{k}^*} \Psi(\mathscr{k}^*)\delta_{(\mathscr{k}^*)} \qquad &\leftrightarrow \qquad \Psi = \sum_{\mathscr{k}^*} \Psi(\mathscr{k}^*)\mathbf{v}_{\mathscr{k}^*}.
\end{aligned}$$

The second viewpoint will be adopted, as it involves only linear algebra. However, it is most helpful to keep the first one in mind and to think of the data or results of a discrete Fourier transform as representing (through their sets of unique weights) two periodic lattice distributions related by the full, distribution-theoretic Fourier transform.

We therefore view $W_{\mathbf{N}}$ (respectively $W^*_{\mathbf{N}}$) as the vector space of complex-valued functions over the finite residual lattice $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$ (respectively $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$) and write:

$$\begin{aligned}
W_{\mathbf{N}} &\cong L(\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}) \cong L(\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n) \\
W^*_{\mathbf{N}} &\cong L(\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}) \cong L(\mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n)
\end{aligned}$$

since a vector such as ψ is in fact the function $\mathscr{k} \longmapsto \psi(\mathscr{k})$.

The two spaces $W_{\mathbf{N}}$ and $W^*_{\mathbf{N}}$ may be equipped with the following Hermitian inner products:

$$\begin{aligned}
(\varphi, \psi)_W &= \sum_{\mathscr{k}} \overline{\varphi(\mathscr{k})}\psi(\mathscr{k}) \\
(\Phi, \Psi)_{W^*} &= \sum_{\mathscr{k}} \overline{\Phi(\mathscr{k}^*)}\Psi(\mathscr{k}^*),
\end{aligned}$$

which makes each of them into a *Hilbert space*. The canonical bases $\{\mathbf{u}_{\mathscr{k}} | \mathscr{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n\}$ and $\{\mathbf{v}_{\mathscr{k}^*} | \mathscr{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n\}$ and $W_{\mathbf{N}}$ and $W^*_{\mathbf{N}}$ are *orthonormal* for their respective product.

1.3.2.7.4. *Matrix representation of the discrete Fourier transform (DFT)*

By virtue of definitions (i) and (ii),

$$\mathscr{F}(\mathbf{N})\mathbf{v}_{\boldsymbol{k}^*} = \frac{1}{|\det \mathbf{N}|}\sum_{\boldsymbol{k}} \exp[-2\pi i \boldsymbol{k}^* \cdot (\mathbf{N}^{-1}\boldsymbol{k})]\mathbf{u}_{\boldsymbol{k}}$$
$$\bar{\mathscr{F}}(\mathbf{N})\mathbf{u}_{\boldsymbol{k}} = \sum_{\boldsymbol{k}^*} \exp[+2\pi i \boldsymbol{k}^* \cdot (\mathbf{N}^{-1}\boldsymbol{k})]\mathbf{v}_{\boldsymbol{k}^*}$$

so that $\mathscr{F}(\mathbf{N})$ and $\bar{\mathscr{F}}(\mathbf{N})$ may be represented, in the canonical bases of $W_{\mathbf{N}}$ and $W^*_{\mathbf{N}}$, by the following matrices:

$$[\mathscr{F}(\mathbf{N})]_{\boldsymbol{k}\boldsymbol{k}^*} = \frac{1}{|\det \mathbf{N}|}\exp[-2\pi i \boldsymbol{k}^* \cdot (\mathbf{N}^{-1}\boldsymbol{k})]$$
$$[\bar{\mathscr{F}}(\mathbf{N})]_{\boldsymbol{k}^*\boldsymbol{k}} = \exp[+2\pi i \boldsymbol{k}^* \cdot (\mathbf{N}^{-1}\boldsymbol{k})].$$

When $\mathbf{N}$ is symmetric, $\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ and $\mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$ may be identified in a natural manner, and the above matrices are symmetric.

When $\mathbf{N}$ is diagonal, say $\mathbf{N} = \mathrm{diag}(\nu_1, \nu_2, \ldots, \nu_n)$, then the tensor product structure of the full multidimensional Fourier transform (Section 1.3.2.4.2.4)

$$\mathscr{F}_{\mathbf{x}} = \mathscr{F}_{x_1} \otimes \mathscr{F}_{x_2} \otimes \ldots \otimes \mathscr{F}_{x_n}$$

gives rise to a tensor product structure for the DFT matrices. The tensor product of matrices is defined as follows:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \ldots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{n1}\mathbf{B} & \ldots & a_{nn}\mathbf{B} \end{pmatrix}.$$

Let the index vectors $\boldsymbol{k}$ and $\boldsymbol{k}^*$ be ordered in the same way as the elements in a Fortran array, *e.g.* for $\boldsymbol{k}$ with k_1 increasing fastest, k_2 next fastest, $\ldots, k_n$ slowest; then

$$\mathscr{F}(\mathbf{N}) = \mathscr{F}(\nu_1) \otimes \mathscr{F}(\nu_2) \otimes \ldots \otimes \mathscr{F}(\nu_n),$$

where

$$[\mathscr{F}(\nu_j)]_{k_j, k_j^*} = \frac{1}{\nu_j}\exp\left(-2\pi i \frac{k_j^* k_j}{\nu_j}\right),$$

and

$$\bar{\mathscr{F}}(\mathbf{N}) = \bar{\mathscr{F}}(\nu_1) \otimes \bar{\mathscr{F}}(\nu_2) \otimes \ldots \otimes \bar{\mathscr{F}}(\nu_n),$$

where

$$[\bar{\mathscr{F}}_{\nu_j}]_{k_j^*, k_j} = \exp\left(+2\pi i \frac{k_j^* k_j}{\nu_j}\right).$$

1.3.2.7.5. *Properties of the discrete Fourier transform*

The DFT inherits most of the properties of the Fourier transforms, but with certain numerical factors ('Jacobians') due to the transition from continuous to discrete measure.

(1) *Linearity* is obvious.

(2) *Shift property*. If $(\tau_a \psi)(\boldsymbol{k}) = \psi(\boldsymbol{k} - \boldsymbol{a})$ and $(\tau_{a^*}\Psi)(\boldsymbol{k}^*) = \Psi(\boldsymbol{k}^* - \boldsymbol{a}^*)$, where subtraction takes place by modular vector arithmetic in $\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ and $\mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$, respectively, then the following identities hold:

$$\bar{\mathscr{F}}(\mathbf{N})[\tau_{\boldsymbol{k}}\psi](\boldsymbol{k}^*) = \exp[+2\pi i \boldsymbol{k}^* \cdot (\mathbf{N}^{-1}\boldsymbol{k})]\bar{\mathscr{F}}(\mathbf{N})[\psi](\boldsymbol{k}^*)$$
$$\mathscr{F}(\mathbf{N})[\tau_{\boldsymbol{k}^*}\Psi](\boldsymbol{k}) = \exp[-2\pi i \boldsymbol{k}^* \cdot (\mathbf{N}^{-1}\boldsymbol{k})]\mathscr{F}(\mathbf{N})[\Psi](\boldsymbol{k}).$$

(3) *Differentiation identities*. Let vectors $\boldsymbol{\psi}$ and $\boldsymbol{\Psi}$ be constructed from $\varphi^0 \in \mathscr{E}(\mathbb{R}^n)$ as in Section 1.3.2.7.3, hence be related by the DFT. If $D^{\mathbf{p}}\boldsymbol{\psi}$ designates the vector of sample values of $D^{\mathbf{p}}_{\mathbf{x}}\varphi^0$ at the points of $\Lambda_{\mathbf{B}}/\Lambda_{\mathbf{A}}$, and $D^{\mathbf{p}}\boldsymbol{\Psi}$ the vector of values of $D^{\mathbf{p}}_{\boldsymbol{\xi}}\Phi^0$ at points of $\Lambda^*_{\mathbf{A}}/\Lambda^*_{\mathbf{B}}$, then for all multi-indices $\mathbf{p} = (p_1, p_2, \ldots, p_n)$

$$(D^{\mathbf{p}}\boldsymbol{\psi})(\boldsymbol{k}) = \bar{\mathscr{F}}(\mathbf{N})[(+2\pi i \boldsymbol{k}^*)^{\mathbf{p}}\boldsymbol{\Psi}](\boldsymbol{k})$$
$$(D^{\mathbf{p}}\boldsymbol{\Psi})(\boldsymbol{k}^*) = \mathscr{F}(\mathbf{N})[(-2\pi i \boldsymbol{k})^{\mathbf{p}}\boldsymbol{\psi}](\boldsymbol{k}^*)$$

or equivalently

$$\mathscr{F}(\mathbf{N})[D^{\mathbf{p}}\boldsymbol{\psi}](\boldsymbol{k}^*) = (+2\pi i \boldsymbol{k}^*)^{\mathbf{p}}\boldsymbol{\Psi}(\boldsymbol{k}^*)$$
$$\bar{\mathscr{F}}(\mathbf{N})[D^{\mathbf{p}}\boldsymbol{\Psi}](\boldsymbol{k}) = (-2\pi i \boldsymbol{k})^{\mathbf{p}}\boldsymbol{\psi}(\boldsymbol{k}).$$

(4) *Convolution property*. Let $\boldsymbol{\varphi} \in W_{\mathbf{N}}$ and $\boldsymbol{\Phi} \in W^*_{\mathbf{N}}$ (respectively $\boldsymbol{\psi}$ and $\boldsymbol{\Psi}$) be related by the DFT, and define

$$(\boldsymbol{\varphi} * \boldsymbol{\psi})(\boldsymbol{k}) = \sum_{\boldsymbol{k}' \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n} \boldsymbol{\varphi}(\boldsymbol{k}')\boldsymbol{\psi}(\boldsymbol{k} - \boldsymbol{k}')$$
$$(\boldsymbol{\Phi} * \boldsymbol{\Psi})(\boldsymbol{k}^*) = \sum_{\boldsymbol{k}^{*\prime} \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n} \boldsymbol{\Phi}(\boldsymbol{k}^{*\prime})\boldsymbol{\Psi}(\boldsymbol{k}^* - \boldsymbol{k}^{*\prime}).$$

Then

$$\bar{\mathscr{F}}(\mathbf{N})[\boldsymbol{\Phi} * \boldsymbol{\Psi}](\boldsymbol{k}) = |\det \mathbf{N}|\boldsymbol{\varphi}(\boldsymbol{k})\boldsymbol{\psi}(\boldsymbol{k})$$
$$\mathscr{F}(\mathbf{N})[\boldsymbol{\varphi} * \boldsymbol{\psi}](\boldsymbol{k}^*) = \boldsymbol{\Phi}(\boldsymbol{k}^*)\boldsymbol{\Psi}(\boldsymbol{k}^*)$$

and

$$\bar{\mathscr{F}}(\mathbf{N})[\boldsymbol{\varphi} \times \boldsymbol{\psi}](\boldsymbol{k}^*) = \frac{1}{|\det \mathbf{N}|}(\boldsymbol{\Phi} * \boldsymbol{\Psi})(\boldsymbol{k}^*)$$
$$\mathscr{F}(\mathbf{N})[\boldsymbol{\Phi} \times \boldsymbol{\Psi}](\boldsymbol{k}) = (\boldsymbol{\varphi} * \boldsymbol{\psi})(\boldsymbol{k}).$$

Since addition on $\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ and $\mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$ is modular, this type of convolution is called *cyclic* convolution.

(5) *Parseval/Plancherel property*. If $\boldsymbol{\varphi}$, $\boldsymbol{\psi}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ are as above, then

$$(\mathscr{F}(\mathbf{N})[\boldsymbol{\Phi}], \mathscr{F}(\mathbf{N})[\boldsymbol{\Psi}])_W = \frac{1}{|\det \mathbf{N}|}(\boldsymbol{\Phi}, \boldsymbol{\Psi})_{W^*}$$
$$(\bar{\mathscr{F}}(\mathbf{N})[\boldsymbol{\varphi}], \bar{\mathscr{F}}(\mathbf{N})[\boldsymbol{\psi}])_W = \frac{1}{|\det \mathbf{N}|}(\boldsymbol{\varphi}, \boldsymbol{\psi})_W.$$

(6) *Period* 4. When $\mathbf{N}$ is symmetric, so that the ranges of indices $\boldsymbol{k}$ and $\boldsymbol{k}^*$ can be identified, it makes sense to speak of powers of $\mathscr{F}(\mathbf{N})$ and $\bar{\mathscr{F}}(\mathbf{N})$. Then the 'standardized' matrices $(1/|\det \mathbf{N}|^{1/2})\mathscr{F}(\mathbf{N})$ and $(1/|\det \mathbf{N}|^{1/2})\bar{\mathscr{F}}(\mathbf{N})$ are *unitary* matrices whose fourth power is the identity matrix (Section 1.3.2.4.3.4); their eigenvalues are therefore ± 1 and $\pm i$.

1.3.3. Numerical computation of the discrete Fourier transform

1.3.3.1. *Introduction*

The Fourier transformation's most remarkable property is undoubtedly that of turning convolution into multiplication. As distribution theory has shown, other valuable properties – such as the shift property, the conversion of differentiation into multiplication by monomials, and the duality between periodicity and sampling – are special instances of the convolution theorem.

This property is exploited in many areas of applied mathematics and engineering (Campbell & Foster, 1948; Sneddon, 1951; Champeney, 1973; Bracewell, 1986). For example, the passing of a signal through a linear filter, which results in its being convolved with the response of the filter to a δ-function 'impulse', may be modelled as a multiplication of the signal's transform by the transform of the impulse response (also called transfer function). Similarly, the solution of systems of partial differential equations may be turned by Fourier transformation into a division problem for distributions. In both cases, the formulations obtained after Fourier transformation are considerably simpler than the initial ones, and lend themselves to constructive solution techniques.

Whenever the functions to which the Fourier transform is applied are band-limited, or can be well approximated by band-limited functions, the discrete Fourier transform (DFT) provides a means of constructing explicit numerical solutions to the problems at hand. A great variety of investigations in physics, engineering and applied mathematics thus lead to DFT calculations, to such a degree that, at the time of writing, about 50% of all supercomputer CPU time is alleged to be spent calculating DFTs.

The straightforward use of the defining formulae for the DFT leads to calculations of size N^2 for N sample points, which become unfeasible for any but the smallest problems. Much ingenuity has therefore been exerted on the design and implementation of faster algorithms for calculating the DFT (McClellan & Rader, 1979; Nussbaumer, 1981; Blahut, 1985; Brigham, 1988). The most famous is that of Cooley & Tukey (1965) which heralded the age of digital signal processing. However, it had been preceded by the prime factor algorithm of Good (1958, 1960), which has lately been the basis of many new developments. Recent historical research (Goldstine, 1977, pp. 249–253; Heideman *et al.*, 1984) has shown that Gauss essentially knew the Cooley–Tukey algorithm as early as 1805 (before Fourier's 1807 work on harmonic analysis!); while it has long been clear that Dirichlet knew of the basis of the prime factor algorithm and used it extensively in his theory of multiplicative characters [see *e.g.* Chapter I of Ayoub (1963), and Chapters 6 and 8 of Apostol (1976)]. Thus the computation of the DFT, far from being a purely technical and rather narrow piece of specialized numerical analysis, turns out to have very rich connections with such central areas of pure mathematics as number theory (algebraic and analytic), the representation theory of certain Lie groups and coding theory – to list only a few. The interested reader may consult Auslander & Tolimieri (1979); Auslander, Feig & Winograd (1982, 1984); Auslander & Tolimieri (1985); Tolimieri (1985).

One-dimensional algorithms are examined first. The Sande mixed-radix version of the Cooley–Tukey algorithm only calls upon the additive structure of congruence classes of integers. The prime factor algorithm of Good begins to exploit some of their multiplicative structure, and the use of relatively prime factors leads to a stronger factorization than that of Sande. Fuller use of the multiplicative structure, *via* the group of units, leads to the Rader algorithm; and the factorization of short convolutions then yields the Winograd algorithms.

Multidimensional algorithms are at first built as tensor products of one-dimensional elements. The problem of factoring the DFT in several dimensions simultaneously is then examined. The section ends with a survey of attempts at formalizing the interplay between algorithm structure and computer architecture for the purpose of automating the design of optimal DFT code.

It was originally intended to incorporate into this section a survey of all the basic notions and results of abstract algebra which are called upon in the course of these developments, but time limitations have made this impossible. This material, however, is adequately covered by the first chapter of Tolimieri *et al.* (1989) in a form tailored for the same purposes. Similarly, the inclusion of numerous detailed examples of the algorithms described here has had to be postponed to a later edition, but an abundant supply of such examples may be found in the signal processing literature, for instance in the books by McClellan & Rader (1979), Blahut (1985), and Tolimieri *et al.* (1989).

1.3.3.2. *One-dimensional algorithms*

Throughout this section we will denote by $e(t)$ the expression $\exp(2\pi i t)$, $t \in \mathbb{R}$. The mapping $t \longmapsto e(t)$ has the following properties:

$$\begin{aligned} e(t_1 + t_2) &= e(t_1)e(t_2) \\ e(-t) &= \overline{e(t)} = [e(t)]^{-1} \\ e(t) &= 1 \Leftrightarrow t \in \mathbb{Z}. \end{aligned}$$

Thus e defines an isomorphism between the additive group $\mathbb{R}/\mathbb{Z}$ (the reals modulo the integers) and the multiplicative group of complex numbers of modulus 1. It follows that the mapping $\ell \longmapsto e(\ell/N)$, where $\ell \in \mathbb{Z}$ and N is a positive integer, defines an isomorphism between the one-dimensional residual lattice $\mathbb{Z}/N\mathbb{Z}$ and the multiplicative group of Nth roots of unity.

The DFT on N points then relates vectors $\mathbf{X}$ and $\mathbf{X}^*$ in W and W^* through the linear transformations:

$$\begin{aligned} F(N): \quad & X(k) = \frac{1}{N} \sum_{k^* \in \mathbb{Z}/N\mathbb{Z}} X^*(k^*) e(-k^* k/N) \\ \bar{F}(N): \quad & X^*(k^*) = \sum_{k \in \mathbb{Z}/N\mathbb{Z}} X(k) e(k^* k/N). \end{aligned}$$

1.3.3.2.1. *The Cooley–Tukey algorithm*

The presentation of Gentleman & Sande (1966) will be followed first [see also Cochran *et al.* (1967)]. It will then be reinterpreted in geometric terms which will prepare the way for the treatment of multidimensional transforms in Section 1.3.3.3.

Suppose that the number of sample points N is composite, say $N = N_1 N_2$. We may write k to the base N_1 and k^* to the base N_2 as follows:

$$\begin{aligned} k &= k_1 + N_1 k_2 \quad k_1 \in \mathbb{Z}/N_1\mathbb{Z}, \quad k_2 \in \mathbb{Z}/N_2\mathbb{Z} \\ k^* &= k_2^* + k_1^* N_2 \quad k_1^* \in \mathbb{Z}/N_1\mathbb{Z}, \quad k_2^* \in \mathbb{Z}/N_2\mathbb{Z}. \end{aligned}$$

The defining relation for $\bar{F}(N)$ may then be written:

$$\begin{aligned} X^*(k_2^* + k_1^* N_2) = \sum_{k_1 \in \mathbb{Z}/N_1\mathbb{Z}} \sum_{k_2 \in \mathbb{Z}/N_2\mathbb{Z}} & X(k_1 + N_1 k_2) \\ & \times e\left[\frac{(k_2^* + k_1^* N_2)(k_1 + N_1 k_2)}{N_1 N_2}\right]. \end{aligned}$$

The argument of $e[.]$ may be expanded as

$$\frac{k_2^* k_1}{N} + \frac{k_1^* k_1}{N_1} + \frac{k_2^* k_2}{N_2} + k_1^* k_2,$$

and the last summand, being an integer, may be dropped:

$$\begin{aligned} & X^*(k_2^* + k_1^* N_2) \\ & \quad = \sum_{k_1} \left\{ e\left(\frac{k_2^* k_1}{N}\right) \left[\sum_{k_2} X(k_1 + N_1 k_2) e\left(\frac{k_2^* k_2}{N_2}\right) \right] \right\} \\ & \qquad \times e\left(\frac{k_1^* k_1}{N_1}\right). \end{aligned}$$

This computation may be decomposed into five stages, as follows:

(i) form the N_1 vectors $\mathbf{Y}_{k_1}$ of length N_2 by the prescription

$$Y_{k_1}(k_2) = X(k_1 + N_1 k_2), \quad k_1 \in \mathbb{Z}/N_1\mathbb{Z}, \quad k_2 \in \mathbb{Z}/N_2\mathbb{Z};$$

(ii) calculate the N_1 transforms $\mathbf{Y}_{k_1}^*$ on N_2 points:

$$\mathbf{Y}_{k_1}^* = \bar{F}(N_2)[\mathbf{Y}_{k_1}], \quad k_1 \in \mathbb{Z}/N_1\mathbb{Z};$$

(iii) form the N_2 vectors $\mathbf{Z}_{k_2^*}$ of length N_1 by the prescription

$$\mathbf{Z}_{k_2^*}(k_1) = e\left(\frac{k_2^* k_1}{N}\right) Y_{k_1}^*(k_2^*), \quad k_1 \in \mathbb{Z}/N_1\mathbb{Z}, \quad k_2^* \in \mathbb{Z}/N_2\mathbb{Z};$$

(iv) calculate the N_2 transforms $\mathbf{Z}^*_{k_2^*}$ on N_1 points:

$$\mathbf{Z}^*_{k_2^*} = \bar{F}(N_1)[\mathbf{Z}_{k_2^*}], \quad k_2^* \in \mathbb{Z}/N_2\mathbb{Z};$$

(v) collect $X^*(k_2^* + k_1^* N_2)$ as $Z^*_{k_2^*}(k_1^*)$.

If the intermediate transforms in stages (ii) and (iv) are performed *in place*, *i.e.* with the results overwriting the data, then at stage (v) the result $X^*(k_2^* + k_1^* N_2)$ will be found at address $k_1^* + N_1 k_2^*$. This phenomenon is called *scrambling* by 'digit reversal', and stage (v) is accordingly known as *unscrambling*.

The initial *N-point* transform $\bar{F}(N)$ has thus been performed as N_1 transforms $\bar{F}(N_2)$ on N_2 points, followed by N_2 transforms $\bar{F}(N_1)$ on N_1 points, thereby reducing the arithmetic cost from $(N_1 N_2)^2$ to $N_1 N_2 (N_1 + N_2)$. The phase shifts applied at stage (iii) are traditionally called 'twiddle factors', and the transposition between k_1 and k_2^* can be performed by the fast recursive technique of Eklundh (1972). Clearly, this procedure can be applied recursively if N_1 and N_2 are themselves composite, leading to an overall arithmetic cost of order $N \log N$ if N has no large prime factors.

The Cooley–Tukey factorization may also be derived from a geometric rather than arithmetic argument. The decomposition $k = k_1 + N_1 k_2$ is associated to a geometric partition of the residual lattice $\mathbb{Z}/N\mathbb{Z}$ into N_1 copies of $\mathbb{Z}/N_2\mathbb{Z}$, each translated by $k_1 \in \mathbb{Z}/N_1\mathbb{Z}$ and 'blown up' by a factor N_1. This partition in turn induces a (direct sum) decomposition of $\mathbf{X}$ as

$$\mathbf{X} = \sum_{k_1} \mathbf{X}_{k_1},$$

where

$$\begin{aligned} X_{k_1}(k) &= X(k) \quad \text{if } k \equiv k_1 \bmod N_1, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

According to (i), $\mathbf{X}_{k_1}$ is related to $\mathbf{Y}_{k_1}$ by *decimation by* N_1 and *offset by* k_1. By Section 1.3.2.7.2, $\bar{F}(N)[\mathbf{X}_{k_1}]$ is related to $\bar{F}(N_2)[\mathbf{Y}_{k_1}]$ by *periodization by* N_2 and *phase shift by* $e(k^* k_1/N)$, so that

$$X^*(k^*) = \sum_{k_1} e\left(\frac{k^* k_1}{N}\right) Y^*_{k_1}(k_2^*),$$

the periodization by N_2 being reflected by the fact that $Y^*_{k_1}$ does not depend on k_1^*. Writing $k^* = k_2^* + k_1^* N_2$ and expanding $k^* k_1$ shows that the phase shift contains both the twiddle factor $e(k_2^* k_1/N)$ and the kernel $e(k_1^* k_1/N_1)$ of $\bar{F}(N_1)$. The Cooley–Tukey algorithm is thus naturally associated to the coset decomposition of a lattice modulo a sublattice (Section 1.3.2.7.2).

It is readily seen that essentially the same factorization can be obtained for $F(N)$, up to the complex conjugation of the twiddle factors. The normalizing constant $1/N$ arises from the normalizing constants $1/N_1$ and $1/N_2$ in $F(N_1)$ and $F(N_2)$, respectively.

Factors of 2 are particularly simple to deal with and give rise to a characteristic computational structure called a 'butterfly loop'. If $N = 2M$, then two options exist:

(*a*) using $N_1 = 2$ and $N_2 = M$ leads to collecting the even-numbered coordinates of $\mathbf{X}$ into $\mathbf{Y}_0$ and the odd-numbered coordinates into $\mathbf{Y}_1$

$$\begin{aligned} Y_0(k_2) &= X(2k_2), \qquad k_2 = 0, \ldots, M-1, \\ Y_1(k_2) &= X(2k_2+1), \qquad k_2 = 0, \ldots, M-1, \end{aligned}$$

and writing:

$$\begin{aligned} X^*(k_2^*) &= Y_0^*(k_2^*) + e(k_2^*/N) Y_1^*(k_2^*), \\ &\qquad k_2^* = 0, \ldots, M-1; \\ X^*(k_2^* + M) &= Y_0^*(k_2^*) - e(k_2^*/N) Y_1^*(k_2^*), \\ &\qquad k_2^* = 0, \ldots, M-1. \end{aligned}$$

This is the original version of Cooley & Tukey, and the process of formation of $\mathbf{Y}_0$ and $\mathbf{Y}_1$ is referred to as 'decimation in time' (*i.e.* decimation along the *data* index $\mathbf{k}$).

(*b*) using $N_1 = M$ and $N_2 = 2$ leads to forming

$$\begin{aligned} Z_0(k_1) &= X(k_1) + X(k_1 + M), \qquad k_1 = 0, \ldots, M-1, \\ Z_1(k_1) &= [X(k_1) - X(k_1 + M)] e\left(\frac{k_1}{N}\right), \qquad k_1 = 0, \ldots, M-1, \end{aligned}$$

then obtaining separately the even-numbered and odd-numbered components of $\mathbf{X}^*$ by transforming $\mathbf{Z}_0$ and $\mathbf{Z}_1$:

$$\begin{aligned} X^*(2k_1^*) &= Z_0^*(k_1^*), \quad k_1^* = 0, \ldots, M-1; \\ X^*(2k_1^* + 1) &= Z_1^*(k_1^*), \quad k_1^* = 0, \ldots, M-1. \end{aligned}$$

This version is due to Sande (Gentleman & Sande, 1966), and the process of separately obtaining even-numbered and odd-numbered results has led to its being referred to as 'decimation in frequency' (*i.e.* decimation along the *result* index k^*).

By repeated factoring of the number N of sample points, the calculation of $F(N)$ and $\bar{F}(N)$ can be reduced to a succession of stages, the smallest of which operate on single prime factors of N. The reader is referred to Gentleman & Sande (1966) for a particularly lucid analysis of the programming considerations which help implement this factorization efficiently; see also Singleton (1969). Powers of two are often grouped together into factors of 4 or 8, which are advantageous in that they require fewer complex multiplications than the repeated use of factors of 2. In this approach, large prime factors P are detrimental, since they require a full P^2-size computation according to the defining formula.

1.3.3.2.2. *The Good (or prime factor) algorithm*

1.3.3.2.2.1. *Ring structure on* $\mathbb{Z}/N\mathbb{Z}$

The set $\mathbb{Z}/N\mathbb{Z}$ of congruence classes of integers modulo an integer N [see *e.g.* Apostol (1976), Chapter 5] inherits from $\mathbb{Z}$ not only the additive structure used in deriving the Cooley–Tukey factorization, but also a *multiplicative* structure in which the product of two congruence classes mod N is uniquely defined as the class of the ordinary product (in $\mathbb{Z}$) of representatives of each class. The multiplication can be distributed over addition in the usual way, endowing $\mathbb{Z}/N\mathbb{Z}$ with the structure of a *commutative ring*.

If N is composite, the ring $\mathbb{Z}/N\mathbb{Z}$ has *zero divisors*. For example, let $N = N_1 N_2$, let $n_1 \equiv N_1 \bmod N$, and let $n_2 \equiv N_2 \bmod N$: then $n_1 n_2 \equiv 0 \bmod N$. In the general case, a product of non-zero elements will be zero whenever these elements collect together all the factors of N. These circumstances give rise to a fundamental theorem in the theory of commutative rings, the *Chinese Remainder Theorem* (CRT), which will now be stated and proved [see Apostol (1976), Chapter 5; Schroeder (1986), Chapter 16].

1.3.3.2.2.2. *The Chinese remainder theorem*

Let $N = N_1 N_2 \ldots N_d$ be factored into a product of pairwise coprime integers, so that g.c.d. $(N_i, N_j) = 1$ for $i \neq j$. Then the system of congruence equations

$$\ell \equiv \ell_j \bmod N_j, \qquad j = 1, \ldots, d,$$

has a unique solution $\ell \bmod N$. In other words, each $\ell \in \mathbb{Z}/N\mathbb{Z}$ is

associated in a one-to-one fashion to the d-tuple $(\ell_1, \ell_2, \ldots, \ell_d)$ of its residue classes in $\mathbb{Z}/N_1\mathbb{Z}, \mathbb{Z}/N_2\mathbb{Z}, \ldots, \mathbb{Z}/N_d\mathbb{Z}$.

The proof of the CRT goes as follows. Let

$$Q_j = \frac{N}{N_j} = \prod_{i \neq j} N_i.$$

Since g.c.d. $(N_j, Q_j) = 1$ there exist integers n_j and q_j such that

$$n_j N_j + q_j Q_j = 1, \qquad j = 1, \ldots, d,$$

then the integer

$$\ell = \sum_{i=1}^{d} \ell_i q_i Q_i \bmod N$$

is the solution. Indeed,

$$\ell \equiv \ell_j q_j Q_j \bmod N_j$$

because all terms with $i \neq j$ contain N_j as a factor; and

$$q_j Q_j \equiv 1 \bmod N_j$$

by the defining relation for q_j.

It may be noted that

$$(q_i Q_i)(q_j Q_j) \equiv 0 \qquad \bmod N \text{ for } i \neq j,$$
$$(q_j Q_j)^2 \equiv q_j Q_j \qquad \bmod N, \ \ j = 1, \ldots, d,$$

so that the $q_j Q_j$ are mutually orthogonal *idempotents* in the ring $\mathbb{Z}/N\mathbb{Z}$, with properties formally similar to those of mutually orthogonal *projectors onto subspaces* in linear algebra. The analogy is exact, since by virtue of the CRT the ring $\mathbb{Z}/N\mathbb{Z}$ may be considered as the direct product

$$\mathbb{Z}/N_1\mathbb{Z} \times \mathbb{Z}/N_2\mathbb{Z} \times \ldots \times \mathbb{Z}/N_d\mathbb{Z}$$

via the two mutually inverse mappings:

(i) $\ell \longmapsto (\ell_1, \ell_2, \ldots, \ell_d)$ by $\ell \equiv \ell_j \bmod N_j$ for each j;

(ii) $(\ell_1, \ell_2, \ldots, \ell_d) \longmapsto \ell$ by $\ell = \sum_{i=1}^{d} \ell_i q_i Q_i \bmod N$.

The mapping defined by (ii) is sometimes called the 'CRT reconstruction' of ℓ from the ℓ_j.

These two mappings have the property of sending sums to sums and products to products, *i.e*:

(i) $\ell + \ell' \longmapsto (\ell_1 + \ell'_1, \ell_2 + \ell'_2, \ldots, \ell_d + \ell'_d)$
$\ell\ell' \longmapsto (\ell_1\ell'_1, \ell_2\ell'_2, \ldots, \ell_d\ell'_d)$

(ii) $(\ell_1 + \ell'_1, \ell_2 + \ell'_2, \ldots, \ell_d + \ell'_d) \longmapsto \ell + \ell'$
$(\ell_1\ell'_1, \ell_2\ell'_2, \ldots, \ell_d\ell'_d) \longmapsto \ell\ell'$

(the last proof requires using the properties of the idempotents $q_j Q_j$). This may be described formally by stating that the CRT establishes a *ring isomorphism*:

$$\mathbb{Z}/N\mathbb{Z} \cong (\mathbb{Z}/N_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/N_d\mathbb{Z}).$$

1.3.3.2.2.3. *The prime factor algorithm*

The CRT will now be used to factor the N-point DFT into a tensor product of d transforms, the jth of length N_j.

Let the indices k and k^* be subjected to the following mappings:

(i) $k \longmapsto (k_1, k_2, \ldots, k_d), k_j \in \mathbb{Z}/N_j\mathbb{Z}$, by $k_j \equiv k \bmod N_j$ for each j, with reconstruction formula

$$k = \sum_{i=1}^{d} k_i q_i Q_i \bmod N;$$

(ii) $k^* \longmapsto (k_1^*, k_2^*, \ldots, k_d^*), k_j^* \in \mathbb{Z}/N_j\mathbb{Z}$, by $k_j^* \equiv q_j k^* \bmod N_j$ for each j, with reconstruction formula

$$k^* = \sum_{i=1}^{d} k_i^* Q_i \bmod N.$$

Then

$$k^* k = \left(\sum_{i=1}^{d} k_i^* Q_i\right)\left(\sum_{j=1}^{d} k_j q_j Q_j\right) \bmod N$$
$$= \sum_{i,j=1}^{d} k_i^* k_j Q_i q_j Q_j \bmod N.$$

Cross terms with $i \neq j$ vanish since they contain all the factors of N, hence

$$k^* k = \sum_{j=1}^{d} q_j Q_j^2 k_j^* k_j \bmod N$$
$$= \sum_{j=1}^{d} (1 - n_j N_j) Q_j k_j^* k_j \bmod N.$$

Dividing by N, which may be written as $N_j Q_j$ for each j, yields

$$\frac{k^* k}{N} = \sum_{j=1}^{d} (1 - n_j N_j) \frac{Q_j}{N_j Q_j} k_j^* k_j \bmod 1$$
$$= \sum_{j=1}^{d} \left(\frac{1}{N_j} - n_j\right) k_j^* k_j \bmod 1,$$

and hence

$$\frac{k^* k}{N} \equiv \sum_{j=1}^{d} \frac{k_j^* k_j}{N_j} \bmod 1.$$

Therefore, by the multiplicative property of $e(.)$,

$$e\left(\frac{k^* k}{N}\right) \equiv \bigotimes_{j=1}^{d} e\left(\frac{k_j^* k_j}{N_j}\right).$$

Let $\mathbf{X} \in L(\mathbb{Z}/N\mathbb{Z})$ be described by a one-dimensional array $X(k)$ indexed by k. The index mapping (i) turns $\mathbf{X}$ into an element of $L(\mathbb{Z}/N_1\mathbb{Z} \times \ldots \times \mathbb{Z}/N_d\mathbb{Z})$ described by a d-dimensional array $X(k_1, \ldots, k_d)$; the latter may be transformed by $\bar{F}(N_1) \bigotimes \ldots \bigotimes \bar{F}(N_d)$ into a new array $X^*(k_1^*, k_2^*, \ldots, k_d^*)$. Finally, the one-dimensional array of results $X^*(k^*)$ will be obtained by reconstructing k^* according to (ii).

The prime factor algorithm, like the Cooley–Tukey algorithm, reindexes a 1D transform to turn it into d separate transforms, but the use of coprime factors and CRT index mapping leads to the further gain that *no twiddle factors* need to be applied between the successive transforms (see Good, 1971). This makes up for the cost of the added complexity of the CRT index mapping.

The natural factorization of N for the prime factor algorithm is thus its factorization into prime powers: $\bar{F}(N)$ is then the tensor product of separate transforms (one for each prime power factor $N_j = p_j^{\nu_j}$) whose results can be reassembled without twiddle factors. The separate factors p_j within each N_j must then be dealt with by another algorithm (*e.g.* Cooley–Tukey, which does require twiddle factors). Thus, the DFT on a prime number of points remains undecomposable.

1.3.3.2.3. *The Rader algorithm*

The previous two algorithms essentially reduce the calculation of the DFT on N points for N composite to the calculation of smaller DFTs on prime numbers of points, the latter remaining irreducible. However, Rader (1968) showed that the p-point DFT for p an odd

prime can itself be factored by invoking some extra arithmetic structure present in $\mathbb{Z}/p\mathbb{Z}$.

1.3.3.2.3.1. *N an odd prime*

The ring $\mathbb{Z}/p\mathbb{Z} = \{0, 1, 2, \ldots, p-1\}$ has the property that its $p-1$ non-zero elements, called *units*, form a *multiplicative group* $U(p)$. In particular, all units $r \in U(p)$ have a unique multiplicative inverse in $\mathbb{Z}/p\mathbb{Z}$, *i.e.* a unit $s \in U(p)$ such that $rs \equiv 1 \bmod p$. This endows $\mathbb{Z}/p\mathbb{Z}$ with the structure of a *finite field*.

Furthermore, $U(p)$ is a *cyclic* group, *i.e.* consists of the successive powers $g^m \bmod p$ of a generator g called a *primitive root mod p* (such a g may not be unique, but it always exists). For instance, for $p = 7$, $U(7) = \{1, 2, 3, 4, 5, 6\}$ is generated by $g = 3$, whose successive powers mod 7 are:

$$g^0 = 1, \quad g^1 = 3, \quad g^2 = 2, \quad g^3 = 6, \quad g^4 = 4, \quad g^5 = 5$$

[see Apostol (1976), Chapter 10].

The basis of Rader's algorithm is to bring to light a hidden regularity in the matrix $F(p)$ by permuting the basis vectors $\mathbf{u}_k$ and $\mathbf{v}_{k^*}$ of $L(\mathbb{Z}/p\mathbb{Z})$ as follows:

$$\begin{aligned}
\mathbf{u}'_0 &= \mathbf{u}_0 \\
\mathbf{u}'_m &= \mathbf{u}_k \quad \text{with } k = g^m, \qquad m = 1, \ldots, p-1; \\
\mathbf{v}'_0 &= \mathbf{v}_0 \\
\mathbf{v}'_{m^*} &= \mathbf{v}_{k^*} \quad \text{with } k^* = g^{m^*}, \quad m^* = 1, \ldots, p-1;
\end{aligned}$$

where g is a primitive root mod p.

With respect to these new bases, the matrix representing $\bar{F}(p)$ will have the following elements:

$$\begin{aligned}
\text{element } (0,0) &= 1 \\
\text{element } (0, m+1) &= 1 \quad \text{for all } m = 0, \ldots p-2, \\
\text{element } (m^*+1, 0) &= 1 \quad \text{for all } m^* = 0, \ldots, p-2, \\
\text{element } (m^*+1, m+1) &= e\left(\frac{k^* k}{p}\right) \\
&= e(g^{(m^*+m)/p}) \\
&\qquad \text{for all } m^* = 0, \ldots, p-2.
\end{aligned}$$

Thus the 'core' $\bar{C}(p)$ of matrix $\bar{F}(p)$, of size $(p-1) \times (p-1)$, formed by the elements with two non-zero indices, has a so-called *skew-circulant* structure because element (m^*, m) depends only on $m^* + m$. Simplification may now occur because multiplication by $\bar{C}(p)$ is closely related to a *cyclic convolution*. Introducing the notation $C(m) = e(g^{m/p})$ we may write the relation $\mathbf{Y}^* = \bar{F}(p)\mathbf{Y}$ in the permuted bases as

$$\begin{aligned}
Y^*(0) &= \sum_k Y(k) \\
Y^*(m^*+1) &= Y(0) + \sum_{m=0}^{p-2} C(m^*+m) Y(m+1) \\
&= Y(0) + \sum_{m=0}^{p-2} C(m^*-m) Z(m) \\
&= Y(0) + (\mathbf{C} * \mathbf{Z})(m^*), \quad m^* = 0, \ldots, p-2,
\end{aligned}$$

where $\mathbf{Z}$ is defined by $Z(m) = Y(p-m-2)$, $m = 0, \ldots, p-2$.

Thus $\mathbf{Y}^*$ may be obtained by cyclic convolution of $\mathbf{C}$ and $\mathbf{Z}$, which may for instance be calculated by

$$\mathbf{C} * \mathbf{Z} = F(p-1)[\bar{F}(p-1)[\mathbf{C}] \times \bar{F}(p-1)[\mathbf{Z}]],$$

where $\times$ denotes the component-wise multiplication of vectors. Since p is odd, $p-1$ is always divisible by 2 and may even be highly composite. In that case, factoring $\bar{F}(p-1)$ by means of the Cooley–Tukey or Good methods leads to an algorithm of complexity $p \log p$ rather than p^2 for $\bar{F}(p)$. An added bonus is that, because $g^{(p-1)/2} = -1$, the elements of $\bar{F}(p-1)[\mathbf{C}]$ can be shown to be either purely real or purely imaginary, which halves the number of real multiplications involved.

1.3.3.2.3.2. *N a power of an odd prime*

This idea was extended by Winograd (1976, 1978) to the treatment of prime powers $N = p^\nu$, using the cyclic structure of the multiplicative group of units $U(p^\nu)$. The latter consists of all those elements of $\mathbb{Z}/p^\nu\mathbb{Z}$ which are not divisible by p, and thus has $q_\nu = p^{\nu-1}(p-1)$ elements. It is cyclic, and there exist primitive roots g modulo p^ν such that

$$U(p^\nu) = \{1, g, g^2, g^3, \ldots, g^{q_\nu - 1}\}.$$

The $p^{\nu-1}$ elements divisible by p, which are divisors of zero, have to be treated separately just as 0 had to be treated separately for $N = p$.

When $k^* \notin U(p^\nu)$, then $k^* = pk_1^*$ with $k_1^* \in \mathbb{Z}/p^{\nu-1}\mathbb{Z}$. The results $X^*(pk_1^*)$ are p-decimated, hence can be obtained *via* the $p^{\nu-1}$-point DFT of the $p^{\nu-1}$-periodized data $\mathbf{Y}$:

$$X^*(pk_1^*) = \bar{F}(p^{\nu-1})[\mathbf{Y}](k_1^*)$$

with

$$Y(k_1) = \sum_{k_2 \in \mathbb{Z}/p\mathbb{Z}} X(k_1 + p^{\nu-1}k_2).$$

When $k^* \in U(p^\nu)$, then we may write

$$X^*(k^*) = X_0^*(k^*) + X_1^*(k^*),$$

where $\mathbf{X}_0^*$ contains the contributions from $k \notin U(p^\nu)$ and $\mathbf{X}_1^*$ those from $k \in U(p^\nu)$. By a converse of the previous calculation, $\mathbf{X}_0^*$ arises from p-decimated data $\mathbf{Z}$, hence is the $p^{\nu-1}$-periodization of the $p^{\nu-1}$-point DFT of these data:

$$X_0^*(p^{\nu-1}k_1^* + k_2^*) = \bar{F}(p^{\nu-1})[\mathbf{Z}](k_2^*)$$

with

$$Z(k_2) = X(pk_2), \qquad k_2 \in \mathbb{Z}/p^{\nu-1}\mathbb{Z}$$

(the $p^{\nu-1}$-periodicity follows implicity from the fact that the transform on the right-hand side is independent of $k_1^* \in \mathbb{Z}/p\mathbb{Z}$).

Finally, the contribution X_1^* from all $k \in U(p^\nu)$ may be calculated by reindexing by the powers of a primitive root g modulo p^ν, *i.e.* by writing

$$X_1^*(g^{m^*}) = \sum_{m=0}^{q_\nu - 1} X(g^m) e(g^{(m^*+m)/p^\nu})$$

then carrying out the multiplication by the skew-circulant matrix core as a convolution.

Thus the DFT of size p^ν may be reduced to two DFTs of size $p^{\nu-1}$ (dealing, respectively, with p-decimated results and p-decimated data) and a convolution of size $q_\nu = p^{\nu-1}(p-1)$. The latter may be 'diagonalized' into a multiplication by purely real or purely imaginary numbers (because $g^{(q_\nu/2)} = -1$) by two DFTs, whose factoring in turn leads to DFTs of size $p^{\nu-1}$ and $p-1$. This method, applied recursively, allows the complete decomposition of the DFT on p^ν points into arbitrarily small DFTs.

1.3.3.2.3.3. *N a power of 2*

When $N = 2^\nu$, the same method can be applied, except for a slight modification in the calculation of $\mathbf{X}_1^*$. There is no primitive root modulo 2^ν for $\nu > 2$: the group $U(2^\nu)$ is the direct product of *two* cyclic groups, the first (of order 2) generated by -1, the second (of order $N/4$) generated by 3 or 5. One then uses a representation

$$k = (-1)^{m_1} 5^{m_2}$$
$$k^* = (-1)^{m_1^*} 5^{m_2^*}$$

and the reindexed core matrix gives rise to a two-dimensional convolution. The latter may be carried out by means of two 2D DFTs on $2 \times (N/4)$ points.

1.3.3.2.4. *The Winograd algorithms*

The cyclic convolutions generated by Rader's multiplicative reindexing may be evaluated more economically than through DFTs if they are re-examined within a new algebraic setting, namely the theory of congruence classes of polynomials [see, for instance, Blahut (1985), Chapter 2; Schroeder (1986), Chapter 24].

The set, denoted $\mathbb{K}[X]$, of polynomials in one variable with coefficients in a given field $\mathbb{K}$ has many of the formal properties of the set $\mathbb{Z}$ of rational integers: it is a *ring* with no zero divisors and has a *Euclidean algorithm* on which a theory of divisibility can be built.

Given a polynomial $P(z)$, then for every $W(z)$ there exist unique polynomials $Q(z)$ and $R(z)$ such that

$$W(z) = P(z)Q(z) + R(z)$$

and

$$\text{degree } (R) \; < \; \text{degree } (P).$$

$R(z)$ is called the *residue* of $H(z)$ modulo $P(z)$. Two polynomials $H_1(z)$ and $H_2(z)$ having the same residue modulo $P(z)$ are said to be *congruent* modulo $P(z)$, which is denoted by

$$H_1(z) \equiv H_2(z) \bmod P(z).$$

If $H(z) \equiv 0 \bmod P(z)$, $H(z)$ is said to be *divisible* by $P(z)$. If $H(z)$ only has divisors of degree zero in $\mathbb{K}[X]$, it is said to be *irreducible over* $\mathbb{K}$ (this notion depends on $\mathbb{K}$). Irreducible polynomials play in $\mathbb{K}[X]$ a role analogous to that of prime numbers in $\mathbb{Z}$, and any polynomial over $\mathbb{K}$ has an essentially unique factorization as a product of irreducible polynomials.

There exists a *Chinese remainder theorem* (CRT) for polynomials. Let $P(z) = P_1(z) \ldots P_d(z)$ be factored into a product of pairwise coprime polynomials [*i.e.* $P_i(z)$ and $P_j(z)$ have no common factor for $i \neq j$]. Then the system of congruence equations

$$H(z) \equiv H_j(z) \bmod P_j(z), \quad j = 1, \ldots, d,$$

has a unique solution $H(z)$ modulo $P(z)$. This solution may be constructed by a procedure similar to that used for integers. Let

$$Q_j(z) = P(z)/P_j(z) = \prod_{i \neq j} P_i(z).$$

Then P_j and Q_j are coprime, and the Euclidean algorithm may be used to obtain polynomials $p_j(z)$ and $q_j(z)$ such that

$$p_j(z)P_j(z) + q_j(z)Q_j(z) = 1.$$

With $S_i(z) = q_i(z)Q_i(z)$, the polynomial

$$H(z) = \sum_{i=1}^{d} S_i(z)H_i(z) \bmod P(z)$$

is easily shown to be the desired solution.

As with integers, it can be shown that the 1:1 correspondence between $H(z)$ and $H_j(z)$ sends sums to sums and products to products, *i.e.* establishes a *ring isomorphism*:

$$\mathbb{K}[X] \bmod P \cong (\mathbb{K}[X] \bmod P_1) \times \ldots \times (\mathbb{K}[X] \bmod P_d).$$

These results will now be applied to the efficient calculation of cyclic convolutions. Let $\mathbf{U} = (u_0, u_1, \ldots, u_{N-1})$ and $\mathbf{V} = (v_0, v_1, \ldots, v_{N-1})$ be two vectors of length N, and let $\mathbf{W} = (w_0, w_1, \ldots, w_{N-1})$ be obtained by cyclic convolution of $\mathbf{U}$ and $\mathbf{V}$:

$$w_n = \sum_{m=0}^{N-1} u_m v_{n-m}, \quad n = 0, \ldots, N-1.$$

The very simple but crucial result is that this cyclic convolution may be carried out by *polynomial multiplication modulo* $(z^N - 1)$: if

$$U(z) = \sum_{l=0}^{N-1} u_l z^l$$
$$V(z) = \sum_{m=0}^{N-1} v_m z^m$$
$$W(z) = \sum_{n=0}^{N-1} w_n z^n$$

then the above relation is equivalent to

$$W(z) \equiv U(z)V(z) \bmod (z^N - 1).$$

Now the polynomial $z^N - 1$ can be *factored* over the field of rational numbers into irreducible factors called *cyclotomic polynomials*: if d is the number of divisors of N, including 1 and N, then

$$z^N - 1 = \prod_{i=1}^{d} P_i(z),$$

where the cyclotomics $P_i(z)$ are well known (Nussbaumer, 1981; Schroeder, 1986, Chapter 22). We may now invoke the CRT, and exploit the ring isomorphism it establishes to simplify the calculation of $W(z)$ from $U(z)$ and $V(z)$ as follows:

(i) compute the d residual polynomials

$$U_i(z) \equiv U(z) \bmod P_i(z), \quad i = 1, \ldots, d,$$
$$V_i(z) \equiv V(z) \bmod P_i(z), \quad i = 1, \ldots, d;$$

(ii) compute the d polynomial products

$$W_i(z) \equiv U_i(z)V_i(z) \bmod P_i(z), \quad i = 1, \ldots, d;$$

(iii) use the CRT reconstruction formula just proved to recover $W(z)$ from the $W_i(z)$:

$$W(z) \equiv \sum_{i=1}^{d} S_i(z)W_i(z) \bmod (z^N - 1).$$

When N is not too large, *i.e.* for 'short cyclic convolutions', the $P_i(z)$ are very simple, with coefficients 0 or ± 1, so that (i) only involves a small number of additions. Furthermore, special techniques have been developed to multiply general polynomials modulo cyclotomic polynomials, thus helping keep the number of multiplications in (ii) and (iii) to a minimum. As a result, cyclic convolutions can be calculated rapidly when N is sufficiently composite.

It will be recalled that Rader's multiplicative indexing often gives rise to cyclic convolutions of length $p - 1$ for p an odd prime. Since $p - 1$ is highly composite for all $p \leq 50$ other than 23 and 47, these cyclic convolutions can be performed more efficiently by the above procedure than by DFT.

These combined algorithms are due to Winograd (1977, 1978, 1980), and are known collectively as 'Winograd small FFT algorithms'. Winograd also showed that they can be thought of as bringing the DFT matrix $\mathbf{F}$ to the following 'normal form':

$$\mathbf{F} = \mathbf{CBA},$$

where

$\mathbf{A}$ is an integer matrix with entries 0, ± 1, defining the 'pre-additions',

B is a diagonal matrix of multiplications,

C is a matrix with entries $0, \pm 1, \pm i$, defining the 'post-additions'.

The elements on the diagonal of **B** can be shown to be either real or pure imaginary, by the same argument as in Section 1.3.3.2.3.1. Matrices **A** and **C** may be rectangular rather than square, so that intermediate results may require extra storage space.

1.3.3.3. *Multidimensional algorithms*

From an algorithmic point of view, the distinction between one-dimensional (1D) and multidimensional DFTs is somewhat blurred by the fact that some factoring techniques turn a 1D transform into a multidimensional one. The distinction made here, however, is a practical one and is based on the dimensionality of the indexing sets for data and results. This section will therefore be concerned with the problem of factoring the DFT when the *indexing sets* for the input data and output results are multidimensional.

1.3.3.3.1. *The method of successive one-dimensional transforms*

The DFT was defined in Section 1.3.2.7.4 in an n-dimensional setting and it was shown that when the decimation matrix **N** is diagonal, say $\mathbf{N} = \mathrm{diag}(N^{(1)}, N^{(2)}, \ldots, N^{(n)})$, then $\bar{F}(N)$ has a tensor product structure:

$$\bar{F}(\mathbf{N}) = \bar{F}(N^{(1)}) \otimes \bar{F}(N^{(2)}) \otimes \ldots \otimes \bar{F}(N^{(n)}).$$

This may be rewritten as follows:

$$\begin{aligned}\bar{F}(\mathbf{N}) = {} & [\bar{F}(N^{(1)}) \otimes I_{N^{(2)}} \otimes \ldots \otimes I_{N^{(n)}}] \\ & \times [I_{N^{(1)}} \otimes \bar{F}(N^{(2)}) \otimes \ldots \otimes I_{N^{(n)}}] \\ & \times \ldots \\ & \times [I_{N^{(1)}} \otimes I_{N^{(2)}} \otimes \ldots \otimes \bar{F}(N^{(n)}],\end{aligned}$$

where the I's are identity matrices and $\times$ denotes ordinary matrix multiplication. The matrix within each bracket represents a one-dimensional DFT along one of the n dimensions, the other dimensions being left untransformed. As these matrices commute, the order in which the successive 1D DFTs are performed is immaterial.

This is the most straightforward method for building an n-dimensional algorithm from existing 1D algorithms. It is known in crystallography under the name of 'Beevers–Lipson factorization' (Section 1.3.4.3.1), and in signal processing as the 'row–column method'.

1.3.3.3.2. *Multidimensional factorization*

Substantial reductions in the arithmetic cost, as well as gains in flexibility, can be obtained if the factoring of the DFT is carried out in several dimensions simultaneously. The presentation given here is a generalization of that of Mersereau & Speake (1981), using the abstract setting established independently by Auslander, Tolimieri & Winograd (1982).

Let us return to the general n-dimensional setting of Section 1.3.2.7.4, where the DFT was defined for an arbitrary decimation matrix **N** by the formulae (where $|\mathbf{N}|$ denotes $|\det \mathbf{N}|$):

$$F(\mathbf{N}): \quad X(\mathbf{k}) = \frac{1}{|\mathbf{N}|} \sum_{\mathbf{k}^*} X^*(\mathbf{k}^*) e[-\mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]$$

$$\bar{F}(\mathbf{N}): \quad X^*(\mathbf{k}^*) = \sum_{\mathbf{k}} X(\mathbf{k}) e[\mathbf{k}^* \cdot (\mathbf{N}^{-1}\mathbf{k})]$$

with

$$\mathbf{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n, \quad \mathbf{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n.$$

1.3.3.3.2.1. *Multidimensional Cooley–Tukey factorization*

Let us now assume that this decimation can be factored into d successive decimations, *i.e.* that

$$\mathbf{N} = \mathbf{N}_1\mathbf{N}_2 \ldots \mathbf{N}_{d-1}\mathbf{N}_d$$

and hence

$$\mathbf{N}^T = \mathbf{N}_d^T\mathbf{N}_{d-1}^T \ldots \mathbf{N}_2^T\mathbf{N}_1^T.$$

Then the coset decomposition formulae corresponding to these successive decimations (Section 1.3.2.7.1) can be combined as follows:

$$\begin{aligned}\mathbb{Z}^n &= \bigcup_{\mathbf{k}_1} (\mathbf{k}_1 + \mathbf{N}_1\mathbb{Z}^n) \\ &= \bigcup_{\mathbf{k}_1} \left\{ \mathbf{k}_1 + \mathbf{N}_1 \left[\bigcup_{\mathbf{k}_2} (\mathbf{k}_2 + \mathbf{N}_2\mathbb{Z}^n) \right] \right\} \\ &= \ldots \\ &= \bigcup_{\mathbf{k}_1} \ldots \bigcup_{\mathbf{k}_d} (\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2 + \ldots + \mathbf{N}_1\mathbf{N}_2 \times \ldots \times \mathbf{N}_{d-1}\mathbf{k}_d + \mathbf{N}\mathbb{Z}^n)\end{aligned}$$

with $\mathbf{k}_j \in \mathbb{Z}^n/\mathbf{N}_j\mathbb{Z}^n$. Therefore, any $\mathbf{k} \in \mathbb{Z}/\mathbf{N}\mathbb{Z}^n$ may be written uniquely as

$$\mathbf{k} = \mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2 + \ldots + \mathbf{N}_1\mathbf{N}_2 \times \ldots \times \mathbf{N}_{d-1}\mathbf{k}_d.$$

Similarly:

$$\begin{aligned}\mathbb{Z}^n &= \bigcup_{\mathbf{k}_d^*} (\mathbf{k}_d^* + \mathbf{N}_d^T\mathbb{Z}^n) \\ &= \ldots \\ &= \bigcup_{\mathbf{k}_d^*} \ldots \bigcup_{\mathbf{k}_1^*} (\mathbf{k}_d^* + \mathbf{N}_d^T\mathbf{k}_{d-1}^* + \ldots + \mathbf{N}_d^T \times \ldots \times \mathbf{N}_2^T\mathbf{k}_1^* \\ &\quad + \mathbf{N}^T\mathbb{Z}^n)\end{aligned}$$

so that any $\mathbf{k}^* \in \mathbb{Z}^n/\mathbf{N}^T\mathbb{Z}^n$ may be written uniquely as

$$\mathbf{k}^* = \mathbf{k}_d^* + \mathbf{N}_d^T\mathbf{k}_{d-1}^* + \ldots + \mathbf{N}_d^T \times \ldots \times \mathbf{N}_2^T\mathbf{k}_1^*$$

with $\mathbf{k}_j^* \in \mathbb{Z}^n/\mathbf{N}_j^T\mathbb{Z}^n$. These decompositions are the vector analogues of the multi-radix number representation systems used in the Cooley–Tukey factorization.

We may then write the definition of $\bar{F}(\mathbf{N})$ with $d = 2$ factors as

$$\begin{aligned}X^*(\mathbf{k}_2^* + \mathbf{N}_2^T\mathbf{k}_1^*) = {} & \sum_{\mathbf{k}_1} \sum_{\mathbf{k}_2} X(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2) \\ & \times e[(\mathbf{k}_2^{*T} + \mathbf{k}_1^{*T}\mathbf{N}_2)\mathbf{N}_2^{-1}\mathbf{N}_1^{-1}(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2)].\end{aligned}$$

The argument of $e(-)$ may be expanded as

$$\mathbf{k}_2^* \cdot (\mathbf{N}^{-1}\mathbf{k}_1) + \mathbf{k}_1^* \cdot (\mathbf{N}_1^{-1}\mathbf{k}_1) + \mathbf{k}_2^* \cdot (\mathbf{N}_2^{-1}\mathbf{k}_2) + \mathbf{k}_1^* \cdot \mathbf{k}_2.$$

The first summand may be recognized as a twiddle factor, the second and third as the kernels of $\bar{F}(\mathbf{N}_1)$ and $\bar{F}(\mathbf{N}_2)$, respectively, while the fourth is an integer which may be dropped. We are thus led to a 'vector-radix' version of the Cooley–Tukey algorithm, in which the successive decimations may be introduced in all n dimensions simultaneously by general integer matrices. The computation may be decomposed into five stages analogous to those of the one-dimensional algorithm of Section 1.3.3.2.1:

(i) form the $|\mathbf{N}_1|$ vectors $\mathbf{Y}_{\mathbf{k}_1}$ of shape $\mathbf{N}_2$ by

$$Y_{\mathbf{k}_1}(\mathbf{k}_2) = X(\mathbf{k}_1 + \mathbf{N}_1\mathbf{k}_2), \quad \mathbf{k}_1 \in \mathbb{Z}^n/\mathbf{N}_1\mathbb{Z}^n, \quad \mathbf{k}_2 \in \mathbb{Z}^n/\mathbf{N}_2\mathbb{Z}^n;$$

(ii) calculate the $|\mathbf{N}_1|$ transforms $\mathbf{Y}^*_{\mathbf{k}_1}$ on $|\mathbf{N}_2|$ points:

$$Y^*_{\mathbf{k}_1}(\mathbf{k}^*_2) = \sum_{\mathbf{k}_2} e[\mathbf{k}^*_2 \cdot (\mathbf{N}_2^{-1}\mathbf{k}_2)] Y_{\mathbf{k}_1}(\mathbf{k}_2), \quad \mathbf{k}_1 \in \mathbb{Z}^n/\mathbf{N}_1\mathbb{Z}^n;$$

(iii) form the $|\mathbf{N}_2|$ vectors $\mathbf{Z}_{\mathbf{k}^*_2}$ of shape $\mathbf{N}_1$ by

$$Z_{\mathbf{k}^*_2}(\mathbf{k}_1) = e[\mathbf{k}^*_2 \cdot (\mathbf{N}^{-1}\mathbf{k}_1)] Y^*_{\mathbf{k}_1}(\mathbf{k}^*_2), \quad \mathbf{k}_1 \in \mathbb{Z}^n/\mathbf{N}_1\mathbb{Z}^n, \quad \mathbf{k}^*_2 \in \mathbb{Z}^n/\mathbf{N}_2^T\mathbb{Z}^n;$$

(iv) calculate the $|\mathbf{N}_2|$ transforms $\mathbf{Z}^*_{\mathbf{k}^*_2}$ on $|\mathbf{N}_1|$ points:

$$Z^*_{\mathbf{k}^*_2}(\mathbf{k}^*_1) = \sum_{\mathbf{k}_1} e[\mathbf{k}^*_1 \cdot (\mathbf{N}_1^{-1}\mathbf{k}_1)] Z_{\mathbf{k}^*_2}(\mathbf{k}_1), \quad \mathbf{k}^*_2 \in \mathbb{Z}^n/\mathbf{N}_2^T\mathbb{Z}^n;$$

(v) collect $X^*(\mathbf{k}^*_2 + \mathbf{N}_2^T\mathbf{k}^*_1)$ as $Z^*_{\mathbf{k}^*_2}(\mathbf{k}^*_1)$.

The initial $|\mathbf{N}|$-point transform $\bar{F}(\mathbf{N})$ can thus be performed as $|\mathbf{N}_1|$ transforms $\bar{F}(\mathbf{N}_2)$ on $|\mathbf{N}_2|$ points, followed by $|\mathbf{N}_2|$ transforms $\bar{F}(\mathbf{N}_1)$ on $|\mathbf{N}_1|$ points. This process can be applied successively to all d factors. The same decomposition applies to $F(\mathbf{N})$, up to the complex conjugation of twiddle factors, the normalization factor $1/|\mathbf{N}|$ being obtained as the product of the factors $1/|\mathbf{N}_j|$ in the successive partial transforms $F(\mathbf{N}_j)$.

The geometric interpretation of this factorization in terms of partial transforms on translates of sublattices applies in full to this n-dimensional setting; in particular, the twiddle factors are seen to be related to the residual translations which place the sublattices in register within the big lattice. If the intermediate transforms are performed *in place*, then the quantity

$$X^*(\mathbf{k}^*_d + \mathbf{N}_d^T\mathbf{k}^*_{d-1} + \ldots + \mathbf{N}_d^T\mathbf{N}_{d-1}^T \times \ldots \times \mathbf{N}_2^T\mathbf{k}^*_1)$$

will eventually be found at location

$$\mathbf{k}^*_1 + \mathbf{N}_1\mathbf{k}^*_2 + \ldots + \mathbf{N}_1\mathbf{N}_2 \times \ldots \times \mathbf{N}_{d-1}\mathbf{k}^*_d,$$

so that the final results will have to be *unscrambled* by a process which may be called 'coset reversal', the vector equivalent of digit reversal.

Factoring by 2 in all n dimensions simultaneously, *i.e.* taking $\mathbf{N} = 2\mathbf{M}$, leads to 'n-dimensional butterflies'. Decimation in time corresponds to the choice $\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$, so that $\mathbf{k}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n$ is an n-dimensional parity class; the calculation then proceeds by

$$Y_{\mathbf{k}_1}(\mathbf{k}_2) = X(\mathbf{k}_1 + 2\mathbf{k}_2), \quad \mathbf{k}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n, \quad \mathbf{k}_2 \in \mathbb{Z}^n/\mathbf{M}\mathbb{Z}^n,$$

$$Y^*_{\mathbf{k}_1} = \bar{F}(\mathbf{M})[\mathbf{Y}_{\mathbf{k}_1}], \quad \mathbf{k}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n;$$

$$X^*(\mathbf{k}^*_2 + \mathbf{M}^T\mathbf{k}^*_1) = \sum_{\mathbf{k}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n} (-1)^{\mathbf{k}^*_1 \cdot \mathbf{k}_1} \times e[\mathbf{k}^*_2 \cdot (\mathbf{N}^{-1}\mathbf{k}_1)] Y^*_{\mathbf{k}_1}(\mathbf{k}^*_2).$$

Decimation in frequency corresponds to the choice $\mathbf{N}_1 = \mathbf{M}$, $\mathbf{N}_2 = 2\mathbf{I}$, so that $\mathbf{k}_2 \in \mathbb{Z}^n/2\mathbb{Z}^n$ labels 'octant' blocks of shape $\mathbf{M}$; the calculation then proceeds through the following steps:

$$Z_{\mathbf{k}^*_2}(\mathbf{k}_1) = \left[\sum_{\mathbf{k}_2 \in \mathbb{Z}^n/2\mathbb{Z}^n} (-1)^{\mathbf{k}^*_2 \cdot \mathbf{k}_2} X(\mathbf{k}_1 + \mathbf{M}\mathbf{k}_2) \right] \times e[\mathbf{k}^*_2 \cdot (\mathbf{N}^{-1}\mathbf{k}_1)],$$

$$\mathbf{Z}^*_{\mathbf{k}^*_2} = \bar{F}(\mathbf{M})[\mathbf{Z}_{\mathbf{k}^*_2}],$$

$$X^*(\mathbf{k}^*_2 + 2\mathbf{k}^*_1) = Z^*_{\mathbf{k}^*_2}(\mathbf{k}^*_1),$$

i.e. the 2^n parity classes of results, corresponding to the different $\mathbf{k}^*_2 \in \mathbb{Z}^n/2\mathbb{Z}^n$, are obtained separately. When the dimension n is 2 and the decimating matrix is diagonal, this analysis reduces to the 'vector radix FFT' algorithms proposed by Rivard (1977) and Harris *et al.* (1977). These lead to substantial reductions in the number M of multiplications compared to the row–column method: M is reduced to $3M/4$ by simultaneous 2×2 factoring, and to $15M/32$ by simultaneous 4×4 factoring.

The use of a non-diagonal decimating matrix may bring savings in computing time if the spectrum of the band-limited function under study is of such a shape as to pack more compactly in a non-rectangular than in a rectangular lattice (Mersereau, 1979). If, for instance, the support K of the spectrum Φ is contained in a sphere, then a decimation matrix producing a close packing of these spheres will yield an aliasing-free DFT algorithm with fewer sample points than the standard algorithm using a rectangular lattice.

1.3.3.3.2.2. *Multidimensional prime factor algorithm*

Suppose that the decimation matrix $\mathbf{N}$ is diagonal

$$\mathbf{N} = \text{diag } (N^{(1)}, N^{(2)}, \ldots, N^{(n)})$$

and let each diagonal element be written in terms of its prime factors:

$$N^{(i)} = \prod_{j=1}^{m} p_j^{\nu(i,\,j)},$$

where m is the total number of distinct prime factors present in the $N^{(i)}$.

The CRT may be used to turn each 1D transform along dimension i $(i = 1, \ldots, n)$ into a multidimensional transform with a separate 'pseudo-dimension' for each distinct prime factor of $N^{(i)}$; the number μ_i, of these pseudo-dimensions is equal to the cardinality of the set:

$$\{ j \in \{1, \ldots, m\} | \nu(i,j) > 0 \text{ for some } i \}.$$

The full n-dimensional transform thus becomes μ-dimensional, with $\mu = \sum_{i=1}^{n} \mu_i$.

We may now permute the μ pseudo-dimensions so as to bring into contiguous position those corresponding to the same prime factor p_j; the m resulting groups of pseudo-dimensions are said to define 'p-primary' blocks. The initial transform is now written as a tensor product of m p-primary transforms, where transform j is on

$$p_j^{\nu(1,\,j)} \times p_j^{\nu(2,\,j)} \times \ldots \times p_j^{\nu(n,\,j)}$$

points [by convention, dimension i is not transformed if $\nu(i,j) = 0$]. These p-primary transforms may be computed, for instance, by multidimensional Cooley–Tukey factorization (Section 1.3.3.3.1), which is faster than the straightforward row–column method. The final results may then be obtained by reversing all the permutations used.

The extra gain with respect to the multidimensional Cooley–Tukey method is that *there are no twiddle factors between p-primary pieces corresponding to different primes p*.

The case where $\mathbf{N}$ is not diagonal has been examined by Guessoum & Mersereau (1986).

1.3.3.3.2.3. *Nesting of Winograd small FFTs*

Suppose that the CRT has been used as above to map an n-dimensional DFT to a μ-dimensional DFT. For each $\kappa = 1, \ldots, \mu$ [κ runs over those pairs (i, j) such that $\nu(i,j) > 0$], the Rader/Winograd procedure may be applied to put the matrix of the κth 1D DFT in the **CBA** normal form of a Winograd small FFT. The full DFT matrix may then be written, up to permutation of data and results, as

$$\bigotimes_{\kappa=1}^{\mu} (\mathbf{C}_\kappa \mathbf{B}_\kappa \mathbf{A}_\kappa).$$

A well known property of the tensor product of matrices allows this to be rewritten as

$$\left(\bigotimes_{\gamma=1}^{\mu}\mathbf{C}_\gamma\right)\times\left(\bigotimes_{\beta=1}^{\mu}\mathbf{B}_\beta\right)\times\left(\bigotimes_{\alpha=1}^{\mu}\mathbf{A}_\alpha\right)$$

and thus to form a matrix in which the *combined* pre-addition, multiplication and post-addition matrices have been *precomputed*. This procedure, called *nesting*, can be shown to afford a reduction of the arithmetic operation count compared to the row–column method (Morris, 1978).

Clearly, the nesting rearrangement need not be applied to all μ dimensions, but can be restricted to any desired subset of them.

1.3.3.3.2.4. *The Nussbaumer–Quandalle algorithm*

Nussbaumer's approach views the DFT as the evaluation of certain polynomials constructed from the data (as in Section 1.3.3.2.4). For instance, putting $\omega = e(1/N)$, the 1D N-point DFT

$$X^*(k^*) = \sum_{k=0}^{N-1} X(k)\omega^{k^*k}$$

may be written

$$X^*(k^*) = Q(\omega^{k^*}),$$

where the polynomial Q is defined by

$$Q(z) = \sum_{k=0}^{N-1} X(k)z^k.$$

Let us consider (Nussbaumer & Quandalle, 1979) a 2D transform of size $N \times N$:

$$X^*(k_1^*, k_2^*) = \sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1} X(k_1,k_2)\omega^{k_1^*k_1+k_2^*k_2}.$$

By introduction of the polynomials

$$Q_{k_2}(z) = \sum_{k_1} X(k_1,k_2)z^{k_1}$$

$$R_{k_2^*}(z) = \sum_{k_2} \omega^{k_2^*k_2}Q_{k_2}(z),$$

this may be rewritten:

$$X^*(k_1^*, k_2^*) = R_{k_2^*}(\omega^{k_1^*}) = \sum_{k_2} \omega^{k_2^*k_2}Q_{k_2}(\omega^{k_1^*}).$$

Let us now suppose that k_1^* is coprime to N. Then k_1^* has a unique inverse modulo N (denoted by $1/k_1^*$), so that multiplication by k_1^* simply permutes the elements of $\mathbb{Z}/N\mathbb{Z}$ and hence

$$\sum_{k_2=0}^{N-1} f(k_2) = \sum_{k_2=0}^{N-1} f(k_1^*k_2)$$

for any function f over $\mathbb{Z}/N\mathbb{Z}$. We may thus write:

$$\begin{aligned} X^*(k_1^*, k_2^*) &= \sum_{k_2} \omega^{k_1^*k_2^*k_2}Q_{k_1^*k_2}(\omega^{k_1^*}) \\ &= S_{k_1^*k_2}(\omega^{k_1^*}) \end{aligned}$$

where

$$S_{k^*}(z) = \sum_{k_2} z^{k^*k_2}Q_{k_2}(z).$$

Since only the value of polynomial $S_{k^*}(z)$ at $z = \omega^{k_1^*}$ is involved in the result, the computation of S_{k^*} may be carried out modulo the unique cyclotomic polynomial $P(z)$ such that $P(\omega^{k_1^*}) = 0$. Thus, if we define:

$$T_{k^*}(z) = \sum_{k_2} z^{k^*k_2}Q_{k_2}(z) \bmod P(z)$$

we may write:

$$X^*(k_1^*, k_2^*) = T_{k_1^*k_2^*}(\omega^{k_1^*})$$

or equivalently

$$X^*\left(k_1^*, \frac{k_2^*}{k_1^*}\right) = T_{k_2^*}(\omega^{k_1^*}).$$

For N an odd prime p, all non-zero values of k_1^* are coprime with p so that the $p \times p$-point DFT may be calculated as follows:

(1) form the polynomials

$$T_{k_2^*}(z) = \sum_{k_1}\sum_{k_2} X(k_1,k_2)z^{k_1+k_2^*k_2} \bmod P(z)$$

for $k_2^* = 0, \ldots, p-1$;

(2) evaluate $T_{k_2^*}(\omega^{k_1^*})$ for $k_1^* = 0, \ldots, p-1$;

(3) put $X^*(k_1^*, k_2^*/k_1^*) = T_{k_2^*}(\omega^{k_1^*})$;

(4) calculate the terms for $k_1^* = 0$ separately by

$$X^*(0, k_2^*) = \sum_{k_2}\left[\sum_{k_1} X(k_1,k_2)\right]\omega^{k_2^*k_2}.$$

Step (1) is a set of p 'polynomial transforms' involving no multiplications; step (2) consists of p DFTs on p points each since if

$$T_{k_2^*}(z) = \sum_{k_1} Y_{k_2^*}(k_1)z^{k_1}$$

then

$$T_{k_2^*}(\omega^{k_1^*}) = \sum_{k_1} Y_{k_2^*}(k_1)\omega^{k_1^*k_1} = Y_{k_2^*}^*(k_1^*);$$

step (3) is a permutation; and step (4) is a p-point DFT. Thus the 2D DFT on $p \times p$ points, which takes $2p$ p-point DFTs by the row–column method, involves only $(p+1)$ p-point DFTs; the other DFTs have been replaced by polynomial transforms involving only additions.

This procedure can be extended to n dimensions, and reduces the number of 1D p-point DFTs from np^{n-1} for the row–column method to $(p^n-1)/(p-1)$, at the cost of introducing extra additions in the polynomial transforms.

A similar algorithm has been formulated by Auslander *et al.* (1983) in terms of Galois theory.

1.3.3.3.3. *Global algorithm design*

1.3.3.3.3.1. *From local pieces to global algorithms*

The mathematical analysis of the structure of DFT computations has brought to light a broad variety of possibilities for reducing or reshaping their arithmetic complexity. All of them are 'analytic' in that they break down large transforms into a succession of smaller ones.

These results may now be considered from the converse 'synthetic' viewpoint as providing a list of procedures for assembling them:

(i) the building blocks are one-dimensional p-point algorithms for p a small prime;

(ii) the low-level connectors are the multiplicative reindexing methods of Rader and Winograd, or the polynomial transform reindexing method of Nussbaumer and Quandalle, which allow the construction of efficient algorithms for larger primes p, for prime powers p^ν, and for p-primary pieces of shape $p^\nu \times \ldots \times p^\nu$;

(iii) the high-level connectors are the additive reindexing scheme of Cooley–Tukey, the Chinese remainder theorem reindexing, and the tensor product construction;

(iv) nesting may be viewed as the 'glue' which seals all elements.

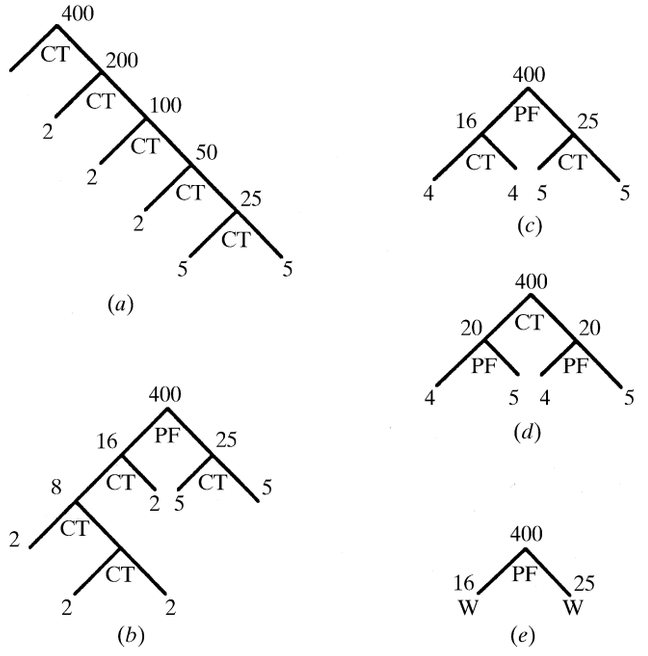

Fig. 1.3.3.1. A few global algorithms for computing a 400-point DFT. CT: Cooley–Tukey factorization. PF: prime factor (or Good) factorization. W: Winograd algorithm.

The simplest DFT may then be carried out into a global algorithm in many different ways. The diagrams in Fig. 1.3.3.1 illustrate a few of the options available to compute a 400-point DFT. They may differ greatly in their arithmetic operation counts.

1.3.3.3.3.2. *Computer architecture considerations*

To obtain a truly useful measure of the computational complexity of a DFT algorithm, its arithmetic operation count must be tempered by computer architecture considerations. Three main types of trade-offs must be borne in mind:

(i) reductions in floating-point (f.p.) arithmetic count are obtained by reindexing, hence at the cost of an increase in integer arithmetic on addresses, although some shortcuts may be found (Uhrich, 1969; Burrus & Eschenbacher, 1981);

(ii) reduction in the f.p. multiplication count usually leads to a large increase in the f.p. addition count (Morris, 1978);

(iii) nesting can increase execution speed, but causes a loss of modularity and hence complicates program development (Silverman, 1977; Kolba & Parks, 1977).

Many of the mathematical developments above took place in the context of single-processor serial computers, where f.p. addition is substantially cheaper than f.p. multiplication but where integer address arithmetic has to compete with f.p. arithmetic for processor cycles. As a result, the alternatives to the Cooley–Tukey algorithm hardly ever led to particularly favourable trade-offs, thus creating the impression that there was little to gain by switching to more exotic algorithms.

The advent of new machine architectures with vector and/or parallel processing features has greatly altered this picture (Pease, 1968; Korn & Lambiotte, 1979; Fornberg, 1981; Swartzrauber, 1984):

(i) *pipelining* equalizes the cost of f.p. addition and f.p. multiplication, and the ideal 'blend' of the two types of operations depends solely on the number of adder and multiplier units available in each machine;

(ii) integer address arithmetic is delegated to specialized arithmetic and logical units (ALUs) operating concurrently with the f.p. units, so that complex reindexing schemes may be used without loss of overall efficiency.

Another major consideration is that of data flow [see *e.g.* Nawab & McClellan (1979)]. Serial machines only have few registers and few paths connecting them, and allow little or no overlap between computation and data movement. New architectures, on the other hand, comprise banks of vector registers (or 'cache memory') besides the usual internal registers, and dedicated ALUs can service data transfers between several of them simultaneously and concurrently with computation.

In this new context, the devices described in Sections 1.3.3.2 and 1.3.3.3 for altering the balance between the various types of arithmetic operations, and reshaping the data flow during the computation, are invaluable. The field of machine-dependent DFT algorithm design is thriving on them [see *e.g.* Temperton (1983*a*,*b*,*c*, 1985); Agarwal & Cooley (1986, 1987)].

1.3.3.3.3.3. *The Johnson–Burrus family of algorithms*

In order to explore systematically all possible algorithms for carrying out a given DFT computation, and to pick the one best suited to a given machine, attempts have been made to develop:

(i) a high-level notation of describing all the ingredients of a DFT computation, including data permutation and data flow;

(ii) a formal calculus capable of operating on these descriptions so as to represent all possible reorganizations of the computation;

(iii) an automatic procedure for evaluating the performance of a given algorithm on a specific architecture.

Task (i) can be accomplished by systematic use of a tensor product notation to represent the various stages into which the DFT can be factored (reindexing, small transforms on subsets of indices, twiddle factors, digit-reversal permutations).

Task (ii) may for instance use the Winograd **CBA** normal form for each small transform, then apply the rules governing the rearrangement of tensor product $\otimes$ and ordinary product $\times$ operations on matrices. The matching of these rearrangements to the architecture of a vector and/or parallel computer can be formalized algebraically [see *e.g.* Chapter 2 of Tolimieri *et al.* (1989)].

Task (iii) is a complex search which requires techniques such as dynamic programming (Bellman, 1958).

Johnson & Burrus (1983) have proposed and tested such a scheme to identify the optimal trade-offs between prime factor nesting and Winograd nesting of small Winograd transforms. In step (ii), they further decomposed the pre-addition matrix **A** and post-addition matrix **C** into several factors, so that the number of design options available becomes very large: the N-point DFT when N has four factors can be calculated in over 10^{12} distinct ways.

This large family of nested algorithms contains the prime factor algorithm and the Winograd algorithms as particular cases, but usually achieves greater efficiency than either by reducing the f.p. multiplication count while keeping the number of f.p. additions small.

There is little doubt that this systematic approach will be extended so as to incorporate all available methods of restructuring the DFT.

1.3.4. Crystallographic applications of Fourier transforms

1.3.4.1. *Introduction*

The central role of the Fourier transformation in X-ray crystallography is a consequence of the kinematic approximation used in the description of the scattering of X-rays by a distribution of electrons (Bragg, 1915; Duane, 1925; Havighurst, 1925*a*,*b*; Zachariasen, 1945; James, 1948*a*, Chapters 1 and 2; Lipson & Cochran, 1953, Chapter 1; Bragg, 1975).

Let $\rho(\mathbf{X})$ be the density of electrons in a sample of matter contained in a finite region V which is being illuminated by a parallel monochromatic X-ray beam with wavevector $\mathbf{K}_0$. Then the far-field amplitude scattered in a direction corresponding to wavevector $\mathbf{K} = \mathbf{K}_0 + \mathbf{H}$ is proportional to

$$\begin{aligned} F(\mathbf{H}) &= \int_V \rho(\mathbf{X}) \exp(2\pi i \mathbf{H}\cdot\mathbf{X})\ \mathrm{d}^3\mathbf{X} \\ &= \bar{\mathscr{F}}[\rho](\mathbf{H}) \\ &= \langle \rho_{\mathbf{x}},\ \exp(2\pi i \mathbf{H}\cdot\mathbf{X})\rangle. \end{aligned}$$

In certain model calculations, the 'sample' may contain not only volume charges, but also point, line and surface charges. These singularities may be accommodated by letting ρ be a distribution, and writing

$$F(\mathbf{H}) = \bar{\mathscr{F}}[\rho](\mathbf{H}) = \langle \rho_{\mathbf{x}},\ \exp(2\pi i \mathbf{H}\cdot\mathbf{X})\rangle.$$

F is still a well behaved function (analytic, by Section 1.3.2.4.2.10) because ρ has been assumed to have compact support.

If the sample is assumed to be an infinite crystal, so that ρ is now a *periodic* distribution, the customary limiting process by which it is shown that F becomes a discrete series of peaks at reciprocal-lattice points (see *e.g.* von Laue, 1936; Ewald, 1940; James, 1948*a* p. 9; Lipson & Taylor, 1958, pp. 14–27; Ewald, 1962, pp. 82–101; Warren, 1969, pp. 27–30) is already subsumed under the treatment of Section 1.3.2.6.

1.3.4.2. *Crystallographic Fourier transform theory*

1.3.4.2.1. *Crystal periodicity*

1.3.4.2.1.1. *Period lattice, reciprocal lattice and structure factors*

Let ρ be the distribution of electrons in a crystal. Then, by definition of a crystal, ρ is Λ-*periodic* for some period lattice Λ (Section 1.3.2.6.5) so that there exists a motif distribution ρ^0 with compact support such that

$$\rho = R * \rho^0,$$

where $R = \sum_{\mathbf{x}\in\Lambda} \delta_{(\mathbf{X})}$. The lattice Λ is usually taken to be the finest for which the above representation holds.

Let Λ have a basis $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ over the integers, these basis vectors being expressed in terms of a standard orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ as

$$\mathbf{a}_k = \sum_{j=1}^{3} a_{jk}\mathbf{e}_j.$$

Then the matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

is the period matrix of Λ (Section 1.3.2.6.5) with respect to the unit lattice with basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, and the volume V of the unit cell is given by $V = |\det \mathbf{A}|$.

By Fourier transformation

$$\bar{\mathscr{F}}[\rho] = R^* \times \bar{\mathscr{F}}[\rho^0],$$

where $R^* = \sum_{\mathbf{H}\in\Lambda^*} \delta_{(\mathbf{H})}$ is the lattice distribution associated to the reciprocal lattice Λ^*. The basis vectors $(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*)$ have coordinates in $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ given by the columns of $(\mathbf{A}^{-1})^T$, whose expression in terms of the cofactors of $\mathbf{A}$ (see Section 1.3.2.6.5) gives the familiar formulae involving the cross product of vectors for $n = 3$. The $\mathbf{H}$-distribution F of scattered amplitudes may be written

$$F = \bar{\mathscr{F}}[\rho]_{\mathbf{H}} = \sum_{\mathbf{H}\in\Lambda^*} \bar{\mathscr{F}}[\rho^0](\mathbf{H})\delta_{(\mathbf{H})} = \sum_{\mathbf{H}\in\Lambda^*} F_{\mathbf{H}}\delta_{(\mathbf{H})}$$

and is thus a weighted reciprocal-lattice distribution, the weight $F_{\mathbf{H}}$ attached to each node $\mathbf{H} \in \Lambda^*$ being the value at $\mathbf{H}$ of the transform $\bar{\mathscr{F}}[\rho^0]$ of the motif ρ^0. Taken in conjunction with the assumption that the scattering is elastic, *i.e.* that $\mathbf{H}$ only changes the direction but not the magnitude of the incident wavevector $\mathbf{K}_0$, this result yields the usual forms (Laue or Bragg) of the diffraction conditions: $\mathbf{H} \in \Lambda^*$, and simultaneously $\mathbf{H}$ lies on the Ewald sphere.

By the reciprocity theorem, ρ^0 can be recovered if F is known for all $\mathbf{H} \in \Lambda^*$ as follows [Section 1.3.2.6.5, *e.g.* (iv)]:

$$\rho_{\mathbf{x}} = \frac{1}{V}\sum_{\mathbf{H}\in\Lambda^*} F_{\mathbf{H}} \exp(-2\pi i \mathbf{H}\cdot\mathbf{X}).$$

These relations may be rewritten in terms of standard, or 'fractional crystallographic', coordinates by putting

$$\mathbf{X} = \mathbf{A}\mathbf{x}, \quad \mathbf{H} = (\mathbf{A}^{-1})^T\mathbf{h},$$

so that a unit cell of the crystal corresponds to $\mathbf{x} \in \mathbb{R}^3/\mathbb{Z}^3$, and that $\mathbf{h} \in \mathbb{Z}^3$. Defining $\mathscr{p}$ and $\mathscr{p}^0$ by

$$\rho = \frac{1}{V}A^{\#}\mathscr{p}, \quad \rho^0 = \frac{1}{V}A^{\#}\mathscr{p}^0$$

so that

$$\rho(\mathbf{X})\ \mathrm{d}^3\mathbf{X} = \mathscr{p}(\mathbf{x})\ \mathrm{d}^3\mathbf{x}, \quad \rho^0(\mathbf{X})\ \mathrm{d}^3\mathbf{X} = \mathscr{p}^0(\mathbf{x})\ \mathrm{d}^3\mathbf{x},$$

we have

$$\begin{aligned} \bar{\mathscr{F}}[\mathscr{p}]_{\mathbf{h}} &= \sum_{\mathbf{h}\in\mathbb{Z}^3} F(\mathbf{h})\delta_{(\mathbf{h})}, \\ F(\mathbf{h}) &= \langle \mathscr{p}^0_{\mathbf{x}},\ \exp(2\pi i \mathbf{h}\cdot\mathbf{x})\rangle \\ &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \mathscr{p}^0(\mathbf{x}) \exp(2\pi i \mathbf{h}\cdot\mathbf{x})\ \mathrm{d}^3\mathbf{x} \quad \text{if } \mathscr{p}^0 \in L^1_{\mathrm{loc}}(\mathbb{R}^3/\mathbb{Z}^3), \\ \mathscr{p}_{\mathbf{x}} &= \sum_{\mathbf{h}\in\mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}). \end{aligned}$$

These formulae are valid for an arbitrary motif distribution $\mathscr{p}^0$, provided the convergence of the Fourier series for $\mathscr{p}$ is considered from the viewpoint of distribution theory (Section 1.3.2.6.10.3).

The experienced crystallographer may notice the absence of the familiar factor $1/V$ from the expression for $\mathscr{p}$ just given. This is because we use the (mathematically) natural unit for $\mathscr{p}$, the *electron per unit cell*, which matches the dimensionless nature of the crystallographic coordinates $\mathbf{x}$ and of the associated volume element $\mathrm{d}^3\mathbf{x}$. The traditional factor $1/V$ was the result of the somewhat inconsistent use of $\mathbf{x}$ as an argument but of $\mathrm{d}^3\mathbf{X}$ as a volume element to obtain ρ in electrons per unit volume (*e.g.* Å^{-3}). A fortunate consequence of the present convention is that nuisance factors of V or $1/V$, which used to abound in convolution or scalar product formulae, are now absent.

It should be noted at this point that the crystallographic terminology regarding $\bar{\mathscr{F}}$ and $\bar{\mathscr{F}}$ differs from the standard mathematical terminology introduced in Section 1.3.2.4.1 and applied to periodic distributions in Section 1.3.2.6.4: F is the *inverse* Fourier transform of ρ rather than its Fourier transform, and the calculation of ρ is called a Fourier *synthesis* in crystallography even though it is mathematically a Fourier analysis. The origin of this discrepancy may be traced to the fact that the mathematical theory of the Fourier transformation originated with the study of *temporal* periodicity, while crystallography deals with *spatial* periodicity; since the expression for the phase factor of a plane wave is $\exp[2\pi i(\nu t - \mathbf{K}\cdot\mathbf{X})]$, the difference in sign between the

contributions from time *versus* spatial displacements makes this conflict unavoidable.

1.3.4.2.1.2. *Structure factors in terms of form factors*

In many cases, ρ^0 is a sum of translates of atomic electron-density distributions. Assume there are n distinct chemical types of atoms, with N_j identical isotropic atoms of type j described by an electron distribution ρ_j about their centre of mass. According to quantum mechanics each ρ_j is a smooth rapidly decreasing function of $\mathbf{x}$, *i.e.* $\rho_j \in \mathscr{S}$, hence $\rho^0 \in \mathscr{S}$ and (ignoring the effect of thermal agitation)

$$\rho^0(\mathbf{x}) = \sum_{j=1}^{n}\left[\sum_{k_j=1}^{N_j}\rho_j(\mathbf{x}-\mathbf{x}_{k_j})\right],$$

which may be written (Section 1.3.2.5.8)

$$\rho^0 = \sum_{j=1}^{n}\left[\rho_j * \left(\sum_{k_j=1}^{N_j}\delta_{(\mathbf{x}_{k_j})}\right)\right].$$

By Fourier transformation:

$$F(\mathbf{h}) = \sum_{j=1}^{n}\left\{\bar{\mathscr{F}}[\rho_j](\mathbf{h}) \times \left[\sum_{k_j=1}^{N_j}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_{k_j})\right]\right\}.$$

Defining the form factor f_j of atom j as a function of $\mathbf{h}$ to be

$$f_j(\mathbf{h}) = \bar{\mathscr{F}}[\rho_j](\mathbf{h})$$

we have

$$F(\mathbf{h}) = \sum_{j=1}^{n} f_j(\mathbf{h}) \times \left[\sum_{k_j=1}^{N_j}\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_{k_j})\right].$$

If $\mathbf{X} = \mathbf{A}\mathbf{x}$ and $\mathbf{H} = (\mathbf{A}^{-1})^T\mathbf{h}$ are the real- and reciprocal-space coordinates in Å and Å^{-1}, and if $\rho_j(\|\mathbf{X}\|)$ is the spherically symmetric electron-density function for atom type j, then

$$f_j(\mathbf{H}) = \int_0^\infty 4\pi\|\mathbf{X}\|^2\rho_j(\|\mathbf{X}\|)\frac{\sin(2\pi\|\mathbf{H}\|\|\mathbf{X}\|)}{2\pi\|\mathbf{H}\|\|\mathbf{X}\|}\,\mathrm{d}\|\mathbf{X}\|.$$

More complex expansions are used for electron-density studies (see Chapter 1.2 in this volume). Anisotropic Gaussian atoms may be dealt with through the formulae given in Section 1.3.2.4.4.2.

1.3.4.2.1.3. *Fourier series for the electron density and its summation*

The convergence of the Fourier series for ρ

$$\rho(\mathbf{x}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} F(\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x})$$

is usually examined from the classical point of view (Section 1.3.2.6.10). The summation of multiple Fourier series meets with considerable difficulties, because there is no natural order in $\mathbb{Z}^n$ to play the role of the natural order in $\mathbb{Z}$ (Ash, 1976). In crystallography, however, the structure factors $F(\mathbf{h})$ are often obtained within spheres $\|\mathbf{H}\| \le \Delta^{-1}$ for increasing resolution (decreasing Δ). Therefore, successive estimates of ρ are most naturally calculated as the corresponding partial sums (Section 1.3.2.6.10.1):

$$S_\Delta(\rho)(\mathbf{x}) = \sum_{\|(\mathbf{A}^{-1})^T\mathbf{h}\|\le\Delta^{-1}} F(\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}).$$

This may be written

$$S_\Delta(\rho)(\mathbf{x}) = (D_\Delta * \rho)(\mathbf{x}),$$

where D_Δ is the 'spherical Dirichlet kernel'

$$D_\Delta(\mathbf{x}) = \sum_{\|(\mathbf{A}^{-1})^T\mathbf{h}\|\le\Delta^{-1}} \exp(-2\pi i\mathbf{h}\cdot\mathbf{x}).$$

D_Δ exhibits numerous negative ripples around its central peak. Thus the 'series termination errors' incurred by using $S_\Delta(\rho)$ instead of ρ consist of negative ripples around each atom, and may lead to a Gibbs-like phenomenon (Section 1.3.2.6.10.1) near a molecular boundary.

As in one dimension, Cesàro sums (arithmetic means of partial sums) have better convergence properties, as they lead to a convolution by a 'spherical Fejér kernel' which is everywhere positive. Thus Cesàro summation will always produce positive approximations to a positive electron density. Other positive summation kernels were investigated by Pepinsky (1952) and by Waser & Schomaker (1953).

1.3.4.2.1.4. *Friedel's law, anomalous scatterers*

If the wavelength λ of the incident X-rays is far from any absorption edge of the atoms in the crystal, there is a constant phase shift in the scattering, and the electron density may be considered to be *real-valued*. Then

$$\begin{aligned} F(\mathbf{h}) &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x})\exp(2\pi i\mathbf{h}\cdot\mathbf{x})\,\mathrm{d}^3\mathbf{x} \\ &= \overline{\int_{\mathbb{R}^3/\mathbb{Z}^3} \overline{\rho(\mathbf{x})}\exp[2\pi i(-\mathbf{h})\cdot\mathbf{x}]\,\mathrm{d}^3\mathbf{x}} \\ &= \overline{F(-\mathbf{h})} \text{ since } \overline{\rho(\mathbf{x})} = \rho(\mathbf{x}). \end{aligned}$$

Thus if

$$F(\mathbf{h}) = |F(\mathbf{h})|\exp(i\varphi(\mathbf{h})),$$

then

$$|F(-\mathbf{h})| = |F(\mathbf{h})| \quad \text{and} \quad \varphi(-\mathbf{h}) = -\varphi(\mathbf{h}).$$

This is Friedel's law (Friedel, 1913). The set $\{F_\mathbf{h}\}$ of Fourier coefficients is said to have *Hermitian symmetry*.

If λ is close to some absorption edge(s), the proximity to resonance induces an extra phase shift, whose effect may be represented by letting $\rho(\mathbf{x})$ take on *complex values*. Let

$$\rho(\mathbf{x}) = \rho^R(\mathbf{x}) + i\rho^I(\mathbf{x})$$

and correspondingly, by termwise Fourier transformation

$$F(\mathbf{h}) = F^R(\mathbf{h}) + iF^I(\mathbf{h}).$$

Since $\rho^R(\mathbf{x})$ and $\rho^I(\mathbf{x})$ are both real, $F^R(\mathbf{h})$ and $F^I(\mathbf{h})$ are both Hermitian symmetric, hence

$$F(-\mathbf{h}) = \overline{F^R(\mathbf{h})} + \overline{iF^I(\mathbf{h})},$$

while

$$\overline{F(\mathbf{h})} = \overline{F^R(\mathbf{h})} - \overline{iF^I(\mathbf{h})}.$$

Thus $F(-\mathbf{h}) \neq \overline{F(\mathbf{h})}$, so that Friedel's law is violated. The components $F^R(\mathbf{h})$ and $F^I(\mathbf{h})$, which do obey Friedel's law, may be expressed as:

$$F^R(\mathbf{h}) = \tfrac{1}{2}[F(\mathbf{h}) + \overline{F(-\mathbf{h})}],$$

$$F^I(\mathbf{h}) = \frac{1}{2i}[F(\mathbf{h}) - \overline{F(-\mathbf{h})}].$$

1.3.4.2.1.5. *Parseval's identity and other L^2 theorems*

By Section 1.3.2.4.3.3 and Section 1.3.2.6.10.2,

$$\sum_{\mathbf{h}\in\mathbb{Z}^3} |F(\mathbf{h})|^2 = \int_{\mathbb{R}^3/\mathbb{Z}^3} |\rho(\mathbf{x})|^2 \, \mathrm{d}^3\mathbf{x} = V \int_{\mathbb{R}^3/\Lambda} |\rho(\mathbf{X})|^2 \, \mathrm{d}^3\mathbf{X}.$$

Usually $\rho(\mathbf{x})$ is real and positive, hence $|\rho(\mathbf{x})| = \rho(\mathbf{x})$, but the identity remains valid even when $\rho(\mathbf{x})$ is made complex-valued by the presence of anomalous scatterers.

If $\{G_\mathbf{h}\}$ is the collection of structure factors belonging to another electron density $\sigma = A^\# \sigma$ with the same period lattice as ρ, then

$$\begin{aligned}\sum_{\mathbf{h}\in\mathbb{Z}^3} \overline{F(\mathbf{h})} G(\mathbf{h}) &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \overline{\rho(\mathbf{x})} \sigma(\mathbf{x}) \, \mathrm{d}^3\mathbf{x} \\ &= V \int_{\mathbb{R}^3/\Lambda} \rho(\mathbf{X}) \sigma(\mathbf{X}) \, \mathrm{d}^3\mathbf{X}.\end{aligned}$$

Thus, norms and inner products may be evaluated either from structure factors or from 'maps'.

1.3.4.2.1.6. *Convolution, correlation and Patterson function*

Let $\rho = r * \rho^0$ and $\sigma = r * \sigma^0$ be two electron densities referred to crystallographic coordinates, with structure factors $\{F_\mathbf{h}\}_{\mathbf{h}\in\mathbb{Z}^3}$ and $\{G_\mathbf{h}\}_{\mathbf{h}\in\mathbb{Z}^3}$, so that

$$\begin{aligned}\rho_\mathbf{x} &= \sum_{\mathbf{h}\in\mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}), \\ \sigma_\mathbf{x} &= \sum_{\mathbf{h}\in\mathbb{Z}^3} G(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}).\end{aligned}$$

The distribution $\omega = r * (\rho^0 * \sigma^0)$ is well defined, since the generalized support condition (Section 1.3.2.3.9.7) is satisfied. The forward version of the convolution theorem implies that if

$$\omega_\mathbf{x} = \sum_{\mathbf{h}\in\mathbb{Z}^3} W(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}),$$

then

$$W(\mathbf{h}) = F(\mathbf{h}) G(\mathbf{h}).$$

If either ρ^0 or σ^0 is infinitely differentiable, then the distribution $\psi = \rho \times \sigma$ exists, and if we analyse it as

$$\psi_\mathbf{x} = \sum_{\mathbf{h}\in\mathbb{Z}^3} Y(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}),$$

then the backward version of the convolution theorem reads:

$$Y(\mathbf{h}) = \sum_{\mathbf{k}\in\mathbb{Z}^3} F(\mathbf{h}) G(\mathbf{h}-\mathbf{k}).$$

The *cross correlation* $\kappa[\rho,\sigma]$ between ρ and σ is the $\mathbb{Z}^3$-periodic distribution defined by:

$$\kappa = \check{\rho}^0 * \sigma.$$

If ρ^0 and σ^0 are locally integrable,

$$\begin{aligned}\kappa[\rho,\sigma](\mathbf{t}) &= \int_{\mathbb{R}^3} \rho^0(\mathbf{x}) \sigma(\mathbf{x}+\mathbf{t}) \, \mathrm{d}^3\mathbf{x} \\ &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x}) \sigma(\mathbf{x}+\mathbf{t}) \, \mathrm{d}^3\mathbf{x}.\end{aligned}$$

Let

$$\kappa(\mathbf{t}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} K(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{t}).$$

The combined use of the shift property and of the forward convolution theorem then gives immediately:

$$K(\mathbf{h}) = \overline{F(\mathbf{h})} G(\mathbf{h});$$

hence the Fourier series representation of $\kappa[\rho,\sigma]$:

$$\kappa[\rho,\sigma](\mathbf{t}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} \overline{F(\mathbf{h})} G(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{t}).$$

Clearly, $\kappa[\rho,\sigma] = (\kappa[\sigma,\rho])\check{}$, as shown by the fact that permuting F and G changes $K(\mathbf{h})$ into its complex conjugate.

The *auto-correlation* of ρ is defined as $\kappa[\rho,\rho]$ and is called the *Patterson function* of ρ. If ρ consists of point atoms, *i.e.*

$$\rho^0 = \sum_{j=1}^N Z_j \delta_{(\mathbf{x}_j)},$$

then

$$\kappa[\rho,\rho] = r * \left[\sum_{j=1}^N \sum_{k=1}^N Z_j Z_k \delta_{(\mathbf{x}_j - \mathbf{x}_k)} \right]$$

contains information about interatomic vectors. It has the Fourier series representation

$$\kappa[\rho,\rho](\mathbf{t}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h}\cdot\mathbf{t}),$$

and is therefore calculable from the diffraction intensities alone. It was first proposed by Patterson (1934, 1935*a*,*b*) as an extension to crystals of the radially averaged correlation function used by Warren & Gingrich (1934) in the study of powders.

1.3.4.2.1.7. *Sampling theorems, continuous transforms, interpolation*

Shannon's sampling and interpolation theorem (Section 1.3.2.7.1) takes two different forms, according to whether the property of finite bandwidth is assumed in real space or in reciprocal space.

(1) The most usual setting is in reciprocal space (see Sayre, 1952*c*). Only a finite number of diffraction intensities can be recorded and phased, and for physical reasons the cutoff criterion is the resolution $\Delta = 1/\|\mathbf{H}\|_{\max}$. Electron-density maps are thus calculated as partial sums (Section 1.3.4.2.1.3), which may be written in Cartesian coordinates as

$$S_\Delta(\rho)(\mathbf{X}) = \sum_{\mathbf{H}\in\Lambda^*,\ \|\mathbf{H}\|\le\Delta^{-1}} F(\mathbf{H}) \exp(-2\pi i \mathbf{H}\cdot\mathbf{X}).$$

$S_\Delta(\rho)$ is band-limited, the support of its spectrum being contained in the solid sphere Σ_Δ defined by $\|\mathbf{H}\| \le \Delta^{-1}$. Let χ_Δ be the indicator function of Σ_Δ. The transform of the normalized version of χ_Δ is (see below, Section 1.3.4.4.3.5)

$$\begin{aligned}I_\Delta(\mathbf{X}) &= \frac{3\Delta^3}{4\pi} \mathscr{F}[\chi_\Delta](\mathbf{X}) \\ &= \frac{3}{u^3}(\sin u - u\cos u) \quad \text{where } u = 2\pi\frac{\|\mathbf{X}\|}{\Delta}.\end{aligned}$$

By Shannon's theorem, it suffices to calculate $S_\Delta(\rho)$ on an integral subdivision Γ of the period lattice Λ such that the sampling criterion is satisfied (*i.e.* that the translates of Σ_Δ by vectors of Γ^* do not overlap). Values of $S_\Delta(\rho)$ may then be calculated at an arbitrary point $\mathbf{X}$ by the interpolation formula:

$$S_\Delta(\rho)(\mathbf{X}) = \sum_{\mathbf{Y}\in\Gamma} I_\Delta(\mathbf{X}-\mathbf{Y}) S_\Delta(\rho)(\mathbf{Y}).$$

(2) The reverse situation occurs whenever the support of the motif ρ^0 does not fill the whole unit cell, *i.e.* whenever there exists a region M (the 'molecular envelope'), strictly smaller than the unit cell, such that the translates of M by vectors of r do not overlap and that

$$\chi_M \times \wp^0 = \wp^0.$$

It then follows that $\wp = r * (\chi_M \times \wp)$. Defining the 'interference function' G as the normalized indicator function of M according to

$$G(\boldsymbol{\eta}) = \frac{1}{\mathrm{vol}(M)} \bar{\mathscr{F}}[\chi_M](\boldsymbol{\eta})$$

we may invoke Shannon's theorem to calculate the value $\bar{\mathscr{F}}[\wp^0](\boldsymbol{\xi})$ at an arbitrary point $\boldsymbol{\xi}$ of reciprocal space from its sample values $F(\mathbf{h}) = \bar{\mathscr{F}}[\wp^0](\mathbf{h})$ at points of the reciprocal lattice as

$$\bar{\mathscr{F}}[\wp^0](\boldsymbol{\xi}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} \mathbf{G}(\boldsymbol{\xi} - \mathbf{h}) F(\mathbf{h}).$$

This aspect of Shannon's theorem constitutes the mathematical basis of phasing methods based on geometric redundancies created by solvent regions and/or noncrystallographic symmetries (Bricogne, 1974). The connection between Shannon's theorem and the phase problem was first noticed by Sayre (1952*b*). He pointed out that the Patterson function of $\wp$, written as $\kappa[\wp, \wp] = r * (\check{\wp}^0 * \wp^0)$, may be viewed as consisting of a motif $\kappa^0 = \check{\wp}^0 * \wp^0$ (containing all the internal interatomic vectors) which is periodized by convolution with r. As the translates of κ^0 by vectors of $\mathbb{Z}^3$ *do* overlap, the sample values of the intensities $|F(\mathbf{h})|^2$ at nodes of the reciprocal lattice do not provide enough data to interpolate intensities $|F(\boldsymbol{\xi})|^2$ at arbitrary points of reciprocal space. Thus the loss of phase is intimately related to the impossibility of intensity interpolation, implying in return that any indication of intensity values attached to non-integral points of the reciprocal lattice is a potential source of phase information.

1.3.4.2.1.8. *Sections and projections*

It was shown at the end of Section 1.3.2.5.8 that the convolution theorem establishes, under appropriate assumptions, a duality between sectioning a smooth function (viewed as a multiplication by a δ-function in the sectioning coordinate) and projecting its transform (viewed as a convolution with the function **1** everywhere equal to 1 as a function of the projection coordinate). This duality follows from the fact that $\mathscr{F}$ and $\bar{\mathscr{F}}$ map $\mathbf{1}_{x_i}$ to δ_{x_i} and δ_{x_i} to $\mathbf{1}_{x_i}$ (Section 1.3.2.5.6), and from the tensor product property (Section 1.3.2.5.5).

In the case of periodic distributions, projection and section must be performed with respect to directions or subspaces which are integral with respect to the period lattice if the result is to be periodic; furthermore, projections must be performed only on the contents of *one* repeating unit along the direction of projection, or else the result would diverge. The same relations then hold between principal central sections and projections of the electron density and the dual principal central projections and sections of the weighted reciprocal lattice, *e.g.*

$$\wp(x_1, 0, 0) \leftrightarrow \sum_{h_1,\, h_2} F(h_1, h_2, h_3),$$

$$\wp(x_1, x_2, 0) \leftrightarrow \sum_{h_3} F(h_1, h_2, h_3),$$

$$\wp_{1,\,2}(x_3) = \int_{\mathbb{R}^2/\mathbb{Z}^2} \wp(x_1, x_2, x_3)\, \mathrm{d}x_1\, \mathrm{d}x_2 \leftrightarrow F(0, 0, h_3),$$

$$\wp_1(x_2, x_3) = \int_{\mathbb{R}/\mathbb{Z}} \wp(x_1, x_2, x_3)\, \mathrm{d}x_1 \qquad \leftrightarrow F(0, h_2, h_3)$$

etc.

When the sections are principal but not central, it suffices to use the shift property of Section 1.3.2.5.5. When the sections or projections are not principal, they can be made principal by changing to new primitive bases B and B^* for Λ and Λ^*, respectively, the transition matrices $\mathbf{P}$ and $\mathbf{P}^*$ to these new bases being related by $\mathbf{P}^* = (\mathbf{P}^{-1})^T$ in order to preserve duality. This change of basis must be such that one of these matrices (say, $\mathbf{P}$) should have a given integer vector $\mathbf{u}$ as its first column, $\mathbf{u}$ being related to the line or plane defining the section or projection of interest.

The problem of constructing a matrix $\mathbf{P}$ given $\mathbf{u}$ received an erroneous solution in Volume II of *International Tables* (Patterson, 1959), which was subsequently corrected in 1962. Unfortunately, the solution proposed there is complicated and does not suggest a general approach to the problem. It therefore seems worthwhile to record here an effective procedure which solves this problem in any dimension n (Watson, 1970).

Let

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

be a primitive integral vector, *i.e.* g.c.d. $(u_1, \ldots, u_n) = 1$. Then an $n \times n$ integral matrix $\mathbf{P}$ with $\det \mathbf{P} = 1$ having $\mathbf{u}$ as its first column can be constructed by induction as follows. For $n = 1$ the result is trivial. For $n = 2$ it can be solved by means of the Euclidean algorithm, which yields z_1, z_2 such that $u_1 z_2 - u_2 z_1 = 1$, so that we may take $\mathbf{P} = \begin{pmatrix} u_1 & z_1 \\ u_2 & z_2 \end{pmatrix}$. Note that, if $\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ is a solution, then $\mathbf{z} + m\mathbf{u}$ is another solution for any $m \in \mathbb{Z}$. For $n \geq 3$, write $\mathbf{u} = \begin{pmatrix} u_1 \\ d\mathbf{z} \end{pmatrix}$ with $d =$ g.c.d. $(u_2, \ldots, u_n)$ so that both $\mathbf{z} = \begin{pmatrix} z_2 \\ \vdots \\ z_n \end{pmatrix}$ and $\begin{pmatrix} u_1 \\ d \end{pmatrix}$ are primitive. By the inductive hypothesis there is an integral 2×2 matrix $\mathbf{V}$ with $\begin{pmatrix} u_1 \\ d \end{pmatrix}$ as its first column, and an integral $(n-1) \times (n-1)$ matrix $\mathbf{Z}$ with $\mathbf{z}$ as its first column, with $\det \mathbf{V} = 1$ and $\det \mathbf{Z} = 1$.

Now put

$$\mathbf{P} = \begin{pmatrix} 1 & \\ & \mathbf{Z} \end{pmatrix} \begin{pmatrix} \mathbf{V} & \\ & \mathbf{I}_{n-2} \end{pmatrix},$$

i.e.

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & . & 0 \\ 0 & z_2 & * & . & * \\ 0 & z_3 & * & . & * \\ . & . & . & . & . \\ 0 & z_n & * & . & * \end{pmatrix} \begin{pmatrix} u_1 & * & 0 & . & 0 \\ d & * & 0 & . & 0 \\ 0 & 0 & 1 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & . & 1 \end{pmatrix}.$$

The first column of $\mathbf{P}$ is

$$\begin{pmatrix} u_1 \\ dz_2 \\ . \\ . \\ dz_n \end{pmatrix} = \mathbf{u},$$

and its determinant is 1, QED.

The incremental step from dimension $n-1$ to dimension n is the construction of 2×2 matrix $\mathbf{V}$, for which there exist infinitely many solutions labelled by an integer m_{n-1}. Therefore, the collection of matrices $\mathbf{P}$ which solve the problem is labelled by $n-1$ arbitrary integers $(m_1, m_2, \ldots, m_{n-1})$. This freedom can be used to adjust the shape of the basis B.

Once **P** has been chosen, the calculation of general sections and projections is transformed into that of *principal* sections and projections by the changes of coordinates:

$$\mathbf{x} = \mathbf{P}\mathbf{x}', \qquad \mathbf{h} = \mathbf{P}^*\mathbf{h}',$$

and an appeal to the tensor product property.

Booth (1945*a*) made use of the convolution theorem to form the Fourier coefficients of 'bounded projections', which provided a compromise between 2D and 3D Fourier syntheses. If it is desired to compute the projection on the (x, y) plane of the electron density lying between the planes $z = z_1$ and $z = z_2$, which may be written as

$$[\rho \times (\mathbf{1}_x \otimes \mathbf{1}_y \otimes \chi_{[z_1,\, z_2]})] * (\delta_x \otimes \delta_y \otimes \mathbf{1}_z).$$

The transform is then

$$[F * (\delta_h \otimes \delta_k \otimes \bar{\mathscr{F}}[\chi_{[z_1,\, z_2]}])] \times (\mathbf{1}_h \otimes \mathbf{1}_k \otimes \delta_l),$$

giving for coefficient (h, k):

$$\sum_{l\in\mathbb{Z}} F(h,k,l) \exp\{2\pi i l[(z_1+z_2)/2]\} \times \frac{\sin \pi l(z_1 - z_2)}{\pi l}.$$

1.3.4.2.1.9. *Differential syntheses*

Another particular instance of the convolution theorem is the duality between differentiation and multiplication by a monomial (Sections 1.3.2.4.2.8, 1.3.2.5.8).

In the present context, this result may be written

$$\bar{\mathscr{F}}\left[\frac{\partial^{m_1+m_2+m_3}\rho}{\partial X_1^{m_1}\partial X_2^{m_2}\partial X_3^{m_3}}\right](\mathbf{H}) = (-2\pi i)^{m_1+m_2+m_3} H_1^{m_1} H_2^{m_2} H_3^{m_3} F(\mathbf{A}^T\mathbf{H})$$

in Cartesian coordinates, and

$$\bar{\mathscr{F}}\left[\frac{\partial^{m_1+m_2+m_3}\rho}{\partial x_1^{m_1}\partial x_2^{m_2}\partial x_3^{m_3}}\right](\mathbf{h}) = (-2\pi i)^{m_1+m_2+m_3} h_1^{m_1} h_2^{m_2} h_3^{m_3} F(\mathbf{h})$$

in crystallographic coordinates.

A particular case of the first formula is

$$-4\pi^2 \sum_{\mathbf{H}\in\Lambda^*} \|\mathbf{H}\|^2 F(\mathbf{A}^T\mathbf{H}) \exp(-2\pi i \mathbf{H}\cdot\mathbf{X}) = \Delta\rho(\mathbf{X}),$$

where

$$\Delta\rho = \sum_{j=1}^{3} \frac{\partial^2\rho}{\partial X_j^2}$$

is the Laplacian of ρ.

The second formula has been used with $|\mathbf{m}| = 1$ or 2 to compute 'differential syntheses' and refine the location of maxima (or other stationary points) in electron-density maps. Indeed, the values at **x** of the gradient vector $\nabla\rho$ and Hessian matrix $(\nabla\nabla^T)\rho$ are readily obtained as

$$(\nabla\rho)(\mathbf{x}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} (-2\pi i\mathbf{h}) F(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}),$$

$$[(\nabla\nabla^T)\rho](\mathbf{x}) = \sum_{\mathbf{h}\in\mathbb{Z}^3} (-4\pi^2 \mathbf{h}\mathbf{h}^T) F(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}),$$

and a step of Newton iteration towards the nearest stationary point of ρ will proceed by

$$\mathbf{x} \longmapsto \mathbf{x} - \{[(\nabla\nabla^T)\rho](\mathbf{x})\}^{-1} (\nabla\rho)(\mathbf{x}).$$

The modern use of Fourier transforms to speed up the computation of derivatives for model refinement will be described in Section 1.3.4.4.7.

The converse property is also useful: it relates the derivatives of the continuous transform $\bar{\mathscr{F}}[\rho^0]$ to the moments of ρ^0:

$$\frac{\partial^{m_1+m_2+m_3}\bar{\mathscr{F}}[\rho^0]}{\partial X_1^{m_1}\partial X_2^{m_2}\partial X_3^{m_3}}(\mathbf{H}) = \bar{\mathscr{F}}[(2\pi i)^{m_1+m_2+m_3} X_1^{m_1} X_2^{m_2} X_3^{m_3} \rho^0_{\mathbf{x}}](\mathbf{H}).$$

For $|\mathbf{m}| = 2$ and $\mathbf{H} = \mathbf{0}$, this identity gives the well known relation between the Hessian matrix of the transform $\bar{\mathscr{F}}[\rho^0]$ at the origin of reciprocal space and the inertia tensor of the motif ρ^0. This is a particular case of the moment-generating properties of $\bar{\mathscr{F}}$, which will be further developed in Section 1.3.4.5.2.

1.3.4.2.1.10. *Toeplitz forms, determinantal inequalities and Szegö's theorem*

The classical results presented in Section 1.3.2.6.9 can be readily generalized to the case of triple Fourier series; no new concept is needed, only an obvious extension of the notation.

Let ρ be real-valued, so that Friedel's law holds and $F(-\mathbf{h}) = \overline{F(\mathbf{h})}$. Let H be a finite set of indices comprising the origin: $\mathsf{H} = \{\mathbf{h}_0 = \mathbf{0}, \mathbf{h}_1, \ldots, \mathbf{h}_n\}$. Then the Hermitian form in $n+1$ complex variables

$$T_{\mathsf{H}}[\rho](\mathbf{u}) = \sum_{j,\ k=0}^{n} F(\mathbf{h}_j - \mathbf{h}_k)\overline{u}_j u_k$$

is called the Toeplitz form of order H associated to ρ. By the convolution theorem and Parseval's identity,

$$T_{\mathsf{H}}[\rho](\mathbf{u}) = \int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x}) \left| \sum_{j=0}^{n} u_j \exp(2\pi i \mathbf{h}_j\cdot\mathbf{x}) \right|^2 \mathrm{d}^3\mathbf{x}.$$

If ρ is almost everywhere non-negative, then for all H the forms $T_{\mathsf{H}}[\rho]$ are positive semi-definite and therefore all Toeplitz determinants $D_{\mathsf{H}}[\rho]$ are non-negative, where

$$D_{\mathsf{H}}[\rho] = \det\{[F(\mathbf{h}_j - \mathbf{h}_k)]\}.$$

The Toeplitz–Carathéodory–Herglotz theorem given in Section 1.3.2.6.9.2 states that the converse is true: if $D_{\mathsf{H}}[\rho] \geq 0$ for all H, then ρ is almost everywhere non-negative. This result is known in the crystallographic literature through the papers of Karle & Hauptman (1950), MacGillavry (1950), and Goedkoop (1950), following previous work by Harker & Kasper (1948) and Gillis (1948*a*,*b*).

Szegö's study of the asymptotic distribution of the eigenvalues of Toeplitz forms as their order tends to infinity remains valid. Some precautions are needed, however, to define the notion of a sequence (H_k) of finite subsets of indices tending to infinity: it suffices that the H_k should consist essentially of the reciprocal-lattice points **h** contained within a domain of the form $k\Omega$ (k-fold dilation of Ω) where Ω is a convex domain in $\mathbb{R}^3$ containing the origin (Widom, 1960). Under these circumstances, the eigenvalues $\lambda_\nu^{(n)}$ of the Toeplitz forms $T_{\mathsf{H}_k}[\rho]$ become equidistributed with the sample values $\rho_{\nu'}^{(n)}$ of ρ on a grid satisfying the Shannon sampling criterion for the data in H_k (*cf.* Section 1.3.2.6.9.3).

A particular consequence of this equidistribution is that the geometric means of the $\lambda_\nu^{(n)}$ and of the $\rho_{\nu'}^{(n)}$ are equal, and hence as in Section 1.3.2.6.9.4

$$\lim_{k\to\infty} \{D_{\mathsf{H}_k}[\rho]\}^{1/|\mathsf{H}_k|} = \exp\left\{ \int_{\mathbb{R}^3/\mathbb{Z}^3} \log\rho(\mathbf{x})\ \mathrm{d}^3\mathbf{x} \right\},$$

where $|\mathsf{H}_k|$ denotes the number of reflections in H_k. Complementary terms giving a better comparison of the two sides were obtained by Widom (1960, 1975) and Linnik (1975).

This formula played an important role in the solution of the 2D Ising model by Onsager (1944) (see Montroll *et al.*, 1963). It is also encountered in phasing methods involving the 'Burg entropy' (Britten & Collins, 1982; Narayan & Nityananda, 1982; Bricogne, 1982, 1984, 1988).

1.3.4.2.2. *Crystal symmetry*

1.3.4.2.2.1. *Crystallographic groups*

The description of a crystal given so far has dealt only with its invariance under the action of the (discrete Abelian) group of translations by vectors of its period lattice Λ.

Let the crystal now be embedded in Euclidean 3-space, so that it may be acted upon by the group $M(3)$ of rigid (*i.e.* distance-preserving) motions of that space. The group $M(3)$ contains a normal subgroup $T(3)$ of translations, and the quotient group $M(3)/T(3)$ may be identified with the 3-dimensional orthogonal group $O(3)$. The period lattice Λ of a crystal is a discrete uniform subgroup of $T(3)$.

The possible invariance properties of a crystal under the action of $M(3)$ are captured by the following definition: a *crystallographic group* is a subgroup Γ of $M(3)$ if

(i) $\Gamma \cap T(3) = \Lambda$, a period lattice and a normal subgroup of Γ;

(ii) the factor group $G = \Gamma/\Lambda$ is finite.

The two properties are not independent: by a theorem of Bieberbach (1911), they follow from the assumption that Λ is a discrete subgroup of $M(3)$ which operates without accumulation point and with a compact fundamental domain (see Auslander, 1965). These two assumptions imply that G acts on Λ through an integral representation, and this observation leads to a complete enumeration of all distinct Γ's. The mathematical theory of these groups is still an active research topic (see, for instance, Farkas, 1981), and has applications to Riemannian geometry (Wolf, 1967).

This classification of crystallographic groups is described elsewhere in these *Tables* (Wondratschek, 1995), but it will be surveyed briefly in Section 1.3.4.2.2.3 for the purpose of establishing further terminology and notation, after recalling basic notions and results concerning groups and group actions in Section 1.3.4.2.2.2.

1.3.4.2.2.2. *Groups and group actions*

The books by Hall (1959) and Scott (1964) are recommended as reference works on group theory.

(*a*) *Left and right actions*

Let G be a group with identity element e, and let X be a set. An *action* of G on X is a mapping from $G \times X$ to X with the property that, if $g\ x$ denotes the image of (g,x), then

(i) $(g_1g_2)x = g_1(g_2x)$ for all $g_1, g_2 \in G$ and all $x \in X$,

(ii) $ex = x$ for all $x \in X$.

An element g of G thus induces a mapping T_g of X into itself defined by $T_g(x) = gx$, with the 'representation property':

(iii) $T_{g_1g_2} = T_{g_1}T_{g_2}$ for all $g_1, g_2 \in G$.

Since G is a group, every g has an inverse g^{-1}; hence every mapping T_g has an inverse $T_{g^{-1}}$, so that each T_g is a permutation of X.

Strictly speaking, what has just been defined is a *left* action. A *right* action of G on X is defined similarly as a mapping $(g,x) \longmapsto xg$ such that

(i′) $x(g_1g_2) = (xg_1)g_2$ for all $g_1, g_2 \in G$ and all $x \in X$,

(ii′) $xe = x$ for all $x \in X$.

The mapping T'_g defined by $T'_g(x) = xg$ then has the 'right-representation' property:

(iii′) $T'_{g_1g_2} = T'_{g_2}T'_{g_1}$ for all $g_1, g_2 \in G$.

The essential difference between left and right actions is of course not whether the elements of G are written on the left or right of those of X: it lies in the difference between (iii) and (iii′). In a left action the product g_1g_2 in G operates on $x \in X$ by g_2 operating first, then g_1 operating on the result; in a right action, g_1 operates first, then g_2. This distinction will be of importance in Sections 1.3.4.2.2.4 and 1.3.4.2.2.5. In the sequel, we will use left actions unless otherwise stated.

(*b*) *Orbits and isotropy subgroups*

Let x be a fixed element of X. Two fundamental entities are associated to x:

(1) the subset of G consisting of all g such that $gx = x$ is a subgroup of G, called the *isotropy subgroup* of x and denoted G_x;

(2) the subset of X consisting of all elements gx with g running through G is called the *orbit* of x under G and is denoted Gx.

Through these definitions, the action of G on X can be related to the internal structure of G, as follows. Let G/G_x denote the collection of distinct left cosets of G_x in G, *i.e.* of distinct subsets of G of the form gG_x. Let $|G|, |G_x|, |Gx|$ and $|G/G_x|$ denote the numbers of elements in the corresponding sets. The number $|G/G_x|$ of distinct cosets of G_x in G is also denoted $[G : G_x]$ and is called the *index* of G_x in G; by Lagrange's theorem

$$[G : G_x] = |G/G_x| = \frac{|G|}{|G_x|}.$$

Now if g_1 and g_2 are in the same coset of G_x, then $g_2 = g_1g'$ with $g' \in G_x$, and hence $g_1x = g_2x$; the converse is obviously true. Therefore, the mapping from cosets to orbit elements

$$gG_x \longmapsto gx$$

establishes a one-to-one correspondence between the distinct left cosets of G_x in G and the elements of the orbit of x under G. It follows that the number of distinct elements in the orbit of x is equal to the index of G_x in G:

$$|Gx| = [G : G_x] = \frac{|G|}{|G_x|},$$

and that the elements of the orbit of x may be listed without repetition in the form

$$Gx = \{\gamma x | \gamma \in G/G_x\}.$$

Similar definitions may be given for a right action of G on X. The set of distinct right cosets G_xg in G, denoted $G_x\backslash G$, is then in one-to-one correspondence with the distinct elements in the orbit xG of x.

(*c*) *Fundamental domain and orbit decomposition*

The group properties of G imply that two orbits under G are either disjoint or equal. The set X may thus be written as the *disjoint* union

$$X = \bigcup_{i \in I} Gx_i,$$

where the x_i are elements of distinct orbits and I is an indexing set labelling them. The subset $D = \{x_i\}_{i \in I}$ is said to constitute a *fundamental domain* (mathematical terminology) or an *asymmetric unit* (crystallographic terminology) for the action of G on X: it contains one representative x_i of each distinct orbit. Clearly, other fundamental domains may be obtained by choosing different representatives for these orbits.

If X is finite and if f is an arbitrary complex-valued function over X, the 'integral' of f over X may be written as a sum of integrals over the distinct orbits, yielding the *orbit decomposition formula*:

$$\sum_{x\in X} f(x) = \sum_{i\in I}\left(\sum_{y_i\in Gx_i} f(y_i)\right) = \sum_{i\in I}\left(\sum_{\gamma_i\in G/G_{x_i}} f(\gamma_i x_i)\right)$$
$$= \sum_{i\in I}\frac{1}{|G_{x_i}|}\left(\sum_{g_i\in G} f(g_i x_i)\right).$$

In particular, taking $f(x) = 1$ for all x and denoting by $|X|$ the number of elements of X:

$$|X| = \sum_{i\in I}|Gx_i| = \sum_{i\in I}|G/G_{x_i}| = \sum_{i\in I}\frac{|G|}{|G_{x_i}|}.$$

(*d*) *Conjugation, normal subgroups, semi-direct products*

A group G acts on itself by *conjugation*, *i.e.* by associating to $g\in G$ the mapping C_g defined by

$$C_g(h) = ghg^{-1}.$$

Indeed, $C_g(hk) = C_g(h)C_g(k)$ and $[C_g(h)]^{-1} = C_{g^{-1}}(h)$. In particular, C_g operates on the set of subgroups of G, two subgroups H and K being called conjugate if $H = C_g(K)$ for some $g\in G$; for example, it is easily checked that $G_{gx} = C_g(G_x)$. The orbits under this action are the *conjugacy classes* of subgroups of G, and the isotropy subgroup of H under this action is called the *normalizer* of H in G.

If $\{H\}$ is a one-element orbit, H is called a *self-conjugate* or *normal* subgroup of G; the cosets of H in G then form a group G/H called the *factor group* of G by H.

Let G and H be two groups, and suppose that G acts on H by *automorphisms* of H, *i.e.* in such a way that

$$g(h_1h_2) = g(h_1)g(h_2)$$
$$g(e_H) = e_H \qquad \text{(where } e_H \text{ is the identity element of } H\text{)}.$$
$$g(h^{-1}) = (g(h))^{-1}$$

Then the symbols $[g,h]$ with $g\in G$, $h\in H$ form a group K under the product rule:

$$[g_1,h_1][g_2,h_2] = [g_1g_2, h_1g_1(h_2)]$$

{associativity checks; $[e_G,e_H]$ is the identity; $[g,h]$ has inverse $[g^{-1},g^{-1}(h^{-1})]$}. The group K is called the *semi-direct product* of H by G, denoted $K = H\rtimes G$.

The elements $[g,e_H]$ form a subgroup of K isomorphic to G, the elements $[e_G,h]$ form a normal subgroup of K isomorphic to H, and the action of G on H may be represented as an action by conjugation in the sense that

$$C_{[g,\,e_H]}([e_G,h]) = [e_G, g(h)].$$

A familiar example of semi-direct product is provided by the group of Euclidean motions $M(3)$ (Section 1.3.4.2.2.1). An element S of $M(3)$ may be written $S = [R,t]$ with $R\in O(3)$, the orthogonal group, and $t\in T(3)$, the translation group, and the product law

$$[R_1,t_1][R_2,t_2] = [R_1R_2, t_1 + R_1(t_2)]$$

shows that $M(3) = T(3)\rtimes O(3)$ with $O(3)$ acting on $T(3)$ by rotating the translation vectors.

(*e*) *Associated actions in function spaces*

For every left action T_g of G in X, there is an associated left action $T_g^{\#}$ of G on the space $L(X)$ of complex-valued functions over X, defined by 'change of variable' (Section 1.3.2.3.9.5):

$$[T_g^{\#}f](x) = f((T_g)^{-1}x) = f(g^{-1}x).$$

Indeed for any g_1, g_2 in G,

$$[T_{g_1}^{\#}[T_{g_2}^{\#}f]](x) = [T_{g_2}^{\#}f]((T_{g_1})^{-1}x) = f[T_{g_2}^{-1}T_{g_1}^{-1}x]$$
$$= f((T_{g_1}T_{g_2})^{-1}x);$$

since $T_{g_1}T_{g_2} = T_{g_1g_2}$, it follows that

$$T_{g_1}^{\#}T_{g_2}^{\#} = T_{g_1g_2}^{\#}.$$

It is clear that the change of variable must involve the action of g^{-1} (not g) if $T^{\#}$ is to define a *left* action; using g instead would yield a *right* action.

The linear representation operators $T_g^{\#}$ on $L(X)$ provide the most natural instrument for stating and exploiting symmetry properties which a function may possess with respect to the action of G. Thus a function $f\in L(X)$ will be called *G-invariant* if $f(gx) = f(x)$ for all $g\in G$ and all $x\in X$. The value $f(x)$ then depends on x only through its orbit Gx, and f is uniquely defined once it is specified on a fundamental domain $D = \{x_i\}_{i\in I}$; its integral over X is then a weighted sum of its values in D:

$$\sum_{x\in X} f(x) = \sum_{i\in I}[G : G_{x_i}]f(x_i).$$

The G-invariance of f may be written:

$$T_g^{\#}f = f \quad \text{for all } g\in G.$$

Thus f is invariant under each $T_g^{\#}$, which obviously implies that f is invariant under the linear operator in $L(X)$

$$A_G = \frac{1}{|G|}\sum_{g\in G} T_g^{\#},$$

which averages an arbitrary function by the action of G. Conversely, if $A_Gf = f$, then

$$T_{g_0}^{\#}f = T_{g_0}^{\#}(A_Gf) = (T_{g_0}^{\#}A_G)f = A_Gf = f \quad \text{for all } g_0\in G,$$

so that the two statements of the G-invariance of f are equivalent. The identity

$$T_{g_0}^{\#}A_G = A_G \quad \text{for all } g_0\in G$$

is easily proved by observing that the map $g\longmapsto g_0g$ (g_0 being any element of G) is a one-to-one map from G into itself, so that

$$\sum_{g\in G} T_g^{\#} = \sum_{g\in G} T_{g_0g}^{\#}$$

as these sums differ only by the order of the terms. The same identity implies that A_G is a *projector*:

$$(A_G)^2 = A_G,$$

and hence that its eigenvalues are either 0 or 1. In summary, we may say that the invariance of f under G is equivalent to f being an eigenfunction of the associated projector A_G for eigenvalue 1.

(*f*) *Orbit exchange*

One final result about group actions which will be used repeatedly later is concerned with the case when X has the structure of a Cartesian product:

$$X = X_1\times X_2\times\ldots\times X_n$$

and when G acts *diagonally* on X, *i.e.* acts on each X_j separately:

$$gx = g(x_1,x_2,\ldots,x_n) = (gx_1, gx_2,\ldots,gx_n).$$

Then complete sets (but not usually minimal sets) of representatives

of the distinct orbits for the action of G in X may be obtained in the form

$$D_k = X_1 \times \ldots \times X_{k-1} \times \{x_{i_k}^{(k)}\}_{i_k \in I_k} \times X_{k+1} \times \ldots \times X_n$$

for each $k = 1, 2, \ldots, n$, *i.e.* by taking a fundamental domain in X_k and all the elements in X_j with $j \neq k$. The action of G on each D_k does indeed generate the whole of X: given an arbitrary element $y = (y_1, y_2, \ldots, y_n)$ of X, there is an index $i_k \in I_k$ such that $y_k \in G x_{i_k}^{(k)}$ and a coset of $G_{x_{i_k}^{(k)}}$ in G such that $y_k = \gamma x_{i_k}^{(k)}$ for any representative γ of that coset; then

$$y = \gamma(\gamma^{-1}y_1, \ldots, \gamma^{-1}y_{k-1}, x_{i_k}^{(k)}, \gamma^{-1}y_{k+1}, \ldots, \gamma^{-1}y_n)$$

which is of the form $y = \gamma d_k$ with $d_k \in D_k$.

The various D_k are related in a simple manner by 'transposition' or 'orbit exchange' (the latter name is due to J. W. Cooley). For instance, D_j may be obtained from $D_k (\, j \neq k)$ as follows: for each $y_j \in X_j$ there exists $g(y_j) \in G$ and $i_j(y_j) \in I_j$ such that $y_j = g(y_j) x_{i_j(y_j)}^{(j)}$; therefore

$$D_j = \bigcup_{y_j \in X_j} [g(y_j)]^{-1} D_k,$$

since the fundamental domain of X_k is thus expanded to the whole of X_k, while X_j is reduced to its fundamental domain. In other words: orbits are simultaneously collapsed in the jth factor and expanded in the kth.

When G operates without fixed points in each X_k (*i.e.* $G_{x_k} = \{e\}$ for all $x_k \in X_k$), then each D_k is a fundamental domain for the action of G in X. The existence of fixed points in some or all of the X_k complicates the situation in that for each k and each $x_k \in X_k$ such that $G_{x_k} \neq \{e\}$ the action of G/G_{x_k} on the other factors must be examined. Shenefelt (1988) has made a systematic study of orbit exchange for space group $P622$ and its subgroups.

Orbit exchange will be encountered, in a great diversity of forms, as the basic mechanism by which intermediate results may be rearranged between the successive stages of the computation of crystallographic Fourier transforms (Section 1.3.4.3).

1.3.4.2.2.3. *Classification of crystallographic groups*

Let Γ be a crystallographic group, Λ the normal subgroup of its lattice translations, and G the finite factor group Γ/Λ. Then G acts on Λ by conjugation [Section 1.3.4.2.2.2(*d*)] and this action, being a mapping of a lattice into itself, is representable by matrices with integer entries.

The classification of crystallographic groups proceeds from this observation in the following three steps:

Step 1: find all possible finite abstract groups G which can be represented by 3×3 integer matrices.

Step 2: for each such G find all its inequivalent representations by 3×3 integer matrices, equivalence being defined by a change of primitive lattice basis (*i.e.* conjugation by a 3×3 integer matrix with determinant ± 1).

Step 3: for each G and each equivalence class of integral representations of G, find all inequivalent extensions of the action of G from Λ to $T(3)$, equivalence being defined by an affine coordinate change [*i.e.* conjugation by an element of $A(3)$].

Step 1 leads to the following groups, listed in association with the crystal system to which they later give rise:

$\mathbb{Z}/2\mathbb{Z}$	monoclinic
$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	orthorhombic
$\mathbb{Z}/3\mathbb{Z}, (\mathbb{Z}/3\mathbb{Z}) \rtimes \{\alpha\}$	trigonal
$\mathbb{Z}/4\mathbb{Z}, (\mathbb{Z}/4\mathbb{Z}) \rtimes \{\alpha\}$	tetragonal
$\mathbb{Z}/6\mathbb{Z}, (\mathbb{Z}/6\mathbb{Z}) \rtimes \{\alpha\}$	hexagonal
$(\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}) \rtimes \{S_3\}$	cubic

and the extension of these groups by a centre of inversion. In this list $\rtimes$ denotes a semi-direct product [Section 1.3.4.2.2.2(*d*)], α denotes the automorphism $g \longmapsto g^{-1}$, and S_3 (the group of permutations on three letters) operates by permuting the copies of $\mathbb{Z}/2\mathbb{Z}$ (using the subgroup A_3 of cyclic permutations gives the tetrahedral subsystem).

Step 2 leads to a list of 73 equivalence classes called *arithmetic classes* of representations $g \longmapsto \mathbf{R}_g$, where $\mathbf{R}_g$ is a 3×3 integer matrix, with $\mathbf{R}_{g_1 g_2} = \mathbf{R}_{g_1}\mathbf{R}_{g_2}$ and $\mathbf{R}_e = \mathbf{I}_3$. This enumeration is more familiar if equivalence is relaxed so as to allow conjugation by *rational* 3×3 matrices with determinant ± 1: this leads to the 32 crystal classes. The difference between an arithmetic class and its rational class resides in the choice of a lattice mode (P, $A/B/C$, I, F or R). Arithmetic classes always refer to a primitive lattice, but may use inequivalent integral representations for a given geometric symmetry element; while crystallographers prefer to change over to a non-primitive lattice, if necessary, in order to preserve the same integral representation for a given geometric symmetry element. The matrices $\mathbf{P}$ and $\mathbf{Q} = \mathbf{P}^{-1}$ describing the changes of basis between primitive and centred lattices are listed in Table 5.1 and illustrated in Figs. 5.3 to 5.9, pp. 76–79, of Volume A of *International Tables* (Arnold, 1995).

Step 3 gives rise to a system of congruences for the systems of non-primitive translations $\{\mathbf{t}_g\}_{g \in G}$ which may be associated to the matrices $\{\mathbf{R}_g\}_{g \in G}$ of a given arithmetic class, namely:

$$\mathbf{t}_{g_1 g_2} \equiv \mathbf{R}_{g_1}\mathbf{t}_{g_2} + \mathbf{t}_{g_1} \bmod \Lambda,$$

first derived by Frobenius (1911). If equivalence under the action of $A(3)$ is taken into account, 219 classes are found. If equivalence is defined with respect to the action of the subgroup $A^+(3)$ of $A(3)$ consisting only of transformations with determinant +1, then 230 classes called *space-group types* are obtained. In particular, associating to each of the 73 arithmetic classes a trivial set of non-primitive translations ($\mathbf{t}_g = \mathbf{0}$ for all $g \in G$) yields the 73 symmorphic space groups. This third step may also be treated as an abstract problem concerning group extensions, using cohomological methods [Ascher & Janner (1965); see Janssen (1973) for a summary]; the connection with Frobenius's approach, as generalized by Zassenhaus (1948), is examined in Ascher & Janner (1968).

The finiteness of the number of space-group types in dimension 3 was shown by Bieberbach (1912) to be the case in arbitrary dimension. The reader interested in N-dimensional space-group theory for $N > 3$ may consult Brown (1969), Brown *et al.* (1978), Schwarzenberger (1980), and Engel (1986). The standard reference for integral representation theory is Curtis & Reiner (1962).

All three-dimensional space groups G have the property of being *solvable*, *i.e.* that there exists a chain of subgroups

$$G = G_r > G_{r-1} > \ldots > G_1 > G_0 = \{e\},$$

where each G_{i-1} is a normal subgroup of G_1 and the factor group G_i/G_{i-1} is a *cyclic* group of some order m_i $(1 \leq i \leq r)$. This property may be established by inspection, or deduced from a famous theorem of Burnside [see Burnside (1911), pp. 322–323] according to which any group G such that $|G| = p^\alpha q^\beta$, with p and q distinct primes, is solvable; in the case at hand, $p = 2$ and $q = 3$.

The whole classification of 3D space groups can be performed swiftly by a judicious use of the solvability property (L. Auslander, personal communication).

Solvability facilitates the indexing of elements of G in terms of generators and relations (Coxeter & Moser, 1972; Magnus *et al.*, 1976) for the purpose of calculation. By definition of solvability, elements $g_1, g_2, \ldots, g_r$ may be chosen in such a way that the cyclic factor group G_i/G_{i-1} is generated by the coset g_iG_{i-1}. The set $\{g_1, g_2, \ldots, g_r\}$ is then a system of generators for G such that the defining relations [see Brown *et al.* (1978), pp. 26–27] have the particularly simple form

$$\begin{aligned} g_1^{m_1} &= e, \\ g_i^{m_i} &= g_{i-1}^{a(i,\,i-1)} g_{i-2}^{a(i,\,i-2)} \cdots g_1^{a(i,\,1)} \qquad \text{for } 2 \le i \le r, \\ g_i^{-1} g_j^{-1} g_i g_j &= g_{j-1}^{b(i,j,j-1)} g_{j-2}^{b(i,j,j-2)} \cdots g_1^{b(i,j,\,1)} \quad \text{for } 1 \le i < j \le r, \end{aligned}$$

with $0 \le a(i,h) < m_h$ and $0 \le b(i,j,h) < m_h$. Each element g of G may then be obtained uniquely as an 'ordered word':

$$g = g_r^{k_r} g_{r-1}^{k_{r-1}} \cdots g_1^{k_1},$$

with $0 \le k_i < m_i$ for all $i = 1, \ldots, r$, using the algorithm of Jürgensen (1970). Such generating sets and defining relations are tabulated in Brown *et al.* (1978, pp. 61–76). An alternative list is given in Janssen (1973, Table 4.3, pp. 121–123, and Appendix D, pp. 262–271).

1.3.4.2.2.4. *Crystallographic group action in real space*

The action of a crystallographic group Γ may be written in terms of standard coordinates in $\mathbb{R}^3/\mathbb{Z}^3$ as

$$(g, \mathbf{x}) \longmapsto S_g(\mathbf{x}) = \mathbf{R}_g\mathbf{x} + \mathbf{t}_g \bmod \Lambda, \quad g \in G,$$

with

$$S_{g_1g_2} = S_{g_1} S_{g_2}.$$

An important characteristic of the representation $\theta : g \longmapsto S_g$ is its *reducibility*, *i.e.* whether or not it has invariant subspaces other than $\{\mathbf{0}\}$ and the whole of $\mathbb{R}^3/\mathbb{Z}^3$. For triclinic, monoclinic and orthorhombic space groups, θ is reducible to a direct sum of three one-dimensional representations:

$$\mathbf{R}_g = \begin{pmatrix} \mathbf{R}_g^{(1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_g^{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_g^{(3)} \end{pmatrix};$$

for trigonal, tetragonal and hexagonal groups, it is reducible to a direct sum of two representations, of dimension 2 and 1, respectively; while for tetrahedral and cubic groups, it is irreducible.

By Schur's lemma (see *e.g.* Ledermann, 1987), any matrix which commutes with all the matrices $\mathbf{R}_g$ for $g \in G$ must be a scalar multiple of the identity in each invariant subspace.

In the reducible cases, the reductions involve changes of basis which will be *rational*, not integral, for those arithmetic classes corresponding to non-primitive lattices. Thus the simplification of having maximally reduced representation has as its counterpart the use of non-primitive lattices.

The notions of orbit, isotropy subgroup and fundamental domain (or asymmetric unit) for the action of G on $\mathbb{R}^3/\mathbb{Z}^3$ are inherited directly from the general setting of Section 1.3.4.2.2.2. Points $\mathbf{x}$ for which $G_\mathbf{x} \neq \{e\}$ are called *special positions*, and the various types of isotropy subgroups which may be encountered in crystallographic groups have been labelled by means of Wyckoff symbols. The representation operators $S_g^\#$ in $L(\mathbb{R}^3/\mathbb{Z}^3)$ have the form:

$$[S_g^\# f](\mathbf{x}) = f[S_g^{-1}(\mathbf{x})] = f[\mathbf{R}_g^{-1}(\mathbf{x} - \mathbf{t}_g)].$$

The operators $R_g^\#$ associated to the purely rotational part of each transformation S_g will also be used. Note the relation: $S_g^\# = \tau_{\mathbf{t}_g} R_g^\#$.

Let a crystal structure be described by the list of the atoms in its unit cell, indexed by $k \in K$. Let the electron-density distribution about the centre of mass of atom k be described by $\wp_k$ with respect to the standard coordinates $\mathbf{x}$. Then the motif $\wp^0$ may be written as a sum of translates:

$$\wp^0 = \sum_{k \in K} \tau_{\mathbf{x}_k} \wp_k$$

and the crystal electron density is $\wp = r^* \wp^0$.

Suppose that $\wp$ is invariant under Γ. If $\mathbf{x}_{k_1}$ and $\mathbf{x}_{k_2}$ are in the same orbit, say $\mathbf{x}_{k_2} = S_g(\mathbf{x}_{k_1})$, then

$$\tau_{\mathbf{x}_{k_2}} \wp_{k_2} = S_g^\#(\tau_{\mathbf{x}_{k_1}} \wp_{k_1}).$$

Therefore if $\mathbf{x}_k$ is a special position and thus $G_{\mathbf{x}_k} \neq \{e\}$, then

$$S_g^\#(\tau_{\mathbf{x}_k} \wp_k) = \tau_{\mathbf{x}_k} \wp_k \quad \text{for all } g \in G_{\mathbf{x}_k}.$$

This identity implies that

$$\mathbf{R}_g\mathbf{x}_k + \mathbf{t}_g \equiv \mathbf{x}_k \bmod \Lambda$$

(the special position condition), and that

$$\wp_k = R_g^\# \wp_k,$$

i.e. that $\wp_k$ must be invariant by the pure rotational part of $G_{\mathbf{x}_k}$. Trueblood (1956) investigated the consequences of this invariance on the thermal vibration tensor of an atom in a special position (see Section 1.3.4.2.2.6 below).

Let J be a subset of K such that $\{\mathbf{x}_j\}_{j \in J}$ contains exactly one atom from each orbit. An orbit decomposition yields an expression for $\wp^0$ in terms of symmetry-unique atoms:

$$\wp^0 = \sum_{j \in J} \left(\sum_{\gamma_j \in G/G_{\mathbf{x}_j}} S_{\gamma_j}^\#(\tau_{\mathbf{x}_j} \wp_j) \right)$$

or equivalently

$$\wp^0(\mathbf{x}) = \sum_{j \in J} \left\{ \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \wp_j[\mathbf{R}_{\gamma_j}^{-1}(\mathbf{x} - \mathbf{t}_{\gamma_j}) - \mathbf{x}_j] \right\}.$$

If the atoms are assumed to be Gaussian, write

$$\rho_j(\mathbf{X}) = \frac{Z_j}{|\det \pi \mathbf{U}_j|^{1/2}} \times \exp(-\tfrac{1}{2}\mathbf{X}^T \mathbf{U}_j^{-1} \mathbf{X}) \text{ in Cartesian Å coordinates,}$$

where Z_j is the total number of electrons, and where the matrix $\mathbf{U}_j$ combines the Gaussian spread of the electrons in atom j at rest with the covariance matrix of the random positional fluctuations of atom j caused by thermal agitation.

In crystallographic coordinates:

$$\wp_j(\mathbf{x}) = \frac{Z_j}{|\det \pi \mathbf{Q}_j|^{1/2}} \times \exp(-\tfrac{1}{2}\mathbf{x}^T \mathbf{Q}_j^{-1} \mathbf{x}) \text{ with } \mathbf{Q}_j = \mathbf{A}^{-1}\mathbf{U}_j(\mathbf{A}^{-1})^T.$$

If atom k is in a special position $\mathbf{x}_k$, then the matrix $\mathbf{Q}_k$ must satisfy the identity

$$\mathbf{R}_g \mathbf{Q}_k \mathbf{R}_g^{-1} = \mathbf{Q}_k$$

for all g in the isotropy subgroup of $\mathbf{x}_k$. This condition may also be written in Cartesian coordinates as

$$\mathbf{T}_g \mathbf{U}_k \mathbf{T}_g^{-1} = \mathbf{U}_k,$$

where

$$\mathbf{T}_g = \mathbf{A}\mathbf{R}_g\mathbf{A}^{-1}.$$

This is a condensed form of the symmetry properties derived by Trueblood (1956).

1.3.4.2.2.5. *Crystallographic group action in reciprocal space*

An elementary discussion of this topic may be found in Chapter 1.4 of this volume.

Having established that the symmetry of a crystal may be most conveniently stated and handled *via* the left representation $g \longmapsto S_g^{\#}$ of G given by its action on electron-density distributions, it is natural to transpose this action by the identity of Section 1.3.2.5.5:

$$\begin{aligned}\bar{\mathscr{F}}[S_g^{\#}T]_{\boldsymbol{\xi}} &= \bar{\mathscr{F}}[\tau_{\mathbf{t}_g}(R_g^{\#}T)]_{\boldsymbol{\xi}} \\ &= \exp(2\pi i\boldsymbol{\xi}\cdot\mathbf{t}_g)[(\mathbf{R}_g^{-1})^{T\#}\bar{\mathscr{F}}[T]]_{\boldsymbol{\xi}}\end{aligned}$$

for any tempered distribution T, *i.e.*

$$\bar{\mathscr{F}}[S_g^{\#}T](\boldsymbol{\xi}) = \exp(2\pi i\boldsymbol{\xi}\cdot\mathbf{t}_g)\bar{\mathscr{F}}[T](\mathbf{R}_g^T\boldsymbol{\xi})$$

whenever the transforms are functions.

Putting $T = \mathscr{p}$, a $\mathbb{Z}^3$-periodic distribution, this relation defines a left action S_g^* of G on $L(\mathbb{Z}^3)$ given by

$$(S_g^*F)(\mathbf{h}) = \exp(2\pi i\boldsymbol{\xi}\cdot\mathbf{t}_g)F(\mathbf{R}_g^T\mathbf{h})$$

which is conjugate to the action $S_g^{\#}$ in the sense that

$$\bar{\mathscr{F}}[S_g^{\#}\mathscr{p}] = S_g^*\bar{\mathscr{F}}[\mathscr{p}], \quad i.e.\ S_g^* = \bar{\mathscr{F}}S_g^{\#}\bar{\mathscr{F}}.$$

The identity $S_g^{\#}\mathscr{p} = \mathscr{p}$ expressing the G-invariance of $\mathscr{p}$ is then equivalent to the identity $S_g^*F = F$ between its structure factors, *i.e.* (Waser, 1955*a*)

$$F(\mathbf{h}) = \exp(2\pi i\mathbf{h}\cdot\mathbf{t}_g)F(\mathbf{R}_g^T\mathbf{h}).$$

If G is made to act on $\mathbb{Z}^3$ *via*

$$\theta^*: \quad (g,\mathbf{h}) \longmapsto (\mathbf{R}_g^{-1})^T\mathbf{h},$$

the usual notions of orbit, isotropy subgroup (denoted $G_{\mathbf{h}}$) and fundamental domain may be attached to this action. The above relation then shows that the spectrum $\{F(\mathbf{h})\}_{\mathbf{h}\in\mathbb{Z}^3}$ is entirely known if it is specified on a fundamental domain D^* containing one reciprocal-lattice point from each orbit of this action.

A reflection $\mathbf{h}$ is called *special* if $G_{\mathbf{h}} \neq \{e\}$. Then for any $g \in G_{\mathbf{h}}$ we have $\mathbf{R}_g^T\mathbf{h} = \mathbf{h}$, and hence

$$F(\mathbf{h}) = \exp(2\pi i\mathbf{h}\cdot\mathbf{t}_g)F(\mathbf{h}),$$

implying that $F(\mathbf{h}) = 0$ unless $\mathbf{h}\cdot\mathbf{t}_g \equiv 0 \bmod 1$. Special reflections $\mathbf{h}$ for which $\mathbf{h}\cdot\mathbf{t}_g \not\equiv 0 \bmod 1$ for some $g \in G_{\mathbf{h}}$ are thus *systematically absent*. This phenomenon is an instance of the duality between periodization and decimation of Section 1.3.2.7.2: if $\mathbf{t}_g \neq \mathbf{0}$, the projection of $\mathscr{p}$ on the direction of $\mathbf{h}$ has period $(\mathbf{t}_g\cdot\mathbf{h})/(\mathbf{h}\cdot\mathbf{h}) < 1$, hence its transform (which is the portion of F supported by the central line through $\mathbf{h}$) will be decimated, giving rise to the above condition.

A reflection $\mathbf{h}$ is called *centric* if $G\mathbf{h} = G(-\mathbf{h})$, *i.e.* if the orbit of $\mathbf{h}$ contains $-\mathbf{h}$. Then $\mathbf{R}_\gamma^T\mathbf{h} = -\mathbf{h}$ for some coset γ in $G/G_{\mathbf{h}}$, so that the following relation must hold:

$$|F(\mathbf{h})|\exp(i\varphi_{\mathbf{h}}) = \exp(2\pi i\mathbf{h}\cdot\mathbf{t}_\gamma)|F(-\mathbf{h})|\exp(i\varphi_{-\mathbf{h}}).$$

In the absence of dispersion, Friedel's law gives rise to the *phase restriction*:

$$\varphi_{\mathbf{h}} \equiv \pi\mathbf{h}\cdot\mathbf{t}_\gamma \bmod \pi.$$

The value of the restricted phase is independent of the choice of coset representative γ. Indeed, if γ' is another choice, then $\gamma' = g\gamma$ with $g \in G_{\mathbf{h}}$ and by the Frobenius congruences $\mathbf{t}_{\gamma'} = \mathbf{R}_g\mathbf{t}_\gamma + \mathbf{t}_g$, so that

$$\mathbf{h}\cdot\mathbf{t}_{\gamma'} \equiv (\mathbf{R}_g^T\mathbf{h})\cdot\mathbf{t}_\gamma + \mathbf{h}\cdot\mathbf{t}_g \bmod 1.$$

Since $g \in G_{\mathbf{h}}$, $\mathbf{R}_g^T\mathbf{h} = \mathbf{h}$ and $\mathbf{h}\cdot\mathbf{t}_g \equiv 0 \bmod 1$ if $\mathbf{h}$ is not a systematic absence: thus

$$\pi\mathbf{h}\cdot\mathbf{t}_\gamma \equiv \pi\mathbf{h}\cdot\mathbf{t}_\gamma \bmod \pi.$$

The treatment of centred lattices may be viewed as another instance of the duality between periodization and decimation (Section 1.3.2.7.2): the periodization of the electron density by the non-primitive lattice translations has as its counterpart in reciprocal space the decimation of the transform by the 'reflection conditions' describing the allowed reflections, the decimation and periodization matrices being each other's contragredient.

The reader may consult the papers by Bienenstock & Ewald (1962) and Wells (1965) for earlier approaches to this material.

1.3.4.2.2.6. *Structure-factor calculation*

Structure factors may be calculated from a list of symmetry-unique atoms by Fourier transformation of the orbit decomposition formula for the motif $\mathscr{p}^0$ given in Section 1.3.4.2.2.4:

$$\begin{aligned}F(\mathbf{h}) &= \bar{\mathscr{F}}[\mathscr{p}^0](\mathbf{h}) \\ &= \bar{\mathscr{F}}\left[\sum_{j\in J}\left(\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} S_{\gamma_j}^{\#}(\tau_{\mathbf{x}_j}\mathscr{p}_j)\right)\right](\mathbf{h}) \\ &= \sum_{j\in J}\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \bar{\mathscr{F}}[\tau_{\mathbf{t}_{\gamma_j}}\mathbf{R}_{\gamma_j}^{\#}\tau_{\mathbf{x}_j}\mathscr{p}_j](\mathbf{h}) \\ &= \sum_{j\in J}\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \exp(2\pi i\mathbf{h}\cdot\mathbf{t}_{\gamma_j}) \\ &\quad\times[(\mathbf{R}_{\gamma_j}^{-1})^{T\#}[\exp(2\pi i\boldsymbol{\xi}\cdot\mathbf{x}_j)\bar{\mathscr{F}}[\mathscr{p}_j]_{\boldsymbol{\xi}}]](\mathbf{h}) \\ &= \sum_{j\in J}\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \exp(2\pi i\mathbf{h}\cdot\mathbf{t}_{\gamma_j}) \\ &\quad\times\exp[2\pi i(\mathbf{R}_{\gamma_j}^T\mathbf{h})\cdot\mathbf{x}_j]\bar{\mathscr{F}}[\mathscr{p}_j](\mathbf{R}_{\gamma_j}^T\mathbf{h});\end{aligned}$$

i.e. finally:

$$F(\mathbf{h}) = \sum_{j\in J}\sum_{\gamma_j\in G/G_{\mathbf{x}_j}} \exp\{2\pi i\mathbf{h}\cdot[S_{\gamma_j}(\mathbf{x}_j)]\}\bar{\mathscr{F}}[\mathscr{p}_j](\mathbf{R}_{\gamma_j}^T\mathbf{h}).$$

In the case of Gaussian atoms, the atomic transforms are

$$\bar{\mathscr{F}}[\mathscr{p}_j](\mathbf{h}) = Z_j\exp[-\tfrac{1}{2}\mathbf{h}^T(4\pi^2\mathbf{Q}_j)\mathbf{h}]$$

or equivalently

$$\bar{\mathscr{F}}[\rho_j](\mathbf{H}) = Z_j\exp[-\tfrac{1}{2}\mathbf{H}^T(4\pi^2\mathbf{U}_j)\mathbf{H}].$$

Two common forms of equivalent temperature factors (incorporating both atomic form and thermal motion) are

(i) isotropic B:

$$\bar{\mathscr{F}}[\mathscr{p}_j](\mathbf{h}) = Z_j\exp(-\tfrac{1}{4}B_j\mathbf{H}^T\mathbf{H}),$$

so that $\mathbf{U}_j = (B_j/8\pi^2)\mathbf{I}$, or $\mathbf{Q}_j = (B_j/8\pi^2)\mathbf{A}^{-1}(\mathbf{A}^{-1})^T$;

(ii) anisotropic β's:

$$\bar{\mathscr{F}}[\rho_j](\mathbf{h}) = Z_j \exp(-\mathbf{h}^T \boldsymbol{\beta}_j \mathbf{h}),$$

so that $\boldsymbol{\beta}_j = 2\pi^2 \mathbf{Q}_j = 2\pi^2 \mathbf{A}^{-1} \mathbf{U}_j (\mathbf{A}^{-1})^T$, or $\mathbf{U}_j = (1/2\pi^2)\mathbf{A}\beta_j \mathbf{A}^T$.

In the first case, $\bar{\mathscr{F}}[\rho_j](\mathbf{R}_{\gamma_j}^T \mathbf{h})$ does not depend on γ_j, and therefore:

$$F(\mathbf{h}) = \sum_{j\in J} Z_j \exp\{-\tfrac{1}{4}B_j \mathbf{h}^T [\mathbf{A}^{-1}(\mathbf{A}^{-1})^T]\mathbf{h}\} \times \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \exp\{2\pi i \mathbf{h}\cdot[S_{\gamma_j}(\mathbf{x}_j)]\}.$$

In the second case, however, no such simplification can occur:

$$F(\mathbf{h}) = \sum_{j\in J} Z_j \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \exp[-\mathbf{h}^T (\mathbf{R}_{\gamma_j} \boldsymbol{\beta}_j \mathbf{R}_{\gamma_j}^T)\mathbf{h}] \times \exp\{2\pi i \mathbf{h}\cdot[S_{\gamma_j}(\mathbf{x}_j)]\}.$$

These formulae, or special cases of them, were derived by Rollett & Davies (1955), Waser (1955*b*), and Trueblood (1956).

The computation of structure factors by applying the discrete Fourier transform to a set of electron-density values calculated on a grid will be examined in Section 1.3.4.4.5.

1.3.4.2.2.7. *Electron-density calculations*

A formula for the Fourier synthesis of electron-density maps from symmetry-unique structure factors is readily obtained by orbit decomposition:

$$\begin{aligned}\rho(\mathbf{x}) &= \sum_{\mathbf{h}\in\mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}) \\ &= \sum_{l\in L}\left[\sum_{\gamma_l \in G/G_{\mathbf{h}_l}} F(\mathbf{R}_{\gamma_l}^T \mathbf{h}_l) \exp[-2\pi i (\mathbf{R}_{\gamma_l}^T \mathbf{h}_l)\cdot\mathbf{x}]\right] \\ &= \sum_{l\in L} F(\mathbf{h}_l)\left[\sum_{\gamma_l \in G/G_{\mathbf{h}_l}} \exp\{-2\pi i \mathbf{h}_l \cdot [S_{\gamma_l}(\mathbf{x})]\}\right],\end{aligned}$$

where L is a subset of $\mathbb{Z}^3$ such that $\{\mathbf{h}_l\}_{l\in L}$ contains exactly one point of each orbit for the action $\theta^* : (g, \mathbf{h}) \longmapsto (\mathbf{R}_g^{-1})^T \mathbf{h}$ of G on $\mathbb{Z}^3$. The physical electron density per cubic ångström is then

$$\rho(\mathbf{X}) = \frac{1}{V}\rho(\mathbf{A}\mathbf{x})$$

with V in Å^3.

In the absence of anomalous scatterers in the crystal and of a centre of inversion $-\mathbf{I}$ in Γ, the spectrum $\{F(\mathbf{h})\}_{\mathbf{h}\in\mathbb{Z}^3}$ has an extra symmetry, namely the Hermitian symmetry expressing Friedel's law (Section 1.3.4.2.1.4). The action of a centre of inversion may be added to that of Γ to obtain further simplification in the above formula: under this extra action, an orbit $G\mathbf{h}_l$ with $\mathbf{h}_l \neq \mathbf{0}$ is either mapped into itself or into the disjoint orbit $G(-\mathbf{h}_l)$; the terms corresponding to $+\mathbf{h}_l$ and $-\mathbf{h}_l$ may then be grouped *within* the common orbit in the first case, and *between* the two orbits in the second case.

Case 1: $G(-\mathbf{h}_l) = G\mathbf{h}_l, \mathbf{h}_l$ is *centric*. The cosets in $G/G_{\mathbf{h}_l}$ may be partitioned into two disjoint classes by picking one coset in each of the two-coset orbits of the action of $-\mathbf{I}$. Let $(G/G_{\mathbf{h}_l})^+$ denote one such class: then the *reduced orbit*

$$\{\mathbf{R}_{\gamma_l}^T \mathbf{h}_l | \gamma_l \in (G/G_{\mathbf{h}_l})^+\}$$

contains exactly once the Friedel-unique half of the full orbit $G\mathbf{h}_l$, and thus

$$|(G/G_{\mathbf{h}_l})^+| = \tfrac{1}{2}|G/G_{\mathbf{h}_l}|.$$

Grouping the summands for $+\mathbf{h}_l$ and $-\mathbf{h}_l$ yields a real-valued summand

$$2F(\mathbf{h}_l) \sum_{\gamma_l \in (G/G_{\mathbf{h}_l})^+} \cos[2\pi \mathbf{h}_l \cdot [S_{\gamma_l}(\mathbf{x})] - \varphi_{\mathbf{h}l}].$$

Case 2: $G(-\mathbf{h}_l) \neq G\mathbf{h}_l$, $\mathbf{h}_l$ is *acentric*. The two orbits are then disjoint, and the summands corresponding to $+\mathbf{h}_l$ and $-\mathbf{h}_l$ may be grouped together into a single real-valued summand

$$2F(\mathbf{h}_l) \sum_{\gamma_l \in G/G_{\mathbf{h}_l}} \cos[2\pi \mathbf{h}_l \cdot [S_{\gamma_l}(\mathbf{x})] - \varphi_{\mathbf{h}_l}].$$

In order to reindex the collection of all summands of ρ, put

$$L = L_c \cup L_a,$$

where L_c labels the Friedel-unique centric reflections in L and L_a the acentric ones, and let L_a^+ stand for a subset of L_a containing a unique element of each pair $\{+\mathbf{h}_l, -\mathbf{h}_l\}$ for $l \in L_a$. Then

$$\begin{aligned}\rho(\mathbf{x}) = F(\mathbf{0}) &+ \sum_{c\in L_c}\left[2F(\mathbf{h}_c) \sum_{\gamma_c \in (G/G_{\mathbf{h}_c})^+} \cos[2\pi \mathbf{h}_c \cdot [S_{\gamma_c}(\mathbf{x})] - \varphi_{\mathbf{h}_c}]\right] \\ &+ \sum_{a\in L_a^+}\left[2F(\mathbf{h}_a) \sum_{\gamma_a \in G/G_{\mathbf{h}_a}} \cos[2\pi \mathbf{h}_a \cdot [S_{\gamma_a}(\mathbf{x})] - \varphi_{\mathbf{h}_a}]\right].\end{aligned}$$

1.3.4.2.2.8. *Parseval's theorem with crystallographic symmetry*

The general statement of Parseval's theorem given in Section 1.3.4.2.1.5 may be rewritten in terms of symmetry-unique structure factors and electron densities by means of orbit decomposition.

In reciprocal space,

$$\sum_{\mathbf{h}\in\mathbb{Z}^3} \overline{F_1(\mathbf{h})} F_2(\mathbf{h}) = \sum_{l\in L} \sum_{\gamma_l \in G/G_{\mathbf{h}_l}} \overline{F_1(\mathbf{R}_{\gamma_l}^T \mathbf{h}_l)} F_2(\mathbf{R}_{\gamma_l}^T \mathbf{h}_l);$$

for each l, the summands corresponding to the various γ_l are equal, so that the left-hand side is equal to

$$\begin{aligned}&F_1(\mathbf{0})F_2(\mathbf{0}) \\ &+ \sum_{c\in L_c} 2|(G/G_{\mathbf{h}_c})^+| |F_1(\mathbf{h}_c)||F_2(\mathbf{h}_c)| \cos[\varphi_1(\mathbf{h}_c) - \varphi_2(\mathbf{h}_c)] \\ &+ \sum_{a\in L_a^+} 2|G/G_{\mathbf{h}_a}| |F_1(\mathbf{h}_a)||F_2(\mathbf{h}_a)| \cos[\varphi_1(\mathbf{h}_a) - \varphi_2(\mathbf{h}_a)].\end{aligned}$$

In real space, the triple integral may be rewritten as

$$\int_{\mathbb{R}^3/\mathbb{Z}^3} \overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x})\ \mathrm{d}^3\mathbf{x} = |G| \int_D \overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x})\ \mathrm{d}^3\mathbf{x}$$

(where D is the asymmetric unit) if ρ_1 and ρ_2 are smooth densities, since the set of special positions has measure zero. If, however, the integral is approximated as a sum over a G-invariant grid defined by decimation matrix $\mathbf{N}$, special positions on this grid must be taken into account:

$$\begin{aligned}\frac{1}{|\mathbf{N}|} \sum_{\mathbf{k}\in\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3} &\overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x}) \\ &= \frac{1}{|\mathbf{N}|}\sum_{\mathbf{x}\in D} [G : G_{\mathbf{x}}]\overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x}) \\ &= \frac{|G|}{|\mathbf{N}|}\sum_{\mathbf{x}\in D} \frac{1}{|G_{\mathbf{x}}|}\overline{\rho_1(\mathbf{x})}\rho_2(\mathbf{x}),\end{aligned}$$

where the discrete asymmetric unit D contains exactly one point in each orbit of G in $\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$.

1.3.4.2.2.9. *Convolution theorems with crystallographic symmetry*

The standard convolution theorems derived in the absence of symmetry are readily seen to follow from simple properties of functions $e^{\pm}(\mathbf{h},\mathbf{x}) = \exp(\pm 2\pi i \mathbf{h}\cdot\mathbf{x})$ (denoted simply e in formulae which are valid for both signs), namely:

$$\text{(i)} \qquad e(\mathbf{h},\mathbf{x}) \times e(\mathbf{k},\mathbf{x}) = e(\mathbf{h}+\mathbf{k},\mathbf{x}),$$
$$\text{(ii)} \qquad e(\mathbf{h},\mathbf{x}) \times e(\mathbf{h},\mathbf{y}) = e(\mathbf{h},\mathbf{x}+\mathbf{y}).$$

These relations imply that the families of functions

$$\{\mathbf{x} \longmapsto e(\mathbf{h},\mathbf{x})\}_{\mathbf{h}\in\mathbb{Z}^3} \qquad \text{in real space}$$

and

$$\{\mathbf{h} \longmapsto e(\mathbf{h},\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^3/\mathbb{Z}^3} \quad \text{in reciprocal space}$$

both generate an *algebra* of functions, *i.e.* a vector space endowed with an internal multiplication, since (i) and (ii) show how to 'linearize products'.

Friedel's law (when applicable) on the one hand, and the Fourier relation between intensities and the Patterson function on the other hand, both follow from the property

$$\text{(iii)} \qquad \overline{e(\mathbf{h},\mathbf{x})} = e(-\mathbf{h},\mathbf{x}) = e(\mathbf{h},-\mathbf{x}).$$

When crystallographic symmetry is present, the convolution theorems remain valid in their original form if written out in terms of 'expanded' data, but acquire a different form when rewritten in terms of symmetry-unique data only. This rewriting is made possible by the extra relation (Section 1.3.4.2.2.5)

$$\text{(iv)} \qquad S^{\#}_{g^{-1}} e(\mathbf{h},\mathbf{x}) \equiv e[\mathbf{h}, S_g(\mathbf{x})] = e(\mathbf{h},\mathbf{t}_g) e(\mathbf{R}_g^T\mathbf{h},\mathbf{x})$$

or equivalently

$$\text{(iv}') \qquad S^{\#}_{g} e(\mathbf{h},\mathbf{x}) \equiv e[\mathbf{h}, S_g^{-1}(\mathbf{x})] = e[(-\mathbf{R}_g^{-1})^T\mathbf{h},\mathbf{t}_g] e[(\mathbf{R}_g^{-1})^T\mathbf{h},\mathbf{x}].$$

The kernels of symmetrized Fourier transforms are not the functions e but rather the symmetrized sums

$$\Xi^{\pm}(\mathbf{h},\mathbf{x}) = \sum_{g\in G} e^{\pm}[\mathbf{h}, S_g(\mathbf{x})] = \sum_{g\in G} e^{\pm}[\mathbf{h}, S_g^{-1}(\mathbf{x})]$$

for which the linearization formulae are readily obtained using (i), (ii) and (iv) as

$$\text{(i)}_G \qquad \Xi^{\pm}(\mathbf{h},\mathbf{x})\Xi^{\pm}(\mathbf{k},\mathbf{x}) = \sum_{g\in G} e^{\pm}(\mathbf{k},\mathbf{t}_g)\Xi^{\pm}(\mathbf{h}+\mathbf{R}_g^T\mathbf{k},\mathbf{x}),$$
$$\text{(ii)}_G \qquad \Xi^{\pm}(\mathbf{h},\mathbf{x})\Xi^{\pm}(\mathbf{h},\mathbf{y}) = \sum_{g\in G} \Xi^{\pm}[\mathbf{h},\mathbf{x}+S_g(\mathbf{y})],$$

where the choice of sign in $\pm$ must be the same throughout each formula.

Formulae $(\text{i})_G$ defining the 'structure-factor algebra' associated to G were derived by Bertaut (1955*c*, 1956*b*,*c*, 1959*a*,*b*) and Bertaut & Waser (1957) in another context.

The forward convolution theorem (in discrete form) then follows. Let

$$F_1(\mathbf{h}) = \sum_{\mathbf{y}\in D} \frac{1}{|G_{\mathbf{y}}|} \rho_1(\mathbf{y}) \Xi^{+}(\mathbf{h},\mathbf{y}),$$
$$F_2(\mathbf{h}) = \sum_{\mathbf{z}\in D} \frac{1}{|G_{\mathbf{z}}|} \rho_2(\mathbf{z}) \Xi^{+}(\mathbf{h},\mathbf{z}),$$

then

$$F_1(\mathbf{h})F_2(\mathbf{h}) = \sum_{\mathbf{x}\in D} \frac{1}{|G_{\mathbf{x}}|} \sigma(\mathbf{x}) \Xi^{+}(\mathbf{h},\mathbf{x})$$

with

$$\sigma(\mathbf{x}) = \frac{1}{|\mathbf{N}|} \sum_{\mathbf{z}\in D} \sum_{g\in G} \frac{|G_{\mathbf{x}}|}{|G_{\mathbf{x}-S_g(\mathbf{z})}| \times |G_{\mathbf{z}}|} \rho_1[\mathbf{x}-S_g(\mathbf{z})]\rho_2(\mathbf{z}).$$

The backward convolution theorem is derived similarly. Let

$$\rho_1(\mathbf{x}) = \sum_{\mathbf{k}\in D^*} \frac{1}{|G_{\mathbf{k}}|} F_1(\mathbf{k}) \Xi^{-}(\mathbf{k},\mathbf{x}),$$
$$\rho_2(\mathbf{x}) = \sum_{\mathbf{l}\in D^*} \frac{1}{|G_{\mathbf{l}}|} F_2(\mathbf{l}) \Xi^{-}(\mathbf{l},\mathbf{x}),$$

then

$$\rho_1(\mathbf{x})\rho_2(\mathbf{x}) = \sum_{\mathbf{h}\in D^*} \frac{1}{|G_{\mathbf{h}}|} F(\mathbf{h}) \Xi^{-}(\mathbf{h},\mathbf{x})$$

with

$$F(\mathbf{h}) = \sum_{\mathbf{l}\in D^*} \sum_{g\in G} \frac{|G_{\mathbf{h}}|}{|G_{\mathbf{h}-\mathbf{R}_g^T(\mathbf{l})}| \times |G_{\mathbf{l}}|} e^{-}(\mathbf{l},\mathbf{t}_g) F_1(\mathbf{h}-\mathbf{R}_g^T\mathbf{l}) F_2(\mathbf{l}).$$

Both formulae are simply orbit decompositions of their symmetry-free counterparts.

1.3.4.2.2.10. *Correlation and Patterson functions*

Consider two model electron densities ρ_1 and ρ_2 with the same period lattice $\mathbb{Z}^3$ and the same space group G. Write their motifs in terms of atomic electron densities (Section 1.3.4.2.2.4) as

$$\rho_1^0 = \sum_{j_1\in J_1} \left(\sum_{\gamma_{j_1}\in G/G_{\mathbf{x}_{j_1}^{(1)}}} S^{\#}_{\gamma_{j_1}} (\tau_{\mathbf{x}_{j_1}^{(1)}} \rho_{j_1}^{(1)}) \right),$$
$$\rho_2^0 = \sum_{j_2\in J_2} \left(\sum_{\gamma_{j_2}\in G/G_{\mathbf{x}_{j_2}^{(2)}}} S^{\#}_{\gamma_{j_2}} (\tau_{\mathbf{x}_{j_2}^{(2)}} \rho_{j_2}^{(2)}) \right),$$

where J_1 and J_2 label the symmetry-unique atoms placed at positions $\{\mathbf{x}_{j_1}^{(1)}\}_{j_1\in J_1}$ and $\{\mathbf{x}_{j_2}^{(2)}\}_{j_2\in J_2}$, respectively.

To calculate the correlation between ρ_1 and ρ_2 we need the following preliminary formulae, which are easily established: if $S(\mathbf{x}) = \mathbf{R}\mathbf{x}+\mathbf{t}$ and f is an arbitrary function on $\mathbb{R}^3$, then

$$(R^{\#}f)^{\check{}} = R^{\#}\check{f}, \quad (\tau_{\mathbf{x}} f)^{\check{}} = \tau_{-\mathbf{x}}\check{f}, \quad R^{\#}(\tau_{\mathbf{x}} f) = \tau_{\mathbf{R}\mathbf{x}} f,$$

hence

$$S^{\#}(\tau_{\mathbf{x}} f) = \tau_{S(\mathbf{x})} R^{\#} f \quad \text{and} \quad [S^{\#}(\tau_{\mathbf{x}} f)]^{\check{}} = \tau_{-S(\mathbf{x})} R^{\#}\check{f};$$

and

$$S_1^{\#} f_1 * S_2^{\#} f_2 = S_1^{\#}[f_1 * (S_1^{-1}S_2)^{\#} f_2] = S_2^{\#}[(S_2^{-1}S_1)^{\#} f_1 * f_2].$$

The cross correlation $\check{\rho}_1^0 * \rho_2^0$ between motifs is therefore

$$\check{\rho}_1^0 * \rho_2^0 = \sum_{j_1}\sum_{j_2}\sum_{\gamma_{j_1}}\sum_{\gamma_{j_2}} [S^{\#}_{\gamma_{j_1}} (\tau_{\mathbf{x}_{j_1}^{(1)}} \rho_{j_1}^{(1)})]^{\check{}} * [S^{\#}_{\gamma_{j_2}} (\tau_{\mathbf{x}_{j_2}^{(2)}} \rho_{j_2}^{(2)})]$$
$$= \sum_{j_1}\sum_{j_2}\sum_{\gamma_{j_1}}\sum_{\gamma_{j_2}} \tau_{S_{\gamma_{j_2}}(\mathbf{x}_{j_2}^{(2)}) - S_{\gamma_{j_1}}(\mathbf{x}_{j_1}^{(1)})} [(R^{\#}_{\gamma_{j_1}} \check{\rho}_{j_1}^{(1)}) * (R^{\#}_{\gamma_{j_2}} \rho_{j_2}^{(2)})]$$

which contains a peak of shape $(R^{\#}_{\gamma_{j_1}} \check{\rho}_{j_1}^{(1)}) * (R^{\#}_{\gamma_{j_2}} \rho_{j_2}^{(2)})$ at the interatomic vector $S_{\gamma_{j_2}}(\mathbf{x}_{j_2}^{(2)}) - S_{\gamma_{j_1}}(\mathbf{x}_{j_1}^{(1)})$ for each $j_1 \in J_1$, $j_2 \in J_2$, $\gamma_{j_1} \in G/G_{\mathbf{x}_{j_1}^{(1)}}$, $\gamma_{j_2} \in G/G_{\mathbf{x}_{j_2}^{(2)}}$.

The cross-correlation $r * \breve{\wp}_1^0 * \wp_2^0$ between the original electron densities is then obtained by further periodizing by $\mathbb{Z}^3$.

Note that these expressions are valid for any choice of 'atomic' density functions $\wp_{j_1}^{(1)}$ and $\wp_{j_2}^{(2)}$, which may be taken as molecular fragments if desired (see Section 1.3.4.4.8).

If G contains elements g such that $\mathbf{R}_g$ has an eigenspace E_1 with eigenvalue 1 and an invariant complementary subspace E_2, while $\mathbf{t}_g$ has a non-zero component $\mathbf{t}_g^{(1)}$ in E_1, then the Patterson function $r * \breve{\wp}^0 * \wp^0$ will contain *Harker peaks* (Harker, 1936) of the form

$$S_g(\mathbf{x}) - \mathbf{x} = \mathbf{t}_g^{(1)} \oplus (S_g^{(2)}(\mathbf{x}) - \mathbf{x})$$

[where $S_g^{(s)}$ represent the action of g in E_2] in the translate of E_1 by $\mathbf{t}_g^{(1)}$.

1.3.4.3. *Crystallographic discrete Fourier transform algorithms*

1.3.4.3.1. *Historical introduction*

In 1929, W. L. Bragg demonstrated the practical usefulness of the Fourier transform relation between electron density and structure factors by determining the structure of diopside from three principal projections calculated numerically by 2D Fourier summation (Bragg, 1929). It was immediately realized that the systematic use of this powerful method, and of its extension to three dimensions, would entail considerable amounts of numerical computation which had to be organized efficiently. As no other branch of applied science had yet needed this type of computation, crystallographers had to invent their own techniques.

The first step was taken by Beevers & Lipson (1934) who pointed out that a 2D summation could be factored into successive 1D summations. This is essentially the tensor product property of the Fourier transform (Sections 1.3.2.4.2.4, 1.3.3.3.1), although its aspect is rendered somewhat complicated by the use of sines and cosines instead of complex exponentials. Computation is economized to the extent that the cost of an $N \times N$ transform grows with N as $2N^3$ rather than N^4. Generalization to 3D is immediate, reducing computation size from N^6 to $3N^4$ for an $N \times N \times N$ transform. The complication introduced by using expressions in terms of sines and cosines is turned to advantage when symmetry is present, as certain families of terms are systematically absent or are simply related to each other; multiplicity corrections must, however, be introduced. The necessary information was tabulated for each space group by Lonsdale (1936), and was later incorporated into Volume I of *International Tables.*

The second step was taken by Beevers & Lipson (1936) and Lipson & Beevers (1936) in the form of the invention of the 'Beevers–Lipson strips', a practical device which was to assist a whole generation of crystallographers in the numerical computation of crystallographic Fourier sums. The strips comprise a set of 'cosine strips' tabulating the functions

$$A\cos\left(\frac{2\pi hm}{60}\right) \quad (A = 1, 2, \ldots, 99;\ h = 1, 2, \ldots, 99)$$

and a set of 'sine strips' tabulating the functions

$$B\sin\left(\frac{2\pi hm}{60}\right) \quad (B = 1, 2, \ldots, 99;\ h = 1, 2, \ldots, 99)$$

for the 16 arguments $m = 0, 1, \ldots, 15$. Function values are rounded to the nearest integer, and those for other arguments m may be obtained by using the symmetry properties of the sine and cosine functions. A Fourier summation of the form

$$Y(m) = \sum_{j=1}^{n}\left[A_j\cos\left(\frac{2\pi h_j m}{60}\right) + B_j\sin\left(\frac{2\pi h_j m}{60}\right)\right]$$

is then performed by selecting the n cosine strips labelled (A_j, h_j) and the n sine strips labelled (B_j, h_j), placing them in register, and adding the tabulated values columnwise. The number 60 was chosen as the l.c.m. of 12 (itself the l.c.m. of the orders of all possible non-primitive translations) and of 10 (for decimal convenience). The limited accuracy imposed by the two-digit tabulation was later improved by Robertson's sorting board (Robertson, 1936*a*,*b*) or by the use of separate strips for each decimal digit of the amplitude (Booth, 1948*b*), which allowed three-digit tabulation while keeping the set of strips within manageable size. Cochran (1948*a*) found that, for most structures under study at the time, the numerical inaccuracies of the method were less than the level of error in the experimental data. The sampling rate was subsequently increased from 60 to 120 (Beevers, 1952) to cope with larger unit cells.

Further gains in speed and accuracy were sought through the construction of special-purpose mechanical, electro-mechanical, electronic or optical devices. Two striking examples are the mechanical computer RUFUS built by Robertson (1954, 1955, 1961) on the principle of previous strip methods (see also Robertson, 1932) and the electronic analogue computer X-RAC built by Pepinsky, capable of real-time calculation and display of 2D and 3D Fourier syntheses (Pepinsky, 1947; Pepinsky & Sayre, 1948; Pepinsky *et al.*, 1961; see also Suryan, 1957). The optical methods of Lipson & Taylor (1951, 1958) also deserve mention. Many other ingenious devices were invented, whose descriptions may be found in Booth (1948*b*), Niggli (1961), and Lipson & Cochran (1968).

Later, commercial punched-card machines were programmed to carry out Fourier summations or structure-factor calculations (Shaffer *et al.*, 1946*a*,*b*; Cox *et al.*, 1947, 1949; Cox & Jeffrey, 1949; Donohue & Schomaker, 1949; Grems & Kasper, 1949; Hodgson *et al.*, 1949; Greenhalgh & Jeffrey, 1950; Kitz & Marchington, 1953).

The modern era of digital electronic computation of Fourier series was initiated by the work of Bennett & Kendrew (1952), Mayer & Trueblood (1953), Ahmed & Cruickshank (1953*b*), Sparks *et al.* (1956) and Fowweather (1955). Their Fourier-synthesis programs used Beevers–Lipson factorization, the program by Sparks *et al.* being the first 3D Fourier program useable for all space groups (although these were treated as $P1$ or $P\bar{1}$ by data expansion). Ahmed & Barnes (1958) then proposed a general programming technique to allow full use of symmetry elements (orthorhombic or lower) in the 3D Beevers–Lipson factorization process, including multiplicity corrections. Their method was later adopted by Shoemaker & Sly (1961), and by crystallographic program writers at large.

The discovery of the FFT algorithm by Cooley & Tukey in 1965, which instantly transformed electrical engineering and several other disciplines, paradoxically failed to have an immediate impact on crystallographic computing. A plausible explanation is that the calculation of large 3D Fourier maps was a relatively infrequent task which was not thought to constitute a bottleneck, as crystallographers had learned to settle most structural questions by means of cheaper 2D sections or projections. It is significant in this respect that the first use of the FFT in crystallography by Barrett & Zwick (1971) should have occurred as part of an iterative scheme for improving protein phases by density modification in real space, which required a much greater number of Fourier transformations than any previous method. Independently, Bondot (1971) had attracted attention to the merits of the FFT algorithm.

The FFT program used by Barrett & Zwick had been written for signal-processing applications. It was restricted to sampling rates of the form 2^n, and was not designed to take advantage of crystallographic symmetry at any stage of the calculation; Bantz & Zwick (1974) later improved this situation somewhat.

It was the work of Ten Eyck (1973) and Immirzi (1973, 1976) which led to the general adoption of the FFT in crystallographic computing. Immirzi treated all space groups as $P1$ by data expansion. Ten Eyck based his program on a versatile multi-radix FFT routine (Gentleman & Sande, 1966) coupled with a flexible indexing scheme for dealing efficiently with multidimensional transforms. He also addressed the problems of incorporating symmetry elements of order 2 into the factorization of 1D transforms, and of transposing intermediate results by other symmetry elements. He was thus able to show that in a large number of space groups (including the 74 space groups having orthorhombic or lower symmetry) it is possible to calculate only the unique results from the unique data within the logic of the FFT algorithm. Ten Eyck wrote and circulated a package of programs for computing Fourier maps and re-analysing them into structure factors in some simple space groups ($P1$, $P\bar{1}$, $P2$, $P2/m$, $P2_1$, $P222$, $P2_12_12_1$, $Pmmm$). This package was later augmented by a handful of new space-group-specific programs contributed by other crystallographers ($P2_12_12$, $I222$, $P3_121$, $P4_12_12$). The writing of such programs is an undertaking of substantial complexity, which has deterred all but the bravest: the usual practice is now to expand data for a high-symmetry space group to the largest subgroup for which a specific FFT program exists in the package, rather than attempt to write a new program. Attempts have been made to introduce more modern approaches to the calculation of crystallographic Fourier transforms (Auslander, Feig & Winograd, 1982; Auslander & Shenefelt, 1987; Auslander *et al.*, 1988) but have not gone beyond the stage of preliminary studies.

The task of fully exploiting the FFT algorithm in crystallographic computations is therefore still unfinished, and it is the purpose of this section to provide a systematic treatment such as that (say) of Ahmed & Barnes (1958) for the Beevers–Lipson algorithm.

Ten Eyck's approach, based on the reducibility of certain space groups, is extended by the derivation of a universal transposition formula for intermediate results. It is then shown that space groups which are not completely reducible may nevertheless be treated by three-dimensional Cooley–Tukey factorization in such a way that their symmetry may be fully exploited, whatever the shape of their asymmetric unit. Finally, new factorization methods with built-in symmetries are presented. The unifying concept throughout this presentation is that of 'group action' on indexing sets, and of 'orbit exchange' when this action has a composite structure; it affords new ways of rationalizing the use of symmetry, or of improving computational speed, or both.

1.3.4.3.2. *Defining relations and symmetry considerations*

A finite set of reflections $\{F_{\mathbf{h}_l}\}_{l\in L}$ can be periodized without aliasing by the translations of a suitable sublattice $\mathbf{N}^T\Lambda^*$ of the reciprocal lattice Λ^*; the converse operation in real space is the sampling of ρ at points $\mathbf{X}$ of a grid of the form $\mathbf{N}^{-1}\Lambda$ (Section 1.3.2.7.3). In standard coordinates, $\{F_{\mathbf{h}_l}\}_{l\in L}$ is periodized by $\mathbf{N}^T\mathbb{Z}^3$, and ρ is sampled at points $\mathbf{x}\in\mathbf{N}^{-1}\mathbb{Z}^3$.

In the absence of symmetry, the unique data are

– the $F_{\mathbf{h}}$ indexed by $\mathbf{h}\in\mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3$ in reciprocal space;

– the $\rho_{\mathbf{x}}$ indexed by $\mathbf{x}\in(\mathbf{N}^{-1}\mathbb{Z}^3)/\mathbb{Z}^3$; or equivalently the $\rho_{\mathbf{m}}$ indexed by $\mathbf{m}\in\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$, where $\mathbf{x}=\mathbf{N}^{-1}\mathbf{m}$.

They are connected by the ordinary DFT relations:

$$F_{\mathbf{h}}=\frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{x}\in(\mathbf{N}^{-1}\mathbb{Z}^3)/\mathbb{Z}^3}\rho_{\mathbf{x}}\exp(2\pi i\mathbf{h}\cdot\mathbf{x})$$

or

$$F_{\mathbf{h}}=\frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{m}\in\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3}\rho_{\mathbf{m}}\exp[2\pi i\mathbf{h}\cdot(\mathbf{N}^{-1}\mathbf{m})]$$

and

$$\rho_{\mathbf{x}}=\sum_{\mathbf{h}=\mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3}F_{\mathbf{h}}\exp(-2\pi i\mathbf{h}\cdot\mathbf{x})$$

or

$$\rho_{\mathbf{m}}=\sum_{\mathbf{h}\in\mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3}F_{\mathbf{h}}\exp[-2\pi i\mathbf{h}\cdot(\mathbf{N}^{-1}\mathbf{m})].$$

In the presence of symmetry, the unique data are

– $\{\rho_{\mathbf{x}}\}_{\mathbf{x}\in D}$ or $\{\rho_{\mathbf{m}}\}_{\mathbf{m}\in D}$ in real space (by abuse of notation, D will denote an asymmetric unit for $\mathbf{x}$ or for $\mathbf{m}$ indifferently);

– $\{F_{\mathbf{h}}\}_{\mathbf{h}\in D^*}$ in reciprocal space.

The previous summations may then be subjected to orbital decomposition, to yield the following 'crystallographic DFT' (CDFT) defining relations:

$$\begin{aligned}F_{\mathbf{h}}&=\frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{x}\in D}\rho_{\mathbf{x}}\left[\sum_{\gamma\in G/G_{\mathbf{x}}}\exp\{2\pi i\mathbf{h}\cdot[S_\gamma(\mathbf{x})]\}\right]\\&=\frac{1}{|\det\mathbf{N}|}\sum_{\mathbf{x}\in D}\rho_{\mathbf{x}}\left[\frac{1}{|G_{\mathbf{x}}|}\sum_{g\in G}\exp\{2\pi i\mathbf{h}\cdot[S_g(\mathbf{x})]\}\right],\\\rho_{\mathbf{x}}&=\sum_{\mathbf{h}\in D^*}F_{\mathbf{h}}\left[\sum_{\gamma\in G/G_{\mathbf{h}}}\exp\{-2\pi i\mathbf{h}\cdot[S_\gamma(\mathbf{x})]\}\right]\\&=\sum_{\mathbf{h}\in D^*}F_{\mathbf{h}}\left[\frac{1}{|G_{\mathbf{h}}|}\sum_{g\in G}\exp\{-2\pi i\mathbf{h}\cdot[S_g(\mathbf{x})]\}\right],\end{aligned}$$

with the obvious alternatives in terms of $\rho_{\mathbf{m}}$, $\mathbf{m}=\mathbf{N}\mathbf{x}$. Our problem is to evaluate the CDFT for a given space group as efficiently as possible, in spite of the fact that the group action has spoilt the simple tensor-product structure of the ordinary three-dimensional DFT (Section 1.3.3.3.1).

Two procedures are available to carry out the 3D summations involved as a succession of smaller summations:

(1) *decomposition* into successive transforms of *fewer dimensions* but on the *same number of points* along these dimensions. This possibility depends on the reducibility of the space group, as defined in Section 1.3.4.2.2.4, and simply invokes the tensor product property of the DFT;

(2) *factorization* of the transform into transforms of the *same number of dimensions* as the original one, but on *fewer points* along each dimension. This possibility depends on the arithmetic factorability of the decimation matrix $\mathbf{N}$, as described in Section 1.3.3.3.2.

Clearly, a symmetry expansion to the largest fully reducible subgroup of the space group will give maximal decomposability, but will require computing more than the unique results from more than the unique data. Economy will follow from factoring the transforms in the subspaces within which the space group acts irreducibly.

For irreducible subspaces of dimension 1, the group action is readily incorporated into the factorization of the transform, as first shown by Ten Eyck (1973).

For irreducible subspaces of dimension 2 or 3, the ease of incorporation of symmetry into the factorization depends on the type of factorization method used. The multidimensional Cooley–Tukey method (Section 1.3.3.3.1) is rather complicated; the multidimensional Good method (Section 1.3.3.3.2.2) is somewhat simpler; and the Rader/Winograd factorization admits a generalization, based on the arithmetic of certain rings of *algebraic*

integers, which accommodates 2D crystallographic symmetries in a most powerful and pleasing fashion.

At each stage of the calculation, it is necessary to keep track of the definition of the asymmetric unit and of the symmetry properties of the numbers being manipulated. This requirement applies not only to the initial data and to the final results, where these are familiar; but also to all the intermediate quantities produced by partial transforms (on subsets of factors, or subsets of dimensions, or both), where they are less familiar. Here, the general formalism of transposition (or 'orbit exchange') described in Section 1.3.4.2.2.2 plays a central role.

1.3.4.3.3. *Interaction between symmetry and decomposition*

Suppose that the space-group action is reducible, *i.e.* that for each $g \in G$

$$\mathbf{R}_g = \begin{pmatrix} \mathbf{R}'_g & \mathbf{0} \\ \mathbf{0} & \mathbf{R}''_g \end{pmatrix}, \qquad \mathbf{t}_g = \begin{pmatrix} \mathbf{t}'_g \\ \mathbf{t}''_g \end{pmatrix};$$

by Schur's lemma, the decimation matrix must then be of the form $\mathbf{N} = \begin{pmatrix} \mathbf{N}' & \mathbf{0} \\ \mathbf{0} & \mathbf{N}'' \end{pmatrix}$ if it is to commute with all the $\mathbf{R}_g$.

Putting $\mathbf{x} = \begin{pmatrix} \mathbf{x}' \\ \mathbf{x}'' \end{pmatrix}$ and $\mathbf{h} = \begin{pmatrix} \mathbf{h}' \\ \mathbf{h}'' \end{pmatrix}$, we may define

$$S'_g(\mathbf{x}') = \mathbf{R}'_g\mathbf{x}' + \mathbf{t}'_g,$$
$$S''_g(\mathbf{x}'') = \mathbf{R}''_g\mathbf{x}'' + \mathbf{t}''_g,$$

and write $S_g = S'_g \oplus S''_g$ (direct sum) as a shorthand for $S_g(\mathbf{x}) = \begin{pmatrix} S'_g(\mathbf{x}') \\ S''_g(\mathbf{x}'') \end{pmatrix}$.

We may also define the representation operators $S'^{\#}_g$ and $S''^{\#}_g$ acting on functions of $\mathbf{x}'$ and $\mathbf{x}''$, respectively (as in Section 1.3.4.2.2.4), and the operators S'^{*}_g and S''^{*}_g acting on functions of $\mathbf{h}'$ and $\mathbf{h}''$, respectively (as in Section 1.3.4.2.2.5). Then we may write

$$S^{\#}_g = (S'_g)^{\#} \oplus (S''_g)^{\#}$$

and

$$S^{*}_g = (S'_g)^{*} \oplus (S''_g)^{*}$$

in the sense that g acts on $f(\mathbf{x}) \equiv f(\mathbf{x}', \mathbf{x}'')$ by

$$(S^{\#}_g f)(\mathbf{x}', \mathbf{x}'') = f[(S'_g)^{-1}(\mathbf{x}'), (S''_g)^{-1}(\mathbf{x}'')]$$

and on $\Phi(\mathbf{h}) \equiv \Phi(\mathbf{h}', \mathbf{h}'')$ by

$$(S^{*}_g \Phi)(\mathbf{h}', \mathbf{h}'') = \exp(2\pi i \mathbf{h}' \cdot \mathbf{t}'_g) \exp(2\pi i \mathbf{h}'' \cdot \mathbf{t}''_g) \times \Phi[\mathbf{R}'^{T}_g \mathbf{h}', \mathbf{R}''^{T}_g \mathbf{h}''].$$

Thus equipped we may now derive concisely a general identity describing the symmetry properties of intermediate quantities of the form

$$T(\mathbf{x}', \mathbf{h}'') = \sum_{\mathbf{h}'} F(\mathbf{h}', \mathbf{h}'') \exp(-2\pi i \mathbf{h}' \cdot \mathbf{x}') = \frac{1}{|\det \mathbf{N}'|} \sum_{\mathbf{x}''} \rho(\mathbf{x}', \mathbf{x}'') \exp(+2\pi i \mathbf{h}'' \cdot \mathbf{x}''),$$

which arise through partial transformation of F on $\mathbf{h}'$ or of ρ on $\mathbf{x}''$. The action of $g \in G$ on these quantities will be

(i) through $(S'_g)^{\#}$ on the function $\mathbf{x}' \longmapsto T(\mathbf{x}', \mathbf{h}'')$,

(ii) through $(S''_g)^{*}$ on the function $\mathbf{h}'' \longmapsto T(\mathbf{x}', \mathbf{h}'')$,

and hence the symmetry properties of T are expressed by the identity

$$T = [(S'_g)^{\#} \oplus (S''_g)^{*}]T.$$

Applying this relation not to T but to $[(S'_{g^{-1}})^{\#} \oplus (S''_e)^{*}]T$ gives

$$[(S'_{g^{-1}})^{\#} \oplus (S''_e)^{*}]T = [(S'_e)^{\#} \oplus (S''_g)^{*}]T,$$

i.e.

$$\boxed{T(S'_g(\mathbf{x}'), \mathbf{h}'') = \exp(2\pi i \mathbf{h}'' \cdot \mathbf{t}''_g) T(\mathbf{x}', \mathbf{R}''^{T}_g \mathbf{h}'').}$$

If the unique $F(\mathbf{h}) \equiv F(\mathbf{h}', \mathbf{h}'')$ were initially indexed by

$$(\text{all } \mathbf{h}') \times (\text{unique } \mathbf{h}'')$$

(see Section 1.3.4.2.2.2), this formula allows the reindexing of the intermediate results $T(\mathbf{x}', \mathbf{h}'')$ from the initial form

$$(\text{all } \mathbf{x}') \times (\text{unique } \mathbf{h}'')$$

to the final form

$$(\text{unique } \mathbf{x}') \times (\text{all } \mathbf{h}''),$$

on which the second transform (on $\mathbf{h}''$) may now be performed, giving the final results $\rho(\mathbf{x}', \mathbf{x}'')$ indexed by

$$(\text{unique } \mathbf{x}') \times (\text{all } \mathbf{x}''),$$

which is an asymmetric unit. An analogous interpretation holds if one is going from ρ to F.

The above formula solves the general problem of transposing from one invariant subspace to another, and is the main device for decomposing the CDFT. Particular instances of this formula were derived and used by Ten Eyck (1973); it is useful for orthorhombic groups, and for dihedral groups containing screw axes n_m with g.c.d. $(m, n) = 1$. For comparison with later uses of orbit exchange, it should be noted that the type of intermediate results just dealt with is obtained after transforming on *all* factors in *one* summand.

A central piece of information for driving such a decomposition is the definition of the full asymmetric unit in terms of the asymmetric units in the invariant subspaces. As indicated at the end of Section 1.3.4.2.2.2, this is straightforward when G acts without fixed points, but becomes more involved if fixed points do exist. To this day, no systematic 'calculus of asymmetric units' exists which can automatically generate a complete description of the asymmetric unit of an arbitrary space group in a form suitable for directing the orbit exchange process, although Shenefelt (1988) has outlined a procedure for dealing with space group *P*622 and its subgroups. The asymmetric unit definitions given in Volume A of *International Tables* are incomplete in this respect, in that they do not specify the possible residual symmetries which may exist on the boundaries of the domains.

1.3.4.3.4. *Interaction between symmetry and factorization*

Methods for factoring the DFT in the absence of symmetry were examined in Sections 1.3.3.2 and 1.3.3.3. They are based on the observation that the finite sets which index both data and results are endowed with certain algebraic structures (*e.g.* are Abelian groups, or rings), and that subsets of indices may be found which are not merely subsets but *substructures* (*e.g.* subgroups or subrings). Summation over these substructures leads to partial transforms, and the way in which substructures fit into the global structure indicates how to reassemble the partial results into the final results. As a rule, the richer the algebraic structure which is identified in the indexing set, the more powerful the factoring method.

The ability of a given factoring method to accommodate crystallographic symmetry will thus be determined by the extent to which the crystallographic group action respects (or fails to respect) the partitioning of the index set into the substructures pertaining to that method. This remark justifies trying to gain an overall view of the algebraic structures involved, and of the possibilities of a crystallographic group acting 'naturally' on them.

The index sets $\{\mathbf{m}|\mathbf{m} \in \mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3\}$ and $\{\mathbf{h}|\mathbf{h} \in \mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3\}$ are finite Abelian groups under component-wise addition. If an iterated addition is viewed as an action of an integer scalar $n \in \mathbb{Z}$ *via*

$$\begin{aligned} n\mathbf{h} &= \mathbf{h} + \mathbf{h} + \ldots + \mathbf{h} \qquad (n \text{ times}) \qquad \text{for } n > 0, \\ &= \mathbf{0} \qquad\qquad\qquad\qquad\quad \text{for } n = 0, \\ &= -(\mathbf{h} + \mathbf{h} + \ldots + \mathbf{h}) \quad (|n| \text{ times}) \quad \text{for } n < 0, \end{aligned}$$

then an Abelian group becomes a *module* over the ring $\mathbb{Z}$ (or, for short, a $\mathbb{Z}$-module), a module being analogous to a vector space but with scalars drawn from a ring rather than a field. The left actions of a crystallographic group G by

$$g: \quad \mathbf{m} \longmapsto \mathbf{R}_g\mathbf{m} + \mathbf{N}\mathbf{t}_g \mod \mathbf{N}\mathbb{Z}^3$$

and by

$$g: \quad \mathbf{h} \longmapsto (\mathbf{R}_g^{-1})^T\mathbf{h} \mod \mathbf{N}^T\mathbb{Z}^3$$

can be combined with this $\mathbb{Z}$ action as follows:

$$\begin{aligned} \textstyle\sum_{g\in G} n_g g: &\quad \mathbf{m} \longmapsto \textstyle\sum_{g\in G} n_g(\mathbf{R}_g\mathbf{m} + \mathbf{N}\mathbf{t}_g) \mod \mathbf{N}\mathbb{Z}^3, \\ \textstyle\sum_{g\in G} n_g g: &\quad \mathbf{h} \longmapsto \textstyle\sum_{g\in G} n_g[(\mathbf{R}_g^{-1})^T\mathbf{h}] \mod \mathbf{N}^T\mathbb{Z}^3. \end{aligned}$$

This provides a left action, on the indexing sets, of the set

$$\mathbb{Z}G = \left\{ \sum_{g\in G} n_g g \middle| n_g \in \mathbb{Z} \text{ for each } g \in G \right\}$$

of symbolic linear combinations of elements of G with integral coefficients. If addition and multiplication are defined in $\mathbb{Z}G$ by

$$\left(\sum_{g_1\in G} a_{g_1}g_1\right) + \left(\sum_{g_2\in G} b_{g_2}g_2\right) = \sum_{g\in G}(a_g + b_g)g$$

and

$$\left(\sum_{g_1\in G} a_{g_1}g_1\right) \times \left(\sum_{g_2\in G} b_{g_2}g_2\right) = \sum_{g\in G} c_g g,$$

with

$$c_g = \sum_{g'\in G} a_{g'} b_{(g')^{-1}g},$$

then $\mathbb{Z}G$ is a *ring*, and the action defined above makes the indexing sets into $\mathbb{Z}G$-*modules*. The ring $\mathbb{Z}G$ is called the *integral group ring* of G (Curtis & Reiner, 1962, p. 44).

From the algebraic standpoint, therefore, the interaction between symmetry and factorization can be expected to be favourable whenever the indexing sets of partial transforms are $\mathbb{Z}G$-submodules of the main $\mathbb{Z}G$-modules.

1.3.4.3.4.1. *Multidimensional Cooley–Tukey factorization*

Suppose, as in Section 1.3.3.3.2.1, that the decimation matrix $\mathbf{N}$ may be factored as $\mathbf{N}_1\mathbf{N}_2$. Then any grid point index $\mathbf{m} \in \mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$ in real space may be written

$$\mathbf{m} = \mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2$$

with $\mathbf{m}_1 \in \mathbb{Z}^3/\mathbf{N}_1\mathbb{Z}^3$ and $\mathbf{m}_2 \in \mathbb{Z}^3/\mathbf{N}_2\mathbb{Z}^3$ determined by

$$\begin{aligned} \mathbf{m}_1 &= \mathbf{m} \mod \mathbf{N}_1\mathbb{Z}^3, \\ \mathbf{m}_2 &= \mathbf{N}_1^{-1}(\mathbf{m} - \mathbf{m}_1) \mod \mathbf{N}_2\mathbb{Z}^3. \end{aligned}$$

These relations establish a one-to-one correspondence $\mathbf{m} \leftrightarrow (\mathbf{m}_1, \mathbf{m}_2)$ between $I = \mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$ and the Cartesian product $I_1 \times I_2$ of $I_1 = \mathbb{Z}^3/\mathbf{N}_1\mathbb{Z}^3$ and $I_2 = \mathbb{Z}^3/\mathbf{N}_2\mathbb{Z}^3$, and hence $I \cong I_1 \times I_2$ as a set. However $I \ncong I_1 \times I_2$ as an Abelian group, since in general $\mathbf{m} + \mathbf{m}' \not\leftrightarrow (\mathbf{m}_1 + \mathbf{m}_1', \mathbf{m}_2 + \mathbf{m}_2')$ because there can be a 'carry' from the addition of the first components into the second components; therefore, $I \ncong I_1 \times I_2$ as a $\mathbb{Z}G$-module, which shows that the incorporation of symmetry into the Cooley–Tukey algorithm is not a trivial matter.

Let $g \in G$ act on I through

$$g: \quad \mathbf{m} \longmapsto S_g(\mathbf{m}) = \mathbf{R}_g\mathbf{m} + \mathbf{N}\mathbf{t}_g \mod \mathbf{N}\mathbb{Z}^3$$

and suppose that $\mathbf{N}$ 'integerizes' all the non-primitive translations $\mathbf{t}_g$ so that we may write

$$\mathbf{N}\mathbf{t}_g = \mathbf{t}_g^{(1)} + \mathbf{N}_1\mathbf{t}_g^{(2)},$$

with $\mathbf{t}_g^{(1)} \in I_1$ and $\mathbf{t}_g^{(2)} \in I_2$ determined as above. Suppose further that $\mathbf{N}$, $\mathbf{N}_1$ and $\mathbf{N}_2$ commute with $\mathbf{R}_g$ for all $g \in G$, *i.e.* (by Schur's lemma, Section 1.3.4.2.2.4) that these matrices are integer multiples of the identity in each G-invariant subspace. The action of g on $\mathbf{m} = \mathbf{N}\mathbf{x} \mod \mathbf{N}\mathbb{Z}^3$ leads to

$$\begin{aligned} S_g(\mathbf{m}) &= \mathbf{N}[\mathbf{R}_g(\mathbf{N}^{-1}\mathbf{m}) + \mathbf{N}\mathbf{t}_g] &&\mod \mathbf{N}\mathbb{Z}^3 \\ &= \mathbf{N}\mathbf{R}_g\mathbf{N}^{-1}(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2) + \mathbf{t}_g^{(1)} + \mathbf{N}_1\mathbf{t}_g^{(2)} &&\mod \mathbf{N}\mathbb{Z}^3 \\ &= \mathbf{R}_g\mathbf{m}_1 + \mathbf{t}_g^{(1)} + \mathbf{N}_1(\mathbf{R}_g\mathbf{m}_2 + \mathbf{t}_g^{(2)}) &&\mod \mathbf{N}\mathbb{Z}^3, \end{aligned}$$

which we may decompose as

$$S_g(\mathbf{m}) = [S_g(\mathbf{m})]_1 + \mathbf{N}_1[S_g(\mathbf{m})]_2$$

with

$$[S_g(\mathbf{m})]_1 \equiv S_g(\mathbf{m}) \mod \mathbf{N}_1\mathbb{Z}^3$$

and

$$[S_g(\mathbf{m})]_2 \equiv \mathbf{N}_1^{-1}\{S_g(\mathbf{m}) - [S_g(\mathbf{m})]_1\} \mod \mathbf{N}_2\mathbb{Z}^3.$$

Introducing the notation

$$\begin{aligned} S_g^{(1)}(\mathbf{m}_1) &= \mathbf{R}_g\mathbf{m}_1 + \mathbf{t}_g^{(1)} \mod \mathbf{N}_1\mathbb{Z}^3, \\ S_g^{(2)}(\mathbf{m}_2) &= \mathbf{R}_g\mathbf{m}_2 + \mathbf{t}_g^{(2)} \mod \mathbf{N}_2\mathbb{Z}^3, \end{aligned}$$

the two components of $S_g(\mathbf{m})$ may be written

$$\begin{aligned} [S_g(\mathbf{m})]_1 &= S_g^{(1)}(\mathbf{m}_1), \\ [S_g(\mathbf{m})]_2 &= S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1) \mod \mathbf{N}_2\mathbb{Z}^3, \end{aligned}$$

with

$$\boldsymbol{\mu}_2(g, \mathbf{m}_1) = \mathbf{N}_1^{-1}\{(\mathbf{R}_g\mathbf{m}_1 + \mathbf{t}_g^{(1)}) - [S_g(\mathbf{m}_1)]_1\} \mod \mathbf{N}_2\mathbb{Z}^3.$$

The term $\boldsymbol{\mu}_2$ is the geometric equivalent of a *carry* or *borrow*: it arises because $\mathbf{R}_g\mathbf{m}_1 + \mathbf{t}_g^{(1)}$, calculated as a vector in $\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$, may be outside the unit cell $\mathbf{N}_1[0, 1]^3$, and may need to be brought back into it by a 'large' translation with a non-zero component in the $\mathbf{m}_2$ space; equivalently, the action of g may need to be applied around different permissible origins for different values of $\mathbf{m}_1$, so as to map the unit cell into itself without any recourse to lattice translations. [Readers familiar with the cohomology of groups (see *e.g.* Hall, 1959; MacLane, 1963) will recognize $\boldsymbol{\mu}_2$ as the cocycle of the extension of $\mathbb{Z}G$-modules described by the exact sequence $0 \to I_2 \to I \to I_1 \to 0$.]

Thus G acts on I in a rather complicated fashion: although $g \longmapsto S_g^{(1)}$ does define a left action in I_1 alone, no action can be defined in I_2 alone because $\boldsymbol{\mu}_2$ depends on $\mathbf{m}_1$. However, because S_g, $S_g^{(1)}$ and $S_g^{(2)}$ *are* left actions, it follows that $\boldsymbol{\mu}_2$ satisfies the identity

$$\boldsymbol{\mu}_2(gg', \mathbf{m}_1) = S_g^{(2)}[\boldsymbol{\mu}_2(g', \mathbf{m}_1)] + \boldsymbol{\mu}_2[g, S_g^{(1)}(\mathbf{m}_1)] \qquad \text{mod } \mathbf{N}_2\mathbb{Z}^3$$

for all g, g' in G and all $\mathbf{m}_1$ in I_1. In particular, $\boldsymbol{\mu}_2(\mathrm{e}, \mathbf{m}_1) = \mathbf{0}$ for all $\mathbf{m}_1$, and

$$\boldsymbol{\mu}_2(g^{-1}, \mathbf{m}_1) = -S_{g^{-1}}^{(2)}\{\boldsymbol{\mu}_2[g, S_{g^{-1}}^{(1)}(\mathbf{m}_1)]\} \text{ mod } \mathbf{N}_2\mathbb{Z}^3.$$

This action will now be used to achieve optimal use of symmetry in the multidimensional Cooley–Tukey algorithm of Section 1.3.3.3.2.1. Let us form an array Y according to

$$Y(\mathbf{m}_1, \mathbf{m}_2) = \rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2)$$

for *all* $\mathbf{m}_2 \in I_2$ but only for the *unique* $\mathbf{m}_1$ under the action $S_g^{(1)}$ of G in I_1. Except in special cases which will be examined later, these vectors contain essentially an asymmetric unit of electron-density data, up to some redundancies on boundaries. We may then compute the partial transform on $\mathbf{m}_2$:

$$Y^*(\mathbf{m}_1, \mathbf{h}_2) = \frac{1}{|\det \mathbf{N}_2|} \sum_{\mathbf{m}_2 \in I_2} Y(\mathbf{m}_1, \mathbf{m}_2) e[\mathbf{h}_2 \cdot (\mathbf{N}_2^{-1}\mathbf{m}_2)].$$

Using the symmetry of ρ in the form $\rho = S_g^{\#}\rho$ yields by the procedure of Section 1.3.3.3.2 the transposition formula

$$Y^*(S_g^{(1)}(\mathbf{m}_1), \mathbf{h}_2) = e\{\mathbf{h}_2 \cdot [\mathbf{N}_2^{-1}(\mathbf{t}_g^{(2)} + \boldsymbol{\mu}_2(g, \mathbf{m}_1))]\} \times Y^*(\mathbf{m}_1, [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2).$$

By means of this identity we can transpose intermediate results Y^* initially indexed by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{h}_2),$$

so as to have them indexed by

$$(\text{all } \mathbf{m}_1) \times (\text{unique } \mathbf{h}_2).$$

We may then apply twiddle factors to get

$$Z(\mathbf{m}_1, \mathbf{h}_2) = e[\mathbf{h}_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] Y^*(\mathbf{m}_1, \mathbf{h}_2)$$

and carry out the second transform

$$Z^*(\mathbf{h}_1, \mathbf{h}_2) = \frac{1}{|\det \mathbf{N}_1|} \sum_{\mathbf{m}_1 \in I_1} Z(\mathbf{m}_1, \mathbf{h}_2) e[\mathbf{h}_1 \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)].$$

The final results are indexed by

$$(\text{all } \mathbf{h}_1) \times (\text{unique } \mathbf{h}_2),$$

which yield essentially an asymmetric unit of structure factors after unscrambling by:

$$F(\mathbf{h}_2 + \mathbf{N}_2^T\mathbf{h}_1) = Z^*(\mathbf{h}_1, \mathbf{h}_2).$$

The transposition formula above applies to intermediate results when going backwards from F to ρ, provided these results are considered *after* the twiddle-factor stage. A transposition formula applicable *before* that stage can be obtained by characterizing the action of G on $\mathbf{h}$ (including the effects of periodization by $\mathbf{N}^T\mathbb{Z}^3$) in a manner similar to that used for $\mathbf{m}$.

Let

$$\mathbf{h} = \mathbf{h}_2 + \mathbf{N}_2^T\mathbf{h}_1,$$

with

$$\mathbf{h}_2 = \mathbf{h} \qquad \text{mod } \mathbf{N}_2^T\mathbb{Z}^3,$$
$$\mathbf{h}_1 = (\mathbf{N}_2^{-1})^T(\mathbf{h} - \mathbf{h}_2) \quad \text{mod } \mathbf{N}_1^T\mathbb{Z}^3.$$

We may then write

$$\mathbf{R}_g^T\mathbf{h} = [\mathbf{R}_g^T\mathbf{h}]_2 + \mathbf{N}_2^T[\mathbf{R}_g^T\mathbf{h}]_1,$$

with

$$[\mathbf{R}_g^T\mathbf{h}]_2 = [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2 \qquad \text{mod } \mathbf{N}_2^T\mathbb{Z}^3,$$
$$[\mathbf{R}_g^T\mathbf{h}]_1 = [\mathbf{R}_g^{(1)}]^T\mathbf{h}_1 + \boldsymbol{\eta}_1(g, \mathbf{h}_2) \quad \text{mod } \mathbf{N}_1^T\mathbb{Z}^3.$$

Here $[\mathbf{R}_g^{(2)}]^T$, $[\mathbf{R}_g^{(1)}]^T$ and $\boldsymbol{\eta}_1$ are defined by

$$[\mathbf{R}_g^{(2)}]^T\mathbf{h}_2 = \mathbf{R}_g^T\mathbf{h} \qquad \text{mod } \mathbf{N}_2^T\mathbb{Z}^3,$$
$$[\mathbf{R}_g^{(1)}]^T\mathbf{h}_1 = \mathbf{R}_g^T\mathbf{h} \qquad \text{mod } \mathbf{N}_1^T\mathbb{Z}^3$$

and

$$\boldsymbol{\eta}_1(g, \mathbf{h}_2) = (\mathbf{N}_2^{-1})^T(\mathbf{R}_g^T\mathbf{h}_2 - [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2) \text{ mod } \mathbf{N}_1^T\mathbb{Z}^3.$$

Let us then form an array Z^* according to

$$Z^*(\mathbf{h}_1', \mathbf{h}_2') = F(\mathbf{h}_2' + \mathbf{N}_2^T\mathbf{h}_1')$$

for *all* $\mathbf{h}_1'$ but only for the *unique* $\mathbf{h}_2'$ under the action of G in $\mathbb{Z}^3/\mathbf{N}_2^T\mathbb{Z}^3$, and transform on $\mathbf{h}_1'$ to obtain

$$Z(\mathbf{m}_1, \mathbf{h}_2) = \sum_{\mathbf{h}_1' \in \mathbb{Z}^3/\mathbf{N}_1^T\mathbb{Z}^3} Z^*(\mathbf{h}_1', \mathbf{h}_2') e[-\mathbf{h}_1' \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)].$$

Putting $\mathbf{h}' = \mathbf{R}_g^T\mathbf{h}$ and using the symmetry of F in the form

$$F(\mathbf{h}') = F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{t}_g),$$

where

$$\mathbf{h} \cdot \mathbf{t}_g = (\mathbf{h}_2^T + \mathbf{h}_1^T\mathbf{N}_2)(\mathbf{N}_2^{-1}\mathbf{N}_1^{-1})(\mathbf{t}_g^{(1)} + \mathbf{N}_1\mathbf{t}_g^{(2)})$$
$$\equiv \mathbf{h}_2 \cdot \mathbf{t}_g + \mathbf{h}_2 \cdot (\mathbf{N}_1^{-1}\mathbf{t}_g^{(1)}) \text{ mod } 1$$

yields by a straightforward rearrangement

$$Z(\mathbf{m}_1, [\mathbf{R}_g^{(2)}]^T\mathbf{h}_2) = e[-\{\mathbf{h}_2 \cdot \mathbf{t}_g + \boldsymbol{\eta}_1(g, \mathbf{h}_2) \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)\}] \times Z\{S_g^{(1)}(\mathbf{m}_1), \mathbf{h}_2\}.$$

This formula allows the transposition of intermediate results Z from an indexing by

$$(\text{all } \mathbf{m}_1) \times (\text{unique } \mathbf{h}_2)$$

to an indexing by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{h}_2).$$

We may then apply the twiddle factors to obtain

$$Y^*(\mathbf{m}_1, \mathbf{h}_2) = e[-\mathbf{h}_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] Z(\mathbf{m}_1, \mathbf{h}_2)$$

and carry out the second transform on $\mathbf{h}_2$

$$Y(\mathbf{m}_1, \mathbf{m}_2) = \sum_{\mathbf{h}_2 \in \mathbb{Z}^3/\mathbf{N}_2^T\mathbb{Z}^3} Y^*(\mathbf{m}_1, \mathbf{h}_2) e[-\mathbf{h}_2 \cdot (\mathbf{N}_2^{-1}\mathbf{m}_2)].$$

The results, indexed by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{m}_2)$$

yield essentially an asymmetric unit of electron densities by the rearrangement

$$\rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2) = Y(\mathbf{m}_1, \mathbf{m}_2).$$

The equivalence of the two transposition formulae up to the intervening twiddle factors is readily established, using the relation

$$\mathbf{h}_2 \cdot [\mathbf{N}_2^{-1}\boldsymbol{\mu}_2(g,\mathbf{m}_1)] = \boldsymbol{\eta}_1(g,\mathbf{h}_2)\cdot(\mathbf{N}_1^{-1}\mathbf{m}_1) \bmod 1$$

which is itself a straightforward consequence of the identity

$$\mathbf{h}\cdot[\mathbf{N}^{-1}S_g(\mathbf{m})] = \mathbf{h}\cdot\mathbf{t}_g + (\mathbf{R}_g^T\mathbf{h})\cdot(\mathbf{N}^{-1}\mathbf{m}).$$

To complete the characterization of the effect of symmetry on the Cooley–Tukey factorization, and of the economy of computation it allows, it remains to consider the possibility that some values of $\mathbf{m}_1$ may be invariant under some transformations $g \in G$ under the action $\mathbf{m}_1 \longmapsto S_g^{(1)}(\mathbf{m}_1)$.

Suppose that $\mathbf{m}_1$ has a non-trivial isotropy subgroup $G_{\mathbf{m}_1}$, and let $g \in G_{\mathbf{m}_1}$. Then each subarray $Y_{\mathbf{m}_1}$ defined by

$$Y_{\mathbf{m}_1}(\mathbf{m}_2) = Y(\mathbf{m}_1,\mathbf{m}_2) = \rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2)$$

satisfies the identity

$$\begin{aligned} Y_{\mathbf{m}_1}(\mathbf{m}_2) &= Y_{S_g^{(1)}(\mathbf{m}_1)}[S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g,\mathbf{m}_1)] \\ &= Y_{\mathbf{m}_1}[S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g,\mathbf{m}_1)] \end{aligned}$$

so that the data for the transform on $\mathbf{m}_2$ have residual symmetry properties. In this case the identity satisfied by $\boldsymbol{\mu}_2$ simplifies to

$$\boldsymbol{\mu}_2(gg',\mathbf{m}_1) = S_g^{(2)}[\boldsymbol{\mu}_2(g',\mathbf{m}_1)] + \boldsymbol{\mu}_2(g,\mathbf{m}_1) \bmod \mathbf{N}_2\mathbb{Z}^3,$$

which shows that the mapping $g \longmapsto \boldsymbol{\mu}_2(g,\mathbf{m}_1)$ satisfies the Frobenius congruences (Section 1.3.4.2.2.3). Thus the internal symmetry of subarray $Y_{\mathbf{m}_1}$ with respect to the action of G on $\mathbf{m}_2$ is given by $G_{\mathbf{m}_1}$ acting on $\mathbb{Z}^3/\mathbf{N}_2\mathbb{Z}^3$ *via*

$$\mathbf{m}_2 \longmapsto S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g,\mathbf{m}_1) \bmod \mathbf{N}_2\mathbb{Z}^3.$$

The transform on $\mathbf{m}_2$ needs only be performed for one out of $[G : G_{\mathbf{m}_1}]$ distinct arrays $Y_{\mathbf{m}_1}$ (results for the others being obtainable by the transposition formula), and this transforms is $G_{\mathbf{m}_1}$-symmetric. In other words, the following cases occur:

(i)	$G_{\mathbf{m}_1} = \{e\}$	maximum saving in computation (by $\lvert G\rvert$); $\mathbf{m}_2$-transform has no symmetry.
(ii)	$G_{\mathbf{m}_1} = G' < G$	saving in computation by a factor of $[G : G']$; $\mathbf{m}_2$-transform is G'-symmetric.
(iii)	$G_{\mathbf{m}_1} = G$	no saving in computation; $\mathbf{m}_2$-transform is G-symmetric.

The symmetry properties of the $\mathbf{m}_2$-transform may themselves be exploited in a similar way if $\mathbf{N}_2$ can be factored as a product of smaller decimation matrices; otherwise, an appropriate symmetrized DFT routine may be provided, using for instance the idea of 'multiplexing/demultiplexing' (Section 1.3.4.3.5). We thus have a recursive *descent procedure*, in which the deeper stages of the recursion deal with transforms on *fewer points*, or of *lower symmetry* (usually both).

The same analysis applies to the $\mathbf{h}_1$-transforms on the subarrays $Z^*_{\mathbf{h}_2}$, and leads to a similar descent procedure.

In conclusion, crystallographic symmetry can be fully exploited to reduce the amount of computation to the minimum required to obtain the unique results from the unique data. No such analysis was so far available in cases where the asymmetric units in real and reciprocal space are not parallelepipeds. An example of this procedure will be given in Section 1.3.4.3.6.5.

1.3.4.3.4.2. *Multidimensional Good factorization*

This procedure was described in Section 1.3.3.3.2.2. The main difference with the Cooley–Tukey factorization is that if $\mathbf{N} = \mathbf{N}_1\mathbf{N}_2\ldots\mathbf{N}_{d-1}\mathbf{N}_d$, where the different factors are pairwise coprime, then the Chinese remainder theorem reindexing makes $\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$ isomorphic to a direct sum.

$$\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3 \cong (\mathbb{Z}^3/\mathbf{N}_1\mathbb{Z}^3)\oplus\ldots\oplus(\mathbb{Z}^3/\mathbf{N}_d\mathbb{Z}^3),$$

where each p-primary piece is endowed with an induced $\mathbb{Z}G$-module structure by letting G operate in the usual way but with the corresponding modular arithmetic. The situation is thus more favourable than with the Cooley–Tukey method, since there is no interference between the factors (no 'carry'). In the terminology of Section 1.3.4.2.2.2, G acts *diagonally* on this direct sum, and results of a partial transform may be transposed by orbit exchange as in Section 1.3.4.3.4.1 but without the extra terms $\boldsymbol{\mu}$ or $\boldsymbol{\eta}$. The analysis of the symmetry properties of partial transforms also carries over, again without the extra terms. Further simplification occurs for all p-primary pieces with p other than 2 or 3, since all non-primitive translations (including those associated to lattice centring) disappear modulo p.

Thus the cost of the CRT reindexing is compensated by the computational savings due to the absence of twiddle factors and of other phase shifts associated with non-primitive translations and with geometric 'carries'.

Within each p-primary piece, however, higher powers of p may need to be split up by a Cooley–Tukey factorization, or carried out directly by a suitably adapted Winograd algorithm.

1.3.4.3.4.3. *Crystallographic extension of the Rader/Winograd factorization*

As was the case in the absence of symmetry, the two previous classes of algorithms can only factor the global transform into partial transforms on prime numbers of points, but cannot break the latter down any further. Rader's idea of using the action of the group of units $U(p)$ to obtain further factorization of a p-primary transform has been used in 'scalar' form by Auslander & Shenefelt (1987), Shenefelt (1988), and Auslander *et al.* (1988). It will be shown here that it can be adapted to the crystallographic case so as to take advantage also of the possible existence of n-fold cyclic symmetry elements ($n = 3, 4, 6$) in a two-dimensional transform (Bricogne & Tolimieri, 1990). This adaptation entails the use of certain rings of *algebraic* integers rather than ordinary integers, whose connection with the handling of cyclic symmetry will now be examined.

Let G be the group associated with a threefold axis of symmetry: $G = \{e, g, g^2\}$ with $g^3 = e$. In a standard trigonal basis, G has matrix representation

$$\mathbf{R}_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}, \quad \mathbf{R}_g = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{R}_{g^2} = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$$

in real space,

$$\mathbf{R}^*_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}, \quad \mathbf{R}^*_g = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{R}^*_{g^2} = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}$$

in reciprocal space. Note that

$$\mathbf{R}^*_{g^2} = [\mathbf{R}_{g^2}^{-1}]^T = \mathbf{R}_g^T,$$

and that

$$\mathbf{R}_g^T = \mathbf{J}^{-1}\mathbf{R}_g\mathbf{J}, \quad \text{where } \mathbf{J} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

so that $\mathbf{R}_g$ and $\mathbf{R}_g^T$ are conjugate in the group of 2×2 unimodular

integer matrices. The group ring $\mathbb{Z}G$ is commutative, and has the structure of the polynomial ring $\mathbb{Z}[X]$ with the single relation $X^2 + X + 1 = 0$ corresponding to the minimal polynomial of $\mathbf{R}_g$. In the terminology of Section 1.3.3.2.4, the ring structure of $\mathbb{Z}G$ is obtained from that of $\mathbb{Z}[X]$ by carrying out polynomial addition and multiplication modulo $X^2 + X + 1$, then replacing X by any generator of G. This type of construction forms the very basis of algebraic number theory [see Artin (1944, Section IIc) for an illustration of this viewpoint], and $\mathbb{Z}G$ as just defined is isomorphic to the ring $\mathbb{Z}[\omega]$ of algebraic integers of the form $a + b\omega$ [$a, b \in \mathbb{Z}, \omega = \exp(2\pi i/3)$] under the identification $X \leftrightarrow \omega$. Addition in this ring is defined component-wise, while multiplication is defined by

$$(a_1 + b_1\omega) \times (a_2 + b_2\omega) = (a_1a_2 - b_1b_2) + [(a_1 - b_1)b_2 + b_1a_2]\omega.$$

In the case of a fourfold axis, $G = \{e, g, g^2, g^3\}$ with $g^4 = e$, and

$$\mathbf{R}_g = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \mathbf{R}_g^*, \quad \text{with again } \mathbf{R}_g^T = \mathbf{J}^{-1}\mathbf{R}_g\mathbf{J}.$$

$\mathbb{Z}G$ is obtained from $\mathbb{Z}[X]$ by carrying out polynomial arithmetic modulo $X^2 + 1$. This identifies $\mathbb{Z}G$ with the ring $\mathbb{Z}[i]$ of Gaussian integers of the form $a + bi$, in which addition takes place component-wise while multiplication is defined by

$$(a_1 + b_1 i) \times (a_2 + b_2 i) = (a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)i.$$

In the case of a sixfold axis, $G = \{e, g, g^2, g^3, g^4, g^5\}$ with $g^6 = e$, and

$$\mathbf{R}_g = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{R}_g^* = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{R}_g^T = \mathbf{J}^{-1}\mathbf{R}_g\mathbf{J}.$$

$\mathbb{Z}G$ is isomorphic to $\mathbb{Z}[\omega]$ under the mapping $g \leftrightarrow 1 + \omega$ since $(1 + \omega)^6 = 1$.

Thus in all cases $\mathbb{Z}G \cong \mathbb{Z}[X]/P(X)$ where $P(X)$ is an irreducible quadratic polynomial with integer coefficients.

The actions of G on lattices in real and reciprocal space (Sections 1.3.4.2.2.4, 1.3.4.2.2.5) extend naturally to actions of $\mathbb{Z}G$ on $\mathbb{Z}^2$ in which an element $z = a + bg$ of $\mathbb{Z}G$ acts *via*

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \longmapsto z\mathbf{m} = (a\mathbf{I} + b\mathbf{R}_g)\begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}$$

in real space, and *via*

$$\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \longmapsto z\mathbf{h} = (a\mathbf{I} + b\mathbf{R}_g^T)\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

in reciprocal space. These two actions are related by conjugation, since

$$(a\mathbf{I} + b\mathbf{R}_g^T) = \mathbf{J}^{-1}(a\mathbf{I} + b\mathbf{R}_g)\mathbf{J}$$

and the following identity (which is fundamental in the sequel) holds:

$$(z\mathbf{h}) \cdot \mathbf{m} = \mathbf{h} \cdot (z\mathbf{m}) \quad \text{for all } \mathbf{m}, \mathbf{h} \in \mathbb{Z}^2.$$

Let us now consider the calculation of a $p \times p$ two-dimensional DFT with n-fold cyclic symmetry ($n = 3, 4, 6$) for an odd prime $p \geq 5$. Denote $\mathbb{Z}/p\mathbb{Z}$ by $\mathbb{Z}_p$. Both the data and the results of the DFT are indexed by $\mathbb{Z}_p \times \mathbb{Z}_p$: hence the action of $\mathbb{Z}G$ on these indices is in fact an action of $\mathbb{Z}_pG$, the latter being obtained from $\mathbb{Z}G$ by carrying out all integer arithmetic in $\mathbb{Z}G$ modulo p. The algebraic structure of $\mathbb{Z}_pG$ combines the symmetry-carrying ring structure of $\mathbb{Z}G$ with the finite field structure of $\mathbb{Z}_p$ used in Section 1.3.3.2.3.1, and holds the key to a symmetry-adapted factorization of the DFT at hand.

The structure of $\mathbb{Z}_pG$ depends on whether $P(X)$ remains irreducible when considered as a polynomial over $\mathbb{Z}_p$. Thus two cases arise:

(1) $P(X)$ remains irreducible mod p, *i.e.* there is no nth root of unity in $\mathbb{Z}_p$;

(2) $P(X)$ factors as $(X - u)(X - v)$, *i.e.* there are nth roots of unity in $\mathbb{Z}_p$.

These two cases require different developments.

Case 1. $\mathbb{Z}_pG$ is a finite field with p^2 elements. There is essentially (*i.e.* up to isomorphism) only one such field, denoted $GF(p^2)$, and its group of units is a cyclic group with $p^2 - 1$ elements. If γ is a generator of this group of units, the input data $\rho_{\mathbf{m}}$ with $\mathbf{m} \neq \mathbf{0}$ may be reordered as

$$\mathbf{m}_0, \gamma\mathbf{m}_0, \gamma^2\mathbf{m}_0, \gamma^3\mathbf{m}_0, \ldots, \gamma^{p^2-2}\mathbf{m}_0$$

by the *real-space action* of γ; while the results $F_{\mathbf{h}}$ with $\mathbf{h} \neq \mathbf{0}$ may be reordered as

$$\mathbf{h}_0, \gamma\mathbf{h}_0, \gamma^2\mathbf{h}_0, \gamma^3\mathbf{h}_0, \ldots, \gamma^{p^2-2}\mathbf{h}_0$$

by the *reciprocal-space action* of γ, where $\mathbf{m}_0$ and $\mathbf{h}_0$ are arbitrary non-zero indices.

The core $\mathbf{C}_{p\times p}$ of the DFT matrix, defined by

$$\mathbf{F}_{p\times p} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & & & \\ \vdots & & \mathbf{C}_{p\times p} & \\ 1 & & & \end{pmatrix},$$

will then have a skew-circulant structure (Section 1.3.3.2.3.1) since

$$(\mathbf{C}_{p\times p})_{jk} = e\left[\frac{(\gamma^j\mathbf{h}_0)\cdot(\gamma^k\mathbf{m}_0)}{p}\right] = e\left[\frac{\mathbf{h}_0\cdot(\gamma^{j+k}\mathbf{m}_0)}{p}\right]$$

depends only on $j + k$. Multiplication by $\mathbf{C}_{p\times p}$ may then be turned into a cyclic convolution of length $p^2 - 1$, which may be factored by two DFTs (Section 1.3.3.2.3.1) or by Winograd's techniques (Section 1.3.3.2.4). The latter factorization is always favourable, as it is easily shown that $p^2 - 1$ is divisible by 24 for any odd prime $p \geq 5$. This procedure is applicable even if no symmetry is present in the data.

Assume now that cyclic symmetry of order $n = 3$, 4 or 6 *is* present. Since n divides 24 hence divides $p^2 - 1$, the generator g of this symmetry is representable as $\gamma^{(p^2-1)/n}$ for a suitable generator γ of the group of units. The reordered data will then be $(p^2 - 1)/n$-periodic rather than simply $(p^2 - 1)$-periodic; hence the reindexed results will be *n-decimated* (Section 1.3.2.7.2), and the $(p^2 - 1)/n$ non-zero results can be calculated by applying the DFT to the $(p^2 - 1)/n$ unique input data. In this way, the n-fold symmetry can be used in full to calculate the core contributions from the unique data to the unique results by a DFT of length $(p^2 - 1)/n$.

It is a simple matter to incorporate non-primitive translations into this scheme. For example, when going from structure factors to electron densities, reordered data items separated by $(p^2 - 1)/n$ are not equal but differ by a phase shift proportional to their index mod p, whose effect is simply to shift the origin of the n-decimated transformed sequence. The same economy of computation can therefore be achieved as in the purely cyclic case.

Dihedral symmetry elements, which map g to g^{-1} (Section 1.3.4.2.2.3), induce extra one-dimensional symmetries of order 2 in the reordered data which can also be fully exploited to reduce computation.

Case 2. If $p \geq 5$, it can be shown that the two roots u and v are always distinct. Then, by the Chinese remainder theorem (CRT) for polynomials (Section 1.3.3.2.4) we have a ring isomorphism

$$\mathbb{Z}_p[X]/P(X) \cong \{\mathbb{Z}_p[X]/(X - u)\} \times \{\mathbb{Z}_p[X]/(X - v)\}$$

defined by sending a polynomial $Q(X)$ from the left-hand-side ring to its two residue classes modulo $X-u$ and $X-v$, respectively. Since the latter are simply the constants $Q(u)$ and $Q(v)$, the CRT reindexing has the particularly simple form

$$a+bX \longmapsto (a+bu, a+bv) = (\alpha, \beta)$$

or equivalently

$$\begin{pmatrix} a \\ b \end{pmatrix} \longmapsto \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \mathbf{M}\begin{pmatrix} a \\ b \end{pmatrix} \bmod p, \quad \text{with } \mathbf{M} = \begin{pmatrix} 1 & u \\ 1 & v \end{pmatrix}.$$

The CRT reconstruction formula similarly simplifies to

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \longmapsto \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{M}^{-1}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \bmod p,$$
$$\text{with } \mathbf{M}^{-1} = \frac{1}{v-u}\begin{pmatrix} v & -u \\ -1 & 1 \end{pmatrix}.$$

The use of the CRT therefore amounts to the *simultaneous diagonalization* (by $\mathbf{M}$) of all the matrices representing the elements of $\mathbb{Z}_pG$ in the basis $(1, X)$.

A first consequence of this diagonalization is that the internal structure of $\mathbb{Z}_pG$ becomes clearly visible. Indeed, $\mathbb{Z}_pG$ is mapped isomorphically to a direct product of two copies of $\mathbb{Z}_p$, in which arithmetic is carried out *component-wise* between eigenvalues α and β. Thus if

$$z = a + bX \overset{\text{CRT}}{\longleftrightarrow} (\alpha, \beta),$$
$$z' = a' + b'X \overset{\text{CRT}}{\longleftrightarrow} (\alpha', \beta'),$$

then

$$z + z' \overset{\text{CRT}}{\longleftrightarrow} (\alpha+\alpha', \beta+\beta'),$$
$$zz' \overset{\text{CRT}}{\longleftrightarrow} (\alpha\alpha', \beta\beta').$$

Taking in particular

$$z \overset{\text{CRT}}{\longleftrightarrow} (\alpha, 0) \neq (0,0),$$
$$z' \overset{\text{CRT}}{\longleftrightarrow} (0, \beta) \neq (0,0),$$

we have $zz' = 0$, so that $\mathbb{Z}_pG$ contains zero divisors; therefore $\mathbb{Z}_pG$ *is not a field.* On the other hand, if $z \overset{\text{CRT}}{\longleftrightarrow} (\alpha, \beta)$ with $\alpha \neq 0$ and $\beta \neq 0$, then α and β belong to the group of units $U(p)$ (Section 1.3.3.2.3.1) and hence have inverses α^{-1} and β^{-1}; it follows that z is a unit in $\mathbb{Z}_pG$, with inverse $z^{-1} \overset{\text{CRT}}{\longleftrightarrow} (\alpha^{-1}, \beta^{-1})$. Therefore, $\mathbb{Z}_pG$ consists of four distinct pieces:

$$0 \overset{\text{CRT}}{\longleftrightarrow} \{(0,0)\},$$
$$D_1 \overset{\text{CRT}}{\longleftrightarrow} \{(\alpha, 0) | \alpha \in U(p)\} \cong U(p),$$
$$D_2 \overset{\text{CRT}}{\longleftrightarrow} \{(0, \beta) | \beta \in U(p)\} \cong U(p),$$
$$U \overset{\text{CRT}}{\longleftrightarrow} \{(\alpha, \beta) | \alpha \in U(p), \beta \in U(p)\} \cong U(p) \times U(p).$$

A second consequence of this diagonalization is that the actions of $\mathbb{Z}_pG$ on indices $\mathbf{m}$ and $\mathbf{h}$ can themselves be brought to diagonal form by basis changes:

$$\mathbf{m} \longmapsto (a\mathbf{I} + b\mathbf{R}_g)\mathbf{m}$$
$$\text{becomes } \boldsymbol{\mu} \longmapsto \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}\boldsymbol{\mu} \quad \text{with } \boldsymbol{\mu} = \mathbf{M}\mathbf{m},$$
$$\mathbf{h} \longmapsto (a\mathbf{I} + b\mathbf{R}_g^T)\mathbf{h}$$
$$\text{becomes } \boldsymbol{\eta} \longmapsto \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}\boldsymbol{\eta} \quad \text{with } \boldsymbol{\eta} = \mathbf{M}\mathbf{J}\mathbf{h}.$$

Thus the sets of indices $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ can be split into four pieces as $\mathbb{Z}_pG$ itself, according as these indices have none, one or two of their coordinates in $U(p)$. These pieces will be labelled by the same symbols – 0, D_1, D_2 and U – as those of $\mathbb{Z}_pG$.

The scalar product $\mathbf{h} \cdot \mathbf{m}$ may be written in terms of $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ as

$$\mathbf{h} \cdot \mathbf{m} = [\boldsymbol{\eta} \cdot ((\mathbf{M}^{-1})^T \mathbf{J} \mathbf{M}^{-1})\boldsymbol{\mu}],$$

and an elementary calculation shows that the matrix $= (\mathbf{M}^{-1})^T\mathbf{J}\mathbf{M}^{-1}$ is *diagonal* by virtue of the relation

$$uv = \text{constant term in } P(X) = 1.$$

Therefore, $\mathbf{h} \cdot \mathbf{m} = 0$ if $\mathbf{h} \in D_1$ and $\boldsymbol{\mu} \in D_2$ or *vice versa.*

We are now in a position to rearrange the DFT matrix $\mathbf{F}_{p\times p}$. Clearly, the structure of $\mathbf{F}_{p\times p}$ is more complex than in case 1, as there are three types of 'core' matrices:

type 1: $D \times D$ (with $D = D_1$ or D_2);
type 2: $D \times U$ or $U \times D$;
type 3: $U \times U$.

(Submatrices of type $D_1 \times D_2$ and $D_2 \times D_1$ have all their elements equal to 1 by the previous remark.)

Let γ be a generator of $U(p)$. We may reorder the elements in D_1, D_2 and U – and hence the data and results indexed by these elements – according to powers of γ. This requires one exponent in each of D_1 and D_2, and two exponents in U. For instance, in the $\mathbf{h}$-index space:

$$D_1 = \left\{ \begin{pmatrix} \gamma & 0 \\ 0 & 0 \end{pmatrix}^j \begin{pmatrix} \eta_1 \\ 0 \end{pmatrix}_0 \Big| \, j = 1, \ldots, p-1 \right\}$$
$$D_2 = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & \gamma \end{pmatrix}^j \begin{pmatrix} 0 \\ \eta_2 \end{pmatrix}_0 \Big| \, j = 1, \ldots, p-1 \right\}$$
$$U = \left\{ \begin{pmatrix} \gamma & 0 \\ 0 & 1 \end{pmatrix}^{j_1} \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}^{j_2} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}_0 \Big| \, j_1 = 1, \ldots, p-1; \; j_2 = 1, \ldots, p-1 \right\}$$

and similarly for the $\boldsymbol{\mu}$ index.

Since the diagonal matrix $\boldsymbol{\Delta}$ commutes with all the matrices representing the action of γ, this rearrangement will induce skew-circulant structures in all the core matrices. The corresponding cyclic convolutions may be carried out by Rader's method, *i.e.* by diagonalizing them by means of two $(p-1)$-point one-dimensional DFTs in the $D \times D$ pieces and of two $(p-1) \times (p-1)$-point two-dimensional DFTs in the $U \times U$ piece (the $U \times D$ and $D \times U$ pieces involve extra section and projection operations).

In the absence of symmetry, no computational saving is achieved, since the same reordering could have been applied to the initial $\mathbb{Z}_p \times \mathbb{Z}_p$ indexing, without the CRT reindexing.

In the presence of n-fold cyclic symmetry, however, the rearranged $\mathbf{F}_{p\times p}$ lends itself to an n-fold reduction in size. The basic fact is that whenever case 2 occurs, $p-1$ is divisible by n (*i.e.* $p-1$ is divisible by 6 when $n = 3$ or 6, and by 4 when $n = 4$), say

$p-1=nq$. If g is a generator of the cyclic symmetry, the generator γ of $U(p)$ may be chosen in such a way that $g=\gamma^q$. The action of g is then to increment the j index in D_1 and D_2 by q, and the (j_1,j_2) index in U by (q,q). Since the data items whose indices are related in this way have identical values, the DFTs used to diagonalize the Rader cyclic convolutions will operate on *periodized data*, hence yield *decimated results*; and the non-zero results will be obtained from the unique data by DFTs n times smaller than their counterparts in the absence of symmetry.

A more thorough analysis is needed to obtain a Winograd factorization into the normal from **CBA** in the presence of symmetry (see Bricogne & Tolimieri, 1990).

Non-primitive translations and dihedral symmetry may also be accommodated within this framework, as in case 1.

This reindexing by means of algebraic integers yields larger orbits, hence more efficient algorithms, than that of Auslander *et al.* (1988) which only uses ordinary integers acting by scalar dilation.

1.3.4.3.5. *Treatment of conjugate and parity-related symmetry properties*

Most crystallographic Fourier syntheses are real-valued and originate from Hermitian-symmetric collections of Fourier coefficients. Hermitian symmetry is closely related to the action of a centre of inversion in reciprocal space, and thus interacts strongly with all other genuinely crystallographic symmetry elements of order 2. All these symmetry properties are best treated by factoring by 2 and reducing the computation of the initial transform to that of a collection of smaller transforms with less symmetry or none at all.

1.3.4.3.5.1. *Hermitian-symmetric or real-valued transforms*

The computation of a DFT with Hermitian-symmetric or real-valued data can be carried out at a cost of half that of an ordinary transform, essentially by 'multiplexing' pairs of special partial transforms into general complex transforms, and then 'demultiplexing' the results on the basis of their symmetry properties. The treatment given below is for general dimension n; a subset of cases for $n=1$ was treated by Ten Eyck (1973).

(*a*) *Underlying group action*

Hermitian symmetry is not a geometric symmetry, but it is defined in terms of the action in reciprocal space of point group $G=\bar{1}$, *i.e.* $G=\{e,-e\}$, where e acts as $\mathbf{I}$ (the $n\times n$ identity matrix) and $-e$ acts as $-\mathbf{I}$.

This group action on $\mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n$ with $\mathbf{N}=\mathbf{N}_1\mathbf{N}_2$ will now be characterized by the calculation of the cocycle $\boldsymbol{\eta}_1$ (Section 1.3.4.3.4.1) under the assumption that $\mathbf{N}_1$ and $\mathbf{N}_2$ are both *diagonal.* For this purpose it is convenient to associate to any integer vector $\mathbf{v}=\begin{pmatrix}v_1\\ \vdots\\ v_n\end{pmatrix}$ in $\mathbb{Z}^n$ the vector $\boldsymbol{\zeta}(\mathbf{v})$ whose jth component is

$$\begin{cases}0 \text{ if} & v_j=0\\ 1 \text{ if} & v_j\neq 0.\end{cases}$$

Let $\mathbf{m}=\mathbf{m}_1+\mathbf{N}_1\mathbf{m}_2$, and hence $\mathbf{h}=\mathbf{h}_2+\mathbf{N}_2\mathbf{h}_1$. Then

$$-\mathbf{h}_2 \bmod \mathbf{N}\mathbb{Z}^n=\mathbf{N}\boldsymbol{\zeta}(\mathbf{h}_2)-\mathbf{h}_2,$$
$$-\mathbf{h}_2 \bmod \mathbf{N}_2\mathbb{Z}^n=\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2)-\mathbf{h}_2,$$

hence

$$\begin{aligned}\boldsymbol{\eta}_1(-e,\mathbf{h}_2)&=\mathbf{N}_2^{-1}\{[\mathbf{N}\boldsymbol{\zeta}(\mathbf{h}_2)-\mathbf{h}_2]-[\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2)-\mathbf{h}_2]\} \bmod \mathbf{N}_1\mathbb{Z}^n\\ &=-\boldsymbol{\zeta}(\mathbf{h}_2) \bmod \mathbf{N}_1\mathbb{Z}^n.\end{aligned}$$

Therefore $-e$ acts by

$$(\mathbf{h}_2,\mathbf{h}_1)\longmapsto[\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2)-\mathbf{h}_2,\mathbf{N}_1\boldsymbol{\zeta}(\mathbf{h}_1)-\mathbf{h}_1-\boldsymbol{\zeta}(\mathbf{h}_2)].$$

Hermitian symmetry is traditionally dealt with by factoring by 2, *i.e.* by assuming $\mathbf{N}=2\mathbf{M}$. If $\mathbf{N}_2=2\mathbf{I}$, then each $\mathbf{h}_2$ is invariant under G, so that each partial vector $\mathbf{Z}^*_{\mathbf{h}_2}$ (Section 1.3.4.3.4.1) inherits the symmetry internally, with a 'modulation' by $\boldsymbol{\eta}_1(g,\mathbf{h}_2)$. The 'multiplexing–demultiplexing' technique provides an efficient treatment of this singular case.

(*b*) *Calculation of structure factors*

The computation may be summarized as follows:

$$\rho \overset{\mathbf{dec}(\mathbf{N}_1)}{\longmapsto} \mathbf{Y} \overset{\bar{F}(\mathbf{N}_2)}{\longmapsto} \mathbf{Y}^* \overset{\text{TW}}{\longmapsto} \mathbf{Z} \overset{\bar{F}(\mathbf{N}_1)}{\longmapsto} \mathbf{Z}^* \overset{\mathbf{rev}(\mathbf{N}_2)}{\longmapsto} \mathbf{F}$$

where $\mathbf{dec}(\mathbf{N}_1)$ is the initial decimation given by $Y_{\mathbf{m}_1}(\mathbf{m}_2)=\rho(\mathbf{m}_1+\mathbf{N}_1\mathbf{m}_2)$, TW is the transposition and twiddle-factor stage, and $\mathbf{rev}(\mathbf{N}_2)$ is the final unscrambling by coset reversal given by $F(\mathbf{h}_2+\mathbf{N}_2\mathbf{h}_1)=\mathbf{Z}^*_{\mathbf{h}_2}(\mathbf{h}_1)$.

(i) *Decimation in time* $(\mathbf{N}_1=2\mathbf{I},\mathbf{N}_2=\mathbf{M})$

The decimated vectors $\mathbf{Y}_{\mathbf{m}_1}$ are real and hence have Hermitian transforms $\mathbf{Y}^*_{\mathbf{m}_1}$. The 2^n values of $\mathbf{m}_1$ may be grouped into 2^{n-1} pairs $(\mathbf{m}_1',\mathbf{m}_1'')$ and the vectors corresponding to each pair may be multiplexed into a general complex vector

$$\mathbf{Y}=\mathbf{Y}_{\mathbf{m}_1'}+i\mathbf{Y}_{\mathbf{m}_1''}.$$

The transform $\mathbf{Y}^*=\bar{F}(\mathbf{M})[\mathbf{Y}]$ can then be resolved into the separate transforms $\mathbf{Y}^*_{\mathbf{m}_1'}$ and $\mathbf{Y}^*_{\mathbf{m}_1''}$ by using the Hermitian symmetry of the latter, which yields the demultiplexing formulae

$$Y^*_{\mathbf{m}_1'}(\mathbf{h}_2)+iY^*_{\mathbf{m}_1''}(\mathbf{h}_2)=Y^*(\mathbf{h}_2)$$
$$\overline{Y^*_{\mathbf{m}_1'}(\mathbf{h}_2)}+i\overline{Y^*_{\mathbf{m}_1''}(\mathbf{h}_2)}=Y^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_2)-\mathbf{h}_2].$$

The number of partial transforms $\bar{F}(\mathbf{M})$ is thus reduced from 2^n to 2^{n-1}. Once this separation has been achieved, the remaining steps need only be carried out for a unique half of the values of $\mathbf{h}_2$.

(ii) *Decimation in frequency* $(\mathbf{N}_1=\mathbf{M},\mathbf{N}_2=2\mathbf{I})$

Since $\mathbf{h}_2\in\mathbb{Z}^n/2\mathbb{Z}^n$ we have $-\mathbf{h}_2=\mathbf{h}_2$ and $\boldsymbol{\zeta}(\mathbf{h}_2)=\mathbf{h}_2 \bmod 2\mathbb{Z}^n$. The vectors of decimated and scrambled results $\mathbf{Z}^*_{\mathbf{h}_2}$ then obey the symmetry relations

$$Z^*_{\mathbf{h}_2}(\mathbf{h}_1-\mathbf{h}_2)=\overline{Z^*_{\mathbf{h}_2}[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1)-\mathbf{h}_1]}$$

which can be used to halve the number of $\bar{F}(\mathbf{M})$ necessary to compute them, as follows.

Having formed the vectors $\mathbf{Z}_{\mathbf{h}_2}$ given by

$$Z_{\mathbf{h}_2}(\mathbf{m}_1)=\left[\sum_{\mathbf{m}_2\in\mathbb{Z}^n/2\mathbb{Z}^n}\frac{(-1)^{\mathbf{h}_2\cdot\mathbf{m}_2}}{2^n}\rho(\mathbf{m}_1+\mathbf{M}\mathbf{m}_2)\right]e[\mathbf{h}_2\cdot(\mathbf{N}^{-1}\mathbf{m}_1)],$$

we may group the 2^n values of $\mathbf{h}_2$ into 2^{n-1} pairs $(\mathbf{h}_2',\mathbf{h}_2'')$ and for each pair form the multiplexed vector:

$$\mathbf{Z}=\mathbf{Z}_{\mathbf{h}_2'}+i\mathbf{Z}_{\mathbf{h}_2''}.$$

After calculating the 2^{n-1} transforms $\mathbf{Z}^*=\bar{F}(\mathbf{M})[\mathbf{Z}]$, the 2^n individual transforms $\mathbf{Z}^*_{\mathbf{h}_2'}$ and $\mathbf{Z}^*_{\mathbf{h}_2''}$ can be separated by using for each pair the demultiplexing formulae

$$Z^*_{\mathbf{h}_2'}(\mathbf{h}_1)+iZ^*_{\mathbf{h}_2''}(\mathbf{h}_1)=Z^*(\mathbf{h}_1)$$
$$Z^*_{\mathbf{h}_2'}(\mathbf{h}_1-\mathbf{h}_2')+iZ^*_{\mathbf{h}_2''}(\mathbf{h}_1-\mathbf{h}_2'')=\overline{Z^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1)-\mathbf{h}_1]}$$

which can be solved recursively. If all pairs are chosen so that they differ only in the jth coordinate $(\mathbf{h}_2)_j$, the recursion is along $(\mathbf{h}_1)_j$ and can be initiated by introducing the (real) values of $Z^*_{\mathbf{h}_2'}$ and $Z^*_{\mathbf{h}_2''}$ at $(\mathbf{h}_1)_j=0$ and $(\mathbf{h}_1)_j=M_j$, accumulated *e.g.* while forming $\mathbf{Z}$ for that pair. Only points with $(\mathbf{h}_1)_j$ going from 0 to $\frac{1}{2}M_j$ need be resolved,

and they contain the unique half of the Hermitian-symmetric transform $\mathbf{F}$.

(*c*) *Calculation of electron densities*

The computation may be summarized as follows:

$$\mathbf{F} \overset{\mathbf{scr}(\mathbf{N}_2)}{\longmapsto} \mathbf{Z}^* \overset{F(\mathbf{N}_1)}{\longmapsto} \mathbf{Z} \overset{\mathrm{TW}}{\longmapsto} \mathbf{Y}^* \overset{F(\mathbf{N}_2)}{\longmapsto} \mathbf{Y} \overset{\mathbf{nat}(\mathbf{N}_1)}{\longmapsto} \rho$$

where $\mathbf{scr}(\mathbf{N}_2)$ is the decimation with coset reversal given by $\mathbf{Z}^*_{\mathbf{h}_2}(\mathbf{h}_1) = F(\mathbf{h}_2 + \mathbf{N}_2\mathbf{h}_1)$, TW is the transposition and twiddle-factor stage, and $\mathbf{nat}(\mathbf{N}_1)$ is the recovery in natural order given by $\rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2) = Y_{\mathbf{m}_1}(\mathbf{m}_2)$.

(i) *Decimation in time* $(\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I})$

The last transformation $F(2\mathbf{I})$ has a real-valued matrix, and the final result ρ is real-valued. It follows that the vectors $\mathbf{Y}^*_{\mathbf{m}_1}$ of intermediate results after the twiddle-factor stage are real-valued, hence lend themselves to multiplexing along the real and imaginary components of half as many general complex vectors.

Let the 2^n initial vectors $\mathbf{Z}^*_{\mathbf{h}_2}$ be multiplexed into 2^{n-1} vectors

$$\mathbf{Z}^* = \mathbf{Z}^*_{\mathbf{h}'_2} + i\mathbf{Z}^*_{\mathbf{h}''_2}$$

[one for each pair $(\mathbf{h}'_2, \mathbf{h}''_2)$], each of which yields by $F(\mathbf{M})$ a vector

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}'_2} + i\mathbf{Z}_{\mathbf{h}''_2}.$$

The real-valuedness of the $\mathbf{Y}^*_{\mathbf{m}_1}$ may be used to recover the separate result vectors for $\mathbf{h}'_2$ and $\mathbf{h}''_2$. For this purpose, introduce the abbreviated notation

$$\begin{aligned} e[-\mathbf{h}'_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] &= (c' + is')(\mathbf{m}_1) \\ e[-\mathbf{h}''_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] &= (c'' + is'')(\mathbf{m}_1) \\ R_{\mathbf{h}_2}(\mathbf{m}_1) &= Y^*_{\mathbf{m}_1}(\mathbf{h}_2) \\ \mathbf{R}' = \mathbf{R}_{\mathbf{h}'_2}, &\quad \mathbf{R}'' = \mathbf{R}_{\mathbf{h}''_2}. \end{aligned}$$

Then we may write

$$\begin{aligned} \mathbf{Z} &= (c' + is')\mathbf{R}' + i(c'' + is'')\mathbf{R}'' \\ &= (c'\mathbf{R}' + s''\mathbf{R}'') + i(-s'\mathbf{R}' + c''\mathbf{R}'') \end{aligned}$$

or, equivalently, for each $\mathbf{m}_1$,

$$\begin{pmatrix} \mathscr{R}e\, Z \\ \mathscr{I}m\, Z \end{pmatrix} = \begin{pmatrix} c' & s'' \\ -s' & c'' \end{pmatrix} \begin{pmatrix} R' \\ R'' \end{pmatrix}.$$

Therefore $\mathbf{R}'$ and $\mathbf{R}''$ may be retrieved from $\mathbf{Z}$ by the 'demultiplexing' formula:

$$\begin{pmatrix} R' \\ R'' \end{pmatrix} = \frac{1}{c'c'' + s's''} \begin{pmatrix} c'' & -s'' \\ s' & c' \end{pmatrix} \begin{pmatrix} \mathscr{R}e\, Z \\ \mathscr{I}m\, Z \end{pmatrix}$$

which is valid at all points $\mathbf{m}_1$ where $c'c'' + s's'' \neq 0$, *i.e.* where

$$\cos[2\pi(\mathbf{h}'_2 - \mathbf{h}''_2) \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] \neq 0.$$

Demultiplexing fails when

$$(\mathbf{h}'_2 - \mathbf{h}''_2) \cdot (\mathbf{N}^{-1}\mathbf{m}_1) = \tfrac{1}{2} \bmod 1.$$

If the pairs $(\mathbf{h}'_2, \mathbf{h}''_2)$ are chosen so that their members differ only in one coordinate (the jth, say), then the exceptional points are at $(\mathbf{m}_1)_j = \frac{1}{2}M_j$ and the missing transform values are easily obtained *e.g.* by accumulation while forming $\mathbf{Z}^*$.

The final stage of the calculation is then

$$\rho(\mathbf{m}_1 + \mathbf{M}\mathbf{m}_2) = \sum_{\mathbf{h}_2 \in \mathbf{Z}^n/2\mathbf{Z}^n} (-1)^{\mathbf{h}_2 \cdot \mathbf{m}_2} R_{\mathbf{h}_2}(\mathbf{m}_1).$$

(ii) *Decimation in frequency* $(\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M})$

The last transformation $F(\mathbf{M})$ gives the real-valued results ρ, therefore the vectors $\mathbf{Y}^*_{\mathbf{m}_1}$ after the twiddle-factor stage each have Hermitian symmetry.

A first consequence is that the intermediate vectors $\mathbf{Z}_{\mathbf{h}_2}$ need only be computed for the unique half of the values of $\mathbf{h}_2$, the other half being related by the Hermitian symmetry of $\mathbf{Y}^*_{\mathbf{m}_1}$.

A second consequence is that the 2^n vectors $\mathbf{Y}^*_{\mathbf{m}_1}$ may be condensed into 2^{n-1} general complex vectors

$$\mathbf{Y}^* = \mathbf{Y}^*_{\mathbf{m}'_1} + i\mathbf{Y}^*_{\mathbf{m}''_1}$$

[one for each pair $(\mathbf{m}'_1, \mathbf{m}''_1)$] to which a general complex $F(\mathbf{M})$ may be applied to yield

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}'_1} + i\mathbf{Y}_{\mathbf{m}''_1}$$

with $\mathbf{Y}_{\mathbf{m}'_1}$ and $\mathbf{Y}_{\mathbf{m}''_1}$ *real-valued.* The final results can therefore be retrieved by the particularly simple demultiplexing formulae:

$$\begin{aligned} \rho(\mathbf{m}'_1 + 2\mathbf{m}_2) &= \mathscr{R}e\, Y(\mathbf{m}_2), \\ \rho(\mathbf{m}''_1 + 2\mathbf{m}_2) &= \mathscr{I}m\, Y(\mathbf{m}_2). \end{aligned}$$

1.3.4.3.5.2. *Hermitian-antisymmetric or pure imaginary transforms*

A vector $\mathbf{X} = \{X(\mathbf{k}) | \mathbf{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n\}$ is said to be Hermitian-antisymmetric if

$$X(\mathbf{k}) = -\overline{X(-\mathbf{k})} \text{ for all } \mathbf{k}.$$

Its transform $\mathbf{X}^*$ then satisfies

$$X^*(\mathbf{k}^*) = -\overline{X^*(\mathbf{k}^*)} \text{ for all } \mathbf{k}^*,$$

i.e. is purely imaginary.

If $\mathbf{X}$ is Hermitian-antisymmetric, then $\mathbf{F} = \pm i\mathbf{X}$ is Hermitian-symmetric, with $\rho = \pm i\mathbf{X}^*$ real-valued. The treatment of Section 1.3.4.3.5.1 may therefore be adapted, with trivial factors of i or -1, or used as such in conjunction with changes of variable by multiplication by $\pm i$.

1.3.4.3.5.3. *Complex symmetric and antisymmetric transforms*

The matrix $-\mathbf{I}$ is its own contragredient, and hence (Section 1.3.2.4.2.2) the transform of a symmetric (respectively antisymmetric) function is symmetric (respectively antisymmetric). In this case the group $G = \{e, -e\}$ acts in both real and reciprocal space as $\{\mathbf{I}, -\mathbf{I}\}$. If $\mathbf{N} = \mathbf{N}_1\mathbf{N}_2$ with both factors diagonal, then $-e$ acts by

$$\begin{aligned} (\mathbf{m}_1, \mathbf{m}_2) &\longmapsto [\mathbf{N}_1\boldsymbol{\zeta}(\mathbf{m}_1) - \mathbf{m}_1, \mathbf{N}_2\boldsymbol{\zeta}(\mathbf{m}_2) - \mathbf{m}_2 - \boldsymbol{\zeta}(\mathbf{m}_1)], \\ (\mathbf{h}_2, \mathbf{h}_1) &\longmapsto [\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2, \mathbf{N}_1\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1 - \boldsymbol{\zeta}(\mathbf{h}_2)], \end{aligned}$$

i.e.

$$\begin{aligned} \boldsymbol{\mu}_2(-e, \mathbf{m}_1) &= -\boldsymbol{\zeta}(\mathbf{m}_1) \bmod \mathbf{N}_2\mathbb{Z}^n, \\ \boldsymbol{\eta}_1(-e, \mathbf{h}_2) &= -\boldsymbol{\zeta}(\mathbf{h}_2) \bmod \mathbf{N}_1\mathbb{Z}^n. \end{aligned}$$

The symmetry or antisymmetry properties of $\mathbf{X}$ may be written

$$X(-\mathbf{m}) = -\varepsilon X(\mathbf{m}) \text{ for all } \mathbf{m},$$

with $\varepsilon = +1$ for symmetry and $\varepsilon = -1$ for antisymmetry.

The computation will be summarized as

$$\mathbf{X} \overset{\mathbf{dec}(\mathbf{N}_1)}{\longmapsto} \mathbf{Y} \overset{\bar{F}(\mathbf{N}_2)}{\longmapsto} \mathbf{Y}^* \overset{\mathrm{TW}}{\longmapsto} \mathbf{Z} \overset{\bar{F}(\mathbf{N}_1)}{\longmapsto} \mathbf{Z}^* \overset{\mathbf{rev}(\mathbf{N}_2)}{\longmapsto} \mathbf{X}^*$$

with the same indexing as that used for structure-factor calculation. In both cases it will be shown that a transform $F(\mathbf{N})$ with $\mathbf{N} = \mathbf{2M}$ and $\mathbf{M}$ diagonal can be computed using only 2^{n-1} partial transforms $F(\mathbf{M})$ instead of 2^n.

(i) *Decimation in time* $(\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M})$

Since $\mathbf{m}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n$ we have $-\mathbf{m}_1 = \mathbf{m}_1$ and $\boldsymbol{\zeta}(\mathbf{m}_1) = \mathbf{m}_1 \bmod 2\mathbb{Z}^n$, so that the symmetry relations for each parity class of data $\mathbf{Y}_{\mathbf{m}_1}$ read

$$Y_{\mathbf{m}_1}[\mathbf{M}\boldsymbol{\zeta}(\mathbf{m}_2) - \mathbf{m}_2 - \mathbf{m}_1] = \varepsilon Y_{\mathbf{m}_1}(\mathbf{m}_2)$$

or equivalently

$$\tau_{\mathbf{m}_1}\mathbf{Y}_{\mathbf{m}_1} = \varepsilon\breve{\mathbf{Y}}_{\mathbf{m}_1}.$$

Transforming by $F(\mathbf{M})$, this relation becomes

$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}_1)]\mathbf{Y}^*_{\mathbf{m}_1} = \varepsilon\mathbf{Y}^*_{\mathbf{m}_1}.$$

Each parity class thus obeys a different symmetry relation, so that we may multiplex them in pairs by forming for each pair $(\mathbf{m}'_1, \mathbf{m}''_1)$ the vector

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}'_1} + \mathbf{Y}_{\mathbf{m}''_1}.$$

Putting

$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}'_1)] = (c' + is')(\mathbf{h}_2)$$
$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}''_1)] = (c'' + is'')(\mathbf{h}_2)$$

we then have the demultiplexing relations for each $\mathbf{h}_2$:

$$Y^*_{\mathbf{m}'_1}(\mathbf{h}_2) + Y^*_{\mathbf{m}''_1}(\mathbf{h}_2) = Y^*(\mathbf{h}_2)$$
$$(c' + is')(\mathbf{h}_2)Y^*_{\mathbf{m}'_1}(\mathbf{h}_2) + (c'' + is'')(\mathbf{h}_2)Y^*_{\mathbf{m}''_1}(\mathbf{h}_2) = \varepsilon Y^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2]$$

which can be solved recursively. Transform values at the exceptional points $\mathbf{h}_2$ where demultiplexing fails (*i.e.* where $c' + is' = c'' + is''$) can be accumulated while forming $\mathbf{Y}$.

Only the unique half of the values of $\mathbf{h}_2$ need to be considered at the demultiplexing stage and at the subsequent TW and $F(2\mathbf{I})$ stages.

(ii) *Decimation in frequency* $(\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I})$

The vectors of final results $\mathbf{Z}^*_{\mathbf{h}_2}$ for each parity class $\mathbf{h}_2$ obey the symmetry relations

$$\tau_{\mathbf{h}_2}\mathbf{Z}^*_{\mathbf{h}_2} = \varepsilon\check{\mathbf{Z}}^*_{\mathbf{h}_2},$$

which are different for each $\mathbf{h}_2$. The vectors $\mathbf{Z}_{\mathbf{h}_2}$ of intermediate results after the twiddle-factor stage may then be multiplexed in pairs as

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}'_2} + \mathbf{Z}_{\mathbf{h}''_2}.$$

After transforming by $F(\mathbf{M})$, the results $\mathbf{Z}^*$ may be demultiplexed by using the relations

$$Z^*_{\mathbf{h}'_2}(\mathbf{h}_1) + Z^*_{\mathbf{h}''_2}(\mathbf{h}_1) = Z^*(\mathbf{h}_1)$$
$$Z^*_{\mathbf{h}'_2}(\mathbf{h}_1 - \mathbf{h}'_2) + Z^*_{\mathbf{h}''_2}(\mathbf{h}_1 - \mathbf{h}''_2) = \varepsilon Z^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1]$$

which can be solved recursively as in Section 1.3.4.3.5.1(*b*)(ii).

1.3.4.3.5.4. *Real symmetric transforms*

Conjugate symmetric (Section 1.3.2.4.2.3) implies that if the data $\mathbf{X}$ are real and symmetric [*i.e.* $X(\mathbf{k}) = \overline{X(\mathbf{k})}$ and $X(-\mathbf{k}) = X(\mathbf{k})$], then so are the results $\mathbf{X}^*$. Thus if $\mathcal{P}$ contains a centre of symmetry, $\mathbf{F}$ is real symmetric. There is no distinction (other than notation) between structure-factor and electron-density calculation; the algorithms will be described in terms of the former. It will be shown that if $\mathbf{N} = 2\mathbf{M}$, a real symmetric transform can be computed with only 2^{n-2} partial transforms $F(\mathbf{M})$ instead of 2^n.

(i) *Decimation in time* $(\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M})$

Since $\mathbf{m}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n$ we have $-\mathbf{m}_1 = \mathbf{m}_1$ and $\boldsymbol{\zeta}(\mathbf{m}_1) = \mathbf{m}_1 \bmod 2\mathbb{Z}^n$. The decimated vectors $\mathbf{Y}_{\mathbf{m}_1}$ are not only real, but have an internal symmetry expressed by

$$\mathbf{Y}_{\mathbf{m}_1}[\mathbf{M}\boldsymbol{\zeta}(\mathbf{m}_2) - \mathbf{m}_2 - \mathbf{m}_1] = \varepsilon\mathbf{Y}_{\mathbf{m}_1}(\mathbf{m}_2).$$

This symmetry, however, is different for each $\mathbf{m}_1$ so that we may multiplex two such vectors $\mathbf{Y}_{\mathbf{m}'_1}$ and $\mathbf{Y}_{\mathbf{m}''_1}$ into a general *real* vector

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}'_1} + \mathbf{Y}_{\mathbf{m}''_1},$$

for each of the 2^{n-1} pairs $(\mathbf{m}'_1, \mathbf{m}''_1)$. The 2^{n-1} Hermitian-symmetric transform vectors

$$\mathbf{Y}^* = \mathbf{Y}^*_{\mathbf{m}'_1} + \mathbf{Y}^*_{\mathbf{m}''_1}$$

can then be evaluated by the methods of Section 1.3.4.3.5.1(*b*) at the cost of only 2^{n-2} general complex $F(\mathbf{M})$.

The demultiplexing relations by which the separate vectors $\mathbf{Y}^*_{\mathbf{m}'_1}$ and $\mathbf{Y}^*_{\mathbf{m}''_1}$ may be recovered are most simply obtained by observing that the vectors $\mathbf{Z}$ after the twiddle-factor stage are real-valued since $F(2\mathbf{I})$ has a real matrix. Thus, as in Section 1.3.4.3.5.1(*c*)(i),

$$\mathbf{Y}^*_{\mathbf{m}'_1} = (c' - is')\mathbf{R}'$$
$$\mathbf{Y}^*_{\mathbf{m}''_1} = (c'' - is'')\mathbf{R}'',$$

where $\mathbf{R}'$ and $\mathbf{R}''$ are real vectors and where the multipliers $(c' - is')$ and $(c'' - is'')$ are the inverse twiddle factors. Therefore,

$$\mathbf{Y}^* = (c' - is')\mathbf{R}' + (c'' - is'')\mathbf{R}''$$
$$= (c'\mathbf{R}' + c''\mathbf{R}'') - i(s'\mathbf{R}' + s''\mathbf{R}'')$$

and hence the demultiplexing relation for each $\mathbf{h}_2$:

$$\begin{pmatrix} R' \\ R'' \end{pmatrix} = \frac{1}{c's'' - s'c''}\begin{pmatrix} s'' & -c'' \\ -s' & c' \end{pmatrix}\begin{pmatrix} \mathscr{R}e\ Y^* \\ -\mathscr{I}m\ Y^* \end{pmatrix}.$$

The values of $R'_{\mathbf{h}_2}$ and $R''_{\mathbf{h}_2}$ at those points $\mathbf{h}_2$ where $c's'' - s'c'' = 0$ can be evaluated directly while forming $\mathbf{Y}$. This demultiplexing and the final stage of the calculation, namely

$$F(\mathbf{h}_2 + \mathbf{M}\mathbf{h}_1) = \frac{1}{2^n}\sum_{\mathbf{m}_1 \in \mathbf{Z}^n/2\mathbf{Z}^n}(-1)^{\mathbf{h}_1 \cdot \mathbf{m}_1}R_{\mathbf{m}_1}(\mathbf{h}_2)$$

need only be carried out for the unique half of the range of $\mathbf{h}_2$.

(ii) *Decimation in frequency* $(\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I})$

Similarly, the vectors $\mathbf{Z}^*_{\mathbf{h}_2}$ of decimated and scrambled results are real and obey internal symmetries

$$\tau_{\mathbf{h}_2}\mathbf{Z}^*_{\mathbf{h}_2} = \varepsilon\breve{\mathbf{Z}}^*_{\mathbf{h}_2}$$

which are different for each $\mathbf{h}_2$. For each of the 2^{n-1} pairs $(\mathbf{h}'_2, \mathbf{h}''_2)$ the multiplexed vector

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}'_2} + \mathbf{Z}_{\mathbf{h}''_2}$$

is a Hermitian-symmetric vector without internal symmetry, and the 2^{n-1} real vectors

$$\mathbf{Z}^* = \mathbf{Z}^*_{\mathbf{h}'_2} + \mathbf{Z}^*_{\mathbf{h}''_2}$$

may be evaluated at the cost of only 2^{n-2} general complex $F(\mathbf{M})$ by the methods of Section 1.3.4.3.5.1(*c*). The individual transforms $\mathbf{Z}_{\mathbf{h}'_2}$ and $\mathbf{Z}_{\mathbf{h}''_2}$ may then be retrieved *via* the demultiplexing relations

$$Z^*_{\mathbf{h}'_2}(\mathbf{h}_1) + Z^*_{\mathbf{h}''_2}(\mathbf{h}_1) = Z^*(\mathbf{h}_1)$$
$$Z^*_{\mathbf{h}'_2}(\mathbf{h}_1 - \mathbf{h}'_2) + Z^*_{\mathbf{h}''_2}(\mathbf{h}_1 - \mathbf{h}''_2) = Z^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1]$$

which can be solved recursively as described in Section 1.3.4.3.5.1(*b*)(ii). This yields the unique half of the real symmetric results $\mathbf{F}$.

1.3.4.4.3.5.5. *Real antisymmetric transforms*

If $\mathbf{X}$ is real antisymmetric, then its transform $\mathbf{X}^*$ is purely imaginary and antisymmetric. The double-multiplexing techniques used for real symmetric transforms may therefore be adapted with only minor changes involving signs and factors of i.

1.3.4.4.3.5.6. *Generalized multiplexing*

So far the multiplexing technique has been applied to pairs of vectors with similar types of parity-related and/or conjugate symmetry properties, in particular the same value of ε.

It can be generalized so as to accommodate mixtures of vectors with different symmetry characteristics. For example if $\mathbf{X}_1$ is Hermitian-symmetric and $\mathbf{X}_2$ is Hermitian-antisymmetric, so that $\mathbf{X}_1^*$ is real-valued while $\mathbf{X}_2^*$ has purely imaginary values, the multiplexing process should obviously form $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ (instead of $\mathbf{X} = \mathbf{X}_1 + i\mathbf{X}_2$ if both had the same type of symmetry), and demultiplexing consists in separating

$$\mathbf{X}_1^* = \mathscr{Re}\,\mathbf{X}^*$$
$$\mathbf{X}_2^* = i\mathscr{Im}\,\mathbf{X}^*.$$

The general multiplexing formula for pairs of vectors may therefore be written

$$\mathbf{X} = \mathbf{X}_1 + \omega\mathbf{X}_2,$$

where ω is a phase factor (*e.g.* 1 or i) chosen in such a way that all non-exceptional components of $\mathbf{X}_1$ and $\mathbf{X}_2$ (or $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$) be embedded in the complex plane $\mathbb{C}$ along linearly independent directions, thus making multiplexing possible.

It is possible to develop a more general form of multiplexing/demultiplexing for more than two vectors, which can be used to deal with symmetry elements of order 3, 4 or 6. It is based on the theory of group characters (Ledermann, 1987).

1.3.4.4.3.6. *Global crystallographic algorithms*

All the necessary ingredients are now available for calculating the CDFT for any given space group.

1.3.4.4.3.6.1. *Triclinic groups*

Space group $P1$ is dealt with by the methods of Section 1.3.4.4.3.5.1 and $P\bar{1}$ by those of Section 1.3.4.4.3.5.4.

1.3.4.4.3.6.2. *Monoclinic groups*

A general monoclinic transformation is of the form

$$S_g : \mathbf{x} \longmapsto \mathbf{R}_g\mathbf{x} + \mathbf{t}_g$$

with $\mathbf{R}_g$ a diagonal matrix whose entries are $+1$ or -1, and $\mathbf{t}_g$ a vector whose entries are 0 or $\frac{1}{2}$. We may thus decompose both real and reciprocal space into a direct sum of a subspace $\mathbb{Z}^{n_+}$ where $\mathbf{R}_g$ acts as the identity, and a subspace $\mathbb{Z}^{n_-}$ where $\mathbf{R}_g$ acts as minus the identity, with $n_+ + n_- = n = 3$. All usual entities may be correspondingly written as direct sums, for instance:

$$\mathbf{R}_g = \mathbf{R}_g^+ \oplus \mathbf{R}_g^-, \quad \mathbf{N} = \mathbf{N}^+ \oplus \mathbf{N}^-, \quad \mathbf{M} = \mathbf{M}^+ \oplus \mathbf{M}^-,$$
$$\mathbf{t}_g = \mathbf{t}_g^+ \oplus \mathbf{t}_g^-, \quad \mathbf{t}_g^{(1)} = \mathbf{t}_g^{(1)+} \oplus \mathbf{t}_g^{(1)-}, \quad \mathbf{t}_g^{(2)} = \mathbf{t}_g^{(2)+} \oplus \mathbf{t}_g^{(2)-},$$
$$\mathbf{m} = \mathbf{m}^+ \oplus \mathbf{m}^-, \quad \mathbf{m}_1 = \mathbf{m}_1^+ \oplus \mathbf{m}_1^-, \quad \mathbf{m}_2 = \mathbf{m}_2^+ \oplus \mathbf{m}_2^-,$$
$$\mathbf{h} = \mathbf{h}^+ \oplus \mathbf{h}^-, \quad \mathbf{h}_1 = \mathbf{h}_1^+ \oplus \mathbf{h}_1^-, \quad \mathbf{h}_2 = \mathbf{h}_2^+ \oplus \mathbf{h}_2^-.$$

We will use factoring by 2, with decimation in frequency when computing structure factors, and decimation in time when computing electron densities; this corresponds to $\mathbf{N} = \mathbf{N}_1\mathbf{N}_2$ with $\mathbf{N}_1 = \mathbf{M}$, $\mathbf{N}_2 = 2\mathbf{I}$. The non-primitive translation vector $\mathbf{N}\mathbf{t}_g$ then belongs to $\mathbf{M}\mathbb{Z}^n$, and thus

$$\mathbf{t}_g^{(1)} = \mathbf{0} \bmod \mathbf{M}\mathbb{Z}^n, \quad \mathbf{t}_g^{(2)} \in \mathbb{Z}^n/2\mathbb{Z}^n.$$

The symmetry relations obeyed by $\wp$ and F are as follows: for electron densities

$$\wp(\mathbf{m}^+, \mathbf{m}^-) = \wp(\mathbf{m}^+ + \mathbf{N}^+\mathbf{t}_g^+, -\mathbf{m}^- - \mathbf{N}^-\mathbf{t}_g^-)$$

or, after factoring by 2,

$$\wp(\mathbf{m}_1^+, \mathbf{m}_2^+, \mathbf{m}_1^-, \mathbf{m}_2^-) = \wp(\mathbf{m}_1^+, \mathbf{m}_2^+ + \mathbf{t}_g^{(2)+}, \mathbf{M}^-\boldsymbol{\zeta}(\mathbf{m}_1^-) - \mathbf{m}_1^- - \mathbf{m}_2^-, \mathbf{m}_2^- + \mathbf{t}_g^{(2)-});$$

while for structure factors

$$F(\mathbf{h}^+, \mathbf{h}^-) = \exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)]F(\mathbf{h}^+, -\mathbf{h}^-)$$

with its Friedel counterpart

$$F(\mathbf{h}^+, \mathbf{h}^-) = \exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)]\overline{F(-\mathbf{h}^+, \mathbf{h}^-)}$$

or, after factoring by 2,

$$F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-) = (-1)^{\mathbf{h}_2^+ \cdot \mathbf{t}_g^{(2)+} + \mathbf{h}_2^- \cdot \mathbf{t}_g^{(2)-}} \times F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{M}^-\boldsymbol{\zeta}(\mathbf{h}_1^-) - \mathbf{h}_1^- - \mathbf{h}_2^-, \mathbf{h}_2^-)$$

with Friedel counterpart

$$F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-) = (-1)^{\mathbf{h}_2^+ \cdot \mathbf{t}_g^{(2)+} + \mathbf{h}_2^- \cdot \mathbf{t}_g^{(2)-}} \overline{F[\mathbf{M}^+\boldsymbol{\zeta}(\mathbf{h}_1^+) - \mathbf{h}_1^+ - \mathbf{h}_2^+, \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-]}.$$

When calculating electron densities, two methods may be used.

(i) Transform on $\mathbf{h}^-$ first.

The partial vectors defined by $X_{\mathbf{h}^+,\,\mathbf{h}_2^-} = F(\mathbf{h}^+, \mathbf{h}_1^-, \mathbf{h}_2^-)$ obey symmetry relations of the form

$$X(\mathbf{h}_1^- - \mathbf{h}_2^-) = \varepsilon X[\mathbf{M}^-\boldsymbol{\zeta}(\mathbf{h}_1^-) - \mathbf{h}_1^-]$$

with $\varepsilon = \pm 1$ independent of $\mathbf{h}_1^-$. This is the same relation as for the same parity class of data for a (complex or real) symmetric ($\varepsilon = +1$) or antisymmetric ($\varepsilon = -1$) transform. The same techniques can be used to decrease the number of $F(\mathbf{M}^-)$ by multiplexing pairs of such vectors and demultiplexing their transforms. Partial vectors with different values of ε may be mixed in this way (Section 1.3.4.4.3.5.6).

Once $F(\mathbf{N}^-)$ is completed, its results have Hermitian symmetry with respect to $\mathbf{h}^+$, and the methods of Section 1.3.4.4.3.5.1 may be used to obtain the unique electron densities.

(ii) Transform on $\mathbf{h}^+$ first.

The partial vectors defined by $X_{\mathbf{h}^-,\,\mathbf{h}_2^+} = F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}^-)$ obey symmetry relations of the form

$$X(\mathbf{h}_1^+ - \mathbf{h}_2^+) = \overline{\varepsilon X[\mathbf{M}^+\boldsymbol{\zeta}(\mathbf{h}_1^+) - \mathbf{h}_1^+]}$$

with $\varepsilon = \pm 1$ independent of $\mathbf{h}_1^+$. This is the same relation as for the same parity class of data for a Hermitian symmetric ($\varepsilon = +1$) or antisymmetric ($\varepsilon = -1$) transform. The same techniques may be used to decrease the number of $F(\mathbf{M}^+)$. This generalizes the procedure described by Ten Eyck (1973) for treating dyad axes, *i.e.* for the case $n_+ = 1, \mathbf{t}_g^{(2)-} = \mathbf{0}$, and $\mathbf{t}_g^{(2)+} = \mathbf{0}$ (simple dyad) or $\mathbf{t}_g^{(2)+} \neq \mathbf{0}$ (screw dyad).

Once $F(\mathbf{N}^+)$ is completed, its results have Hermitian symmetry properties with respect to $\mathbf{h}^-$ which can be used to obtain the unique electron densities.

Structure factors may be computed by applying the reverse procedures in the reverse order.

1.3.4.4.3.6.3. *Orthorhombic groups*

Almost all orthorhombic space groups are generated by two monoclinic transformations g_1 and g_2 of the type described in Section 1.3.4.4.3.6.2, with the addition of a centre of inversion $-e$ for centrosymmetric groups. The only exceptions are *Fdd*2 and *Fddd*

which contain diamond glides, in which some non-primitive translations are 'square roots' not of primitive lattice translations, but of centring translations. The generic case will be examined first.

To calculate electron densities, the unique octant of data may first be transformed on $\mathbf{h}^+$ (respectively $\mathbf{h}^-$) as in Section 1.3.4.3.6.2 using the symmetry pertaining to generator g_1. These intermediate results may then be expanded by generator g_2 by the formula of Section 1.3.4.3.3 prior to the final transform on $\mathbf{h}^-$ (respectively $\mathbf{h}^+$). To calculate structure factors, the reverse operations are applied in the reverse order.

The two exceptional groups *Fdd*2 and *Fddd* only require a small modification. The *F*-centring causes the systematic absence of parity classes with mixed parities, leaving only (000) and (111). For the former, the phase factors $\exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)]$ in the symmetry relations of Section 1.3.4.3.6.2 become powers of (-1) so that one is back to the generic case. For the latter, these phase factors are odd powers of i which it is a simple matter to incorporate into a modified multiplexing/demultiplexing procedure.

1.3.4.3.6.4. *Trigonal, tetragonal and hexagonal groups*

All the symmetries in this class of groups can be handled by the generalized Rader/Winograd algorithms of Section 1.3.4.3.4.3, but no implementation of these is yet available.

In groups containing axes of the form n_m with g.c.d. $(m, n) = 1$ (*i.e.* $3_1, 3_2, 4_1, 4_3, 6_1, 6_5$) along the c direction, the following procedure may be used (Ten Eyck, 1973):

(i) to calculate electron densities, the unique structure factors indexed by

$$[\text{unique } (h,k)] \times (\text{all } l)$$

are transformed on l; the results are rearranged by the transposition formula of Section 1.3.4.3.3 so as to be indexed by

$$[\text{all } (h,k)] \times \left[\text{unique } \left(\frac{1}{n}\right)\text{th of } z\right]$$

and are finally transformed on (h, k) to produce an asymmetric unit. For a dihedral group, the extra twofold axis may be used in the transposition to produce a unique $(1/2n)$th of z.

(ii) to calculate structure factors, the unique densities in $(1/n)$th of z [or $(1/2n)$th for a dihedral group] are first transformed on x and y, then transposed by the formula of Section 1.3.4.3.3 to reindex the intermediate results by

$$[\text{unique } (h,k)] \times (\text{all } z);$$

the last transform on z is then carried out.

1.3.4.3.6.5. *Cubic groups*

These are usually treated as their orthorhombic or tetragonal subgroups, as the body-diagonal threefold axis cannot be handled by ordinary methods of decomposition.

The three-dimensional factorization technique of Section 1.3.4.3.4.1 allows a complete treatment of cubic symmetry. Factoring by 2 along all three dimensions gives four types (*i.e.* orbits) of parity classes:

(000)	with residual threefold symmetry,
(100), (010), (001)	related by threefold axis,
(110), (101), (011)	related by threefold axis,
(111)	with residual threefold symmetry.

Orbit exchange using the threefold axis thus allows one to reduce the number of partial transforms from 8 to 4 (one per orbit). Factoring by 3 leads to a reduction from 27 to 11 (in this case, further reduction to 9 can be gained by multiplexing the three diagonal classes with residual threefold symmetry into a single class; see Section 1.3.4.3.5.6). More generally, factoring by q leads to a reduction from q^3 to $\frac{1}{3}(q^3 - q) - q$. Each of the remaining transforms then has a symmetry induced from the orthorhombic or tetragonal subgroup, which can be treated as above.

No implementation of this procedure is yet available.

1.3.4.3.6.6. *Treatment of centred lattices*

Lattice centring is an instance of the duality between periodization and decimation: the extra translational periodicity of ρ induces a decimation of $\mathbf{F} = \{F_\mathbf{h}\}$ described by the 'reflection conditions' on $\mathbf{h}$. As was pointed out in Section 1.3.4.2.2.3, non-primitive lattices are introduced in order to retain the same matrix representation for a given geometric symmetry operation in all the arithmetic classes in which it occurs. From the computational point of view, therefore, the main advantage in using centred lattices is that it *maximizes decomposability* (Section 1.3.4.2.2.4); reindexing to a primitive lattice would for instance often destroy the diagonal character of the matrix representing a dyad.

In the usual procedure involving three successive one-dimensional transforms, the loss of efficiency caused by the duplication of densities or the systematic vanishing of certain classes of structure factors may be avoided by using a multiplexing/demultiplexing technique (Ten Eyck, 1973):

(i) for base-centred or body-centred lattices, two successive planes of structure factors may be overlaid into a single plane; after transformation, the results belonging to each plane may be separated by parity considerations;

(ii) for face-centred lattices the same method applies, using four successive planes (the third and fourth with an origin translation);

(iii) for rhombohedral lattices in hexagonal coordinates, three successive planes may be overlaid, and the results may be separated by linear combinations involving cube roots of unity.

The three-dimensional factorization technique of Section 1.3.4.3.4.1 is particularly well suited to the treatment of centred lattices: if the decimation matrix of $\mathbf{N}$ contains as a factor $\mathbf{N}_1$ a matrix which 'integerizes' all the non-primitive lattice vectors, then centring is reflected by the systematic vanishing of certain classes of vectors of decimated data or results, which can simply be omitted from the calculation. An alternative possibly is to reindex on a primitive lattice and use different representative matrices for the symmetry operations: the loss of decomposability is of little consequence in this three-dimensional scheme, although it substantially complicates the definition of the cocycles $\boldsymbol{\mu}_2$ and $\boldsymbol{\eta}_1$.

1.3.4.3.6.7. *Programming considerations*

The preceding sections have been devoted to showing how the raw computational efficiency of a crystallographic Fourier transform *algorithm* can be maximized. This section will briefly discuss another characteristic (besides speed) which a crystallographic Fourier transform *program* may be required to possess if it is to be useful in various applications: a convenient and versatile mode of presentation of input data or output results.

The standard crystallographic FFT programs (Ten Eyck, 1973, 1985) are rather rigid in this respect, and use rather rudimentary data structures (lists of structure-factor values, and two-dimensional arrays containing successive sections of electron-density maps). It is frequently the case that considerable reformatting of these data or results must be carried out before they can be used in other computations; for instance, maps have to be converted from 2D sections to 3D 'bricks' before they can be inspected on a computer graphics display.

The explicitly three-dimensional approach to the factorization of the DFT and the use of symmetry offers the possibility of richer and more versatile data structures. For instance, the use of 'decimation in frequency' in real space and of 'decimation in time' in reciprocal

space leads to data structures in which real-space coordinates are handled by blocks (thus preserving, at least locally, the three-dimensional topological connectivity of the maps) while reciprocal-space indices are handled by parity classes or their generalizations for factors other than 2 (thus making the treatment of centred lattices extremely easy). This global three-dimensional indexing also makes it possible to carry symmetry and multiplicity characteristics for each subvector of intermediate results for the purpose of automating the use of the orbit exchange mechanism.

Brünger (1989) has described the use of a similar three-dimensional factoring technique in the context of structure-factor calculations for the refinement of macromolecular structures.

1.3.4.4. *Basic crystallographic computations*

1.3.4.4.1. *Introduction*

Fourier transform (FT) calculations play an indispensable role in crystallography, because the Fourier transformation is inherent in the diffraction phenomenon itself.

Besides this obligatory use, the FT has numerous other applications, motivated more often by its mathematical properties than by direct physical reasoning (although the latter can be supplied after the fact). Typically, many crystallographic computations turn out to be convolutions in disguise, which can be speeded up by orders of magnitude through a judicious use of the FT. Several recent advances in crystallographic computation have been based on this kind of observation.

1.3.4.4.2. *Fourier synthesis of electron-density maps*

Bragg (1929) was the first to use this type of calculation to assist structure determination. Progress in computing techniques since that time was reviewed in Section 1.3.4.3.1.

The usefulness of the maps thus obtained can be adversely affected by three main factors:

(i) limited resolution;

(ii) errors in the data;

(iii) computational errors.

Limited resolution causes 'series-termination errors' first investigated by Bragg & West (1930), who used an optical analogy with the numerical aperture of a microscope. James (1948*b*) gave a quantitative description of this phenomenon as a convolution with the 'spherical Dirichlet kernel' (Section 1.3.4.2.1.3), which reflects the truncation of the Fourier spectrum by multiplication with the indicator function of the limiting resolution sphere. Bragg & West (1930) suggested that the resulting ripples might be diminished by applying an artificial temperature factor to the data, which performs a further convolution with a Gaussian point-spread function. When the electron-density map is to be used for model refinement, van Reijen (1942) suggested using Fourier coefficients calculated from the model when no observation is available, as a means of combating series-termination effects.

Errors in the data introduce errors in the electron-density maps, with the same mean-square value by virtue of Parseval's theorem. Special positions accrue larger errors (Cruickshank & Rollett, 1953; Cruickshank, 1965*a*). To minimize the mean-square electron-density error due to large phase uncertainties, Blow & Crick (1959) introduced the 'best Fourier' which uses centroid Fourier coefficients; the associated error level in the electron-density map was evaluated by Blow & Crick (1959) and Dickerson *et al.* (1961*a*,*b*).

Computational errors used to be a serious concern when Beevers–Lipson strips were used, and Cochran (1948*a*) carried out a critical evaluation of the accuracy limitations imposed by strip methods. Nowadays, the FFT algorithm implemented on digital computers with a word size of at least 32 bits gives results accurate to six decimal places or better in most applications (see Gentleman & Sande, 1966).

1.3.4.4.3. *Fourier analysis of modified electron-density maps*

Various approaches to the phase problem are based on certain modifications of the electron-density map, followed by Fourier analysis of the modified map and extraction of phase information from the resulting Fourier coefficients.

1.3.4.4.3.1. *Squaring*

Sayre (1952*a*) derived his 'squaring method equation' for structures consisting of equal, resolved and spherically symmetric atoms by observing that squaring such an electron density is equivalent merely to sharpening each atom into its square. Thus

$$F_{\mathbf{h}} = \theta_{\mathbf{h}} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}},$$

where $\theta_{\mathbf{h}} = f(\mathbf{h})/f^{\mathrm{sq}}(\mathbf{h})$ is the ratio between the form factor $f(\mathbf{h})$ common to all the atoms and the form factor $f^{\mathrm{sq}}(\mathbf{h})$ for the squared version of that atom.

Most of the central results of direct methods, such as the tangent formula, are an immediate consequence of Sayre's equation. Phase refinement for a macromolecule by enforcement of the squaring method equation was demonstrated by Sayre (1972, 1974).

1.3.4.4.3.2. *Other non-linear operations*

A category of phase improvement procedures known as 'density modification' is based on the pointwise application of various quadratic or cubic 'filters' to electron-density maps after removal of negative regions (Hoppe & Gassmann, 1968; Hoppe *et al.*, 1970; Barrett & Zwick, 1971; Gassmann & Zechmeister, 1972; Collins, 1975; Collins *et al.*, 1976; Gassmann, 1976). These operations are claimed to be equivalent to reciprocal-space phase-refinement techniques such as those based on the tangent formula. Indeed the replacement of

$$\rho(\mathbf{x}) = \sum_{\mathbf{h}} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

by $P[\rho(\mathbf{x})]$, where P is a polynomial

$$P(\rho) = a_0 + a_1\rho + a_2\rho^2 + a_3\rho^3 + \ldots$$

yields

$$P[\rho(\mathbf{x})] = a_0 + \sum_{\mathbf{h}} \left[a_1 F_{\mathbf{h}} + a_2 \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}} + a_3 \sum_{\mathbf{k}} \sum_{\mathbf{l}} F_{\mathbf{k}} F_{\mathbf{l}} F_{\mathbf{h}-\mathbf{k}-\mathbf{l}} + \ldots \right] \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

and hence gives rise to the convolution-like families of terms encountered in direct methods. This equivalence, however, has been shown to be rather superficial (Bricogne, 1982) because the 'uncertainty principle' embodied in Heisenberg's inequality (Section 1.3.2.4.4.3) imposes severe limitations on the effectiveness of any procedure which operates *pointwise* in *both* real and reciprocal space.

In applying such methods, sampling considerations must be given close attention. If the spectrum of ρ extends to resolution Δ and if the pointwise non-linear filter involves a polynomial P of degree n, then $P(\rho)$ should be sampled at intervals of at most $\Delta/2n$ to accommodate the full bandwidth of its spectrum.

1.3.4.4.3.3. *Solvent flattening*

Crystals of proteins and nucleic acids contain large amounts of mother liquor, often in excess of 50% of the unit-cell volume,

occupying connected channels. The well ordered electron density $\rho_M(\mathbf{x})$ corresponding to the macromolecule thus occupies only a periodic subregion $\mathscr{U}$ of the crystal. Thus

$$\rho_M = \chi_{\mathscr{U}} \times \rho_M,$$

implying the convolution identity between structure factors (Main & Woolfson, 1963):

$$F_M(\mathbf{h}) = \sum_{\mathbf{k}} \bar{\mathscr{F}}\left[\frac{1}{\mathscr{U}}\chi_{\mathscr{U}}\right](\mathbf{h}-\mathbf{k})F_M(\mathbf{k})$$

which is a form of the Shannon interpolation formula (Sections 1.3.2.7.1, 1.3.4.2.1.7; Bricogne, 1974; Colman, 1974).

It is often possible to obtain an approximate 'molecular envelope' $\mathscr{U}$ from a poor electron-density map ρ, either interactively by computer graphics (Bricogne, 1976) or automatically by calculating a moving average of the electron density within a small sphere S. The latter procedure can be implemented in real space (Wang, 1985). However, as it is a convolution of ρ with χ_S, it can be speeded up considerably (Leslie, 1987) by computing the moving average ρ_{mav} as

$$\rho_{\text{mav}}(\mathbf{x}) = \mathscr{F}[\bar{\mathscr{F}}[\rho] \times \bar{\mathscr{F}}[\chi_S]](\mathbf{x}).$$

This remark is identical in substance to Booth's method of computation of 'bounded projections' (Booth, 1945*a*) described in Section 1.3.4.2.1.8, except that the summation is kept three-dimensional.

The iterative use of the estimated envelope $\mathscr{U}$ for the purpose of phase improvement (Wang, 1985) is a submethod of the previously developed method of molecular averaging, which is described below. Sampling rules for the Fourier analysis of envelope-truncated maps will be given there.

1.3.4.4.3.4. *Molecular averaging by noncrystallographic symmetries*

Macromolecules and macromolecular assemblies frequently crystallize with several identical subunits in the asymmetric metric unit, or in several crystal forms containing the same molecule in different arrangements. Rossmann & Blow (1963) recognized that intensity data collected from such structures are redundant (Sayre, 1952*b*) and that their redundancy could be a source of phase information.

The phase constraints implied by the consistency of geometrically redundant intensities were first derived by Rossmann & Blow (1963), and were generalized by Main & Rossmann (1966). Crowther (1967, 1969) reformulated them as linear eigenvalue equations between structure factors, for which he proposed an iterative matrix solution method. Although useful in practice (Jack, 1973), this reciprocal-space approach required computations of size $\propto N^2$ for N reflections, so that N could not exceed a few thousands.

The theory was then reformulated in real space (Bricogne, 1974), showing that the most costly step in Crowther's procedure could be carried out much more economically by averaging the electron densities of all crystallographically independent subunits, then rebuilding the crystal(s) from this averaged subunit, flattening the density in the solvent region(s) by resetting it to its average value. This operation is a *projection* [by virtue of Section 1.3.4.2.2.2(*d*)]. The overall complexity was thus reduced from N^2 to $N \log N$. The design and implementation of a general-purpose program package for averaging, reconstructing and solvent-flattening electron-density maps (Bricogne, 1976) led rapidly to the first high-resolution determinations of virus structures (Bloomer *et al.*, 1978; Harrison *et al.*, 1978), with $N \sim 200\,000$.

The considerable gain in speed is a consequence of the fact that the masking operations used to retrieve the various copies of the common subunit are carried out by simple *pointwise multiplication* by an indicator function χ_U in real space, whereas they involve a *convolution* with $\bar{\mathscr{F}}[\chi_U]$ in reciprocal space.

The averaging by noncrystallographic symmetries of an electron-density map calculated by FFT – hence sampled on a grid which is an integral subdivision of the period lattice – necessarily entails the interpolation of densities at non-integral points of that grid. The effect of interpolation on the structure factors recalculated from an averaged map was examined by Bricogne (1976). This study showed that, if linear interpolation is used, the initial map should be calculated on a fine grid, of size $\Delta/5$ or $\Delta/6$ at resolution Δ (instead of the previously used value of $\Delta/3$). The analysis about to be given applies to all interpolation schemes which consist in a *convolution* of the sampled density with a fixed *interpolation kernel* function K.

Let ρ be a $\mathbb{Z}^3$-periodic function. Let K be the interpolation kernel in 'normalized' form, *i.e.* such that $\int_{\mathbb{R}^3} K(\mathbf{x})\ \mathrm{d}^3\mathbf{x} = 1$ and scaled so as to interpolate between sample values given on a unit grid $\mathbb{Z}^3$; in the case of linear interpolation, K is the 'trilinear wedge'

$$K(\mathbf{x}) = W(x)W(y)W(z),$$

where

$$\begin{aligned} W(t) &= 1 - |t| \quad \text{if } |t| \le 1, \\ &= 0 \qquad \text{if } |t| \ge 1. \end{aligned}$$

Let ρ be sampled on a grid $\mathscr{G}_1 = \mathbf{N}_1^{-1}\mathbb{Z}^3$, and let $I_{\mathbf{N}_1}\rho$ denote the function interpolated from this sampled version of ρ. Then:

$$I_{\mathbf{N}_1}\rho = \left[\rho \times \sum_{\mathbf{m}\in\mathbb{Z}^3} \delta_{(\mathbf{N}_1^{-1}\mathbf{m})}\right] * [(\mathbf{N}_1^{-1})^{\#}K],$$

where $[(N_1^{-1})^{\#}K](\mathbf{x}) = K(\mathbf{N}_1\mathbf{x})$, so that

$$\begin{aligned} \bar{\mathscr{F}}[I_{\mathbf{N}_1}\rho] &= \left[\bar{\mathscr{F}}[\rho] * |\det \mathbf{N}_1| \sum_{\mathbf{k}_1\in\mathbb{Z}^3} \delta_{(\mathbf{N}_1^T\mathbf{k}_1)}\right] \\ &\quad \times \left[\frac{1}{|\det \mathbf{N}_1|}(N_1^T)^{\#}\bar{\mathscr{F}}[K]\right] \\ &= \left[\sum_{\mathbf{k}_1\in\mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_1^T\mathbf{k}_1}\bar{\mathscr{F}}[\rho]\right] \times (N_1^T)^{\#}\bar{\mathscr{F}}[K]. \end{aligned}$$

The transform of $I_{\mathbf{N}_1}\rho$ thus consists of

(i) a 'main band' corresponding to $\mathbf{k}_1 = \mathbf{0}$, which consists of the true transform $\bar{\mathscr{F}}[\rho](\boldsymbol{\xi})$ attenuated by multiplication by the central region of $\bar{\mathscr{F}}[K][(\mathbf{N}^{-1})^T\boldsymbol{\xi}]$; in the case of linear interpolation, for example,

$$\bar{\mathscr{F}}[K](\xi,\eta,\zeta) = \left(\frac{\sin\pi\xi}{\pi\xi}\right)^2\left(\frac{\sin\pi\eta}{\pi\eta}\right)^2\left(\frac{\sin\pi\zeta}{\pi\zeta}\right)^2;$$

(ii) a series of 'ghost bands' corresponding to $\mathbf{k}_1 \neq \mathbf{0}$, which consist of translates of $\bar{\mathscr{F}}[\rho]$ multiplied by the tail regions of $(N_1^T)^{\#}\bar{\mathscr{F}}[K]$.

Thus $I_{\mathbf{N}_1}\rho$ is not band-limited even if ρ is. Supposing, however, that ρ is band-limited and that grid $\mathscr{G}_1$ satisfies the Shannon sampling criterion, we see that there will be no overlap between the different bands: $\bar{\mathscr{F}}[\rho]$ may therefore be recovered from the main band by compensating its attenuation, which is approximately a temperature-factor correction.

For numerical work, however, $I_{\mathbf{N}_1}\rho$ must be resampled onto another grid $\mathscr{G}_2$, which causes its transform to become periodized into

$$|\det \mathbf{N}_2| \sum_{\mathbf{k}_2\in\mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_2^T\mathbf{k}_2}\left\{\left[\sum_{\mathbf{k}_1\in\mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_1^T\mathbf{k}_1}\bar{\mathscr{F}}[\rho]\right](N_1^T)^{\#}\bar{\mathscr{F}}[K]\right\}.$$

This now causes the main band $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{0}$ to become contaminated by the ghost bands ($\mathbf{k}_1 \neq \mathbf{0}$) of the translates ($\mathbf{k}_2 \neq \mathbf{0}$) of $I_{\mathbf{N}_1 \mathcal{P}}$.

Aliasing errors may be minimized by increasing the sampling rate in grid $\mathcal{G}_1$ well beyond the Shannon minimum, which rapidly reduces the r.m.s. content of the ghost bands.

The sampling rate in grid $\mathcal{G}_2$ needs only exceed the Shannon minimum to the extent required to accommodate the increase in bandwidth due to convolution with $\bar{\mathcal{F}}[\chi_U]$, which is the reciprocal-space counterpart of envelope truncation (or solvent flattening) in real space.

1.3.4.4.3.5. *Molecular-envelope transforms via Green's theorem*

Green's theorem stated in terms of distributions (Section 1.3.2.3.9.1) is particularly well suited to the calculation of the Fourier transforms $\bar{\mathcal{F}}[\chi_U]$ of indicator functions. Let f be the indicator function χ_U and let S be the boundary of U (assumed to be a smooth surface). The jump σ_0 in the value of f across S along the outer normal vector is $\sigma_0 = -1$, the jump σ_ν in the normal derivative of f across S is $\sigma_\nu = 0$, and the Laplacian of f as a function is (almost everywhere) 0 so that $T_{\Delta f} = 0$. Green's theorem then reads:

$$\begin{aligned}\Delta(T_f) &= T_{\Delta f} + \sigma_\nu \delta_{(S)} + \partial_\nu[\sigma_0 \delta_{(S)}] \\ &= -\partial_\nu[\delta_{(S)}].\end{aligned}$$

The function $e_{\mathbf{H}}(\mathbf{X}) = \exp(2\pi i \mathbf{H}\cdot\mathbf{X})$ satisfies the identity $\Delta e_{\mathbf{H}} = -4\pi^2 \|\mathbf{H}\|^2 e_{\mathbf{H}}$. Therefore, in Cartesian coordinates:

$$\begin{aligned}\bar{F}[\chi_U](\mathbf{H}) &= \langle T_{\chi_U}, e_{\mathbf{H}} \rangle \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \langle T_{\chi_U}, \Delta e_{\mathbf{H}} \rangle \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \langle \Delta(T_{\chi_U}), e_{\mathbf{H}} \rangle \qquad \text{[Section 1.3.2.3.9.1}(a)\text{]} \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \langle -\partial_\nu[\delta_{(S)}], e_{\mathbf{H}} \rangle \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \int_S \partial_\nu e_{\mathbf{H}} \, \mathrm{d}^2 S \qquad \text{[Section 1.3.2.3.9.1}(c)\text{]} \\ &= -\frac{1}{4\pi^2\|\mathbf{H}\|^2} \int_S 2\pi i \mathbf{H}\cdot\mathbf{n} \exp(2\pi i \mathbf{H}\cdot\mathbf{X}) \, \mathrm{d}^2 S,\end{aligned}$$

i.e.

$$\bar{\mathcal{F}}[\chi_U](\mathbf{H}) = \frac{1}{2\pi i \|\mathbf{H}\|^2} \int_S \mathbf{H}\cdot\mathbf{n} \exp(2\pi i \mathbf{H}\cdot\mathbf{X}) \, \mathrm{d}^2 S,$$

where $\mathbf{n}$ is the outer normal to S. This formula was used by von Laue (1936) for a different purpose, namely to calculate the transforms of crystal shapes (see also Ewald, 1940). If the surface S is given by a triangulation, the surface integral becomes a sum over all faces, since $\mathbf{n}$ is constant on each face. If U is a solid sphere with radius R, an integration by parts gives immediately:

$$\frac{1}{\mathrm{vol}(U)} \bar{\mathcal{F}}[\chi_U](\mathbf{H}) = \frac{3}{X^3}[\sin X - X\cos X] \quad \text{with } X = 2\pi\|\mathbf{H}\|R.$$

1.3.4.4.4. *Structure factors from model atomic parameters*

An atomic model of a crystal structure consists of a list of symmetry-unique atoms described by their positions, their thermal agitation and their chemical identity (which can be used as a pointer to form-factor tables). Form factors are usually parameterized as sums of Gaussians, and thermal agitation by a Gaussian temperature factor or tensor. The formulae given in Section 1.3.4.2.2.6 for Gaussian atoms are therefore adequate for most purposes. High-resolution electron-density studies use more involved parameterizations.

Early calculations were carried out by means of Bragg–Lipson charts (Bragg & Lipson, 1936) which gave a graphical representation of the symmetrized trigonometric sums Ξ of Section 1.3.4.2.2.9. The approximation of form factors by Gaussians goes back to the work of Vand *et al.* (1957) and Forsyth & Wells (1959). Agarwal (1978) gave simplified expansions suitable for medium-resolution modelling of macromolecular structures.

This method of calculating structure factors is expensive because *each* atom sends contributions of essentially equal magnitude to *all* structure factors in a resolution shell. The calculation is therefore of size $\propto N\mathcal{N}$ for N atoms and $\mathcal{N}$ reflections. Since N and $\mathcal{N}$ are roughly proportional at a given resolution, this method is very costly for large structures.

Two distinct programming strategies are available (Rollett, 1965) according to whether the fast loop is on all atoms for each reflection, or on all reflections for each atom. The former method was favoured in the early times when computers were unreliable. The latter was shown by Burnett & Nordman (1974) to be more amenable to efficient programming, as no multiplication is required in calculating the arguments of the sine/cosine terms: these can be accumulated by integer addition, and used as subscripts in referencing a trigonometric function table.

1.3.4.4.5. *Structure factors via model electron-density maps*

Robertson (1936*b*) recognized the similarity between the calculation of structure factors by Fourier summation and the calculation of Fourier syntheses, the main difference being of course that atomic coordinates do not usually lie exactly on a grid obtained by integer subdivision of the crystal lattice. He proposed to address this difficulty by the use of his sorting board, which could extend the scale of subdivision and thus avoid phase errors. In this way the calculation of structure factors became amenable to Beevers–Lipson strip methods, with considerable gain of speed.

Later, Beevers & Lipson (1952) proposed that trigonometric functions attached to atomic positions falling between the grid points on which Beevers–Lipson strips were based should be obtained by linear interpolation from the values found on the strips for the closest grid points. This amounts (Section 1.3.4.4.3.4) to using atoms in the shape of a trilinear wedge, whose form factor was indicated in Section 1.3.4.4.3.4 and gives rise to aliasing effects (see below) not considered by Beevers & Lipson.

The correct formulation of this idea came with the work of Sayre (1951), who showed that structure factors could be calculated by Fourier analysis of a sampled electron-density map previously generated on a subdivision $\mathbf{N}^{-1}\Lambda$ of the crystal lattice Λ. When generating such a map, care must be taken to distribute onto the sample grid not only the electron densities of all the atoms in the asymmetric motif, but also those of their images under space-group symmetries and lattice translations. Considerable savings in computation occur, especially for large structures, because atoms are *localized*: each atom sends contributions to only a few grid points in real space, rather than to all reciprocal-lattice points. The generation of the sampled electron-density map is still of complexity $\propto N\mathcal{N}$ for N atoms and $\mathcal{N}$ reflections, but the proportionality constant is smaller than that in Section 1.3.4.4.4 by orders of magnitude; the extra cost of Fourier analysis, proportional to $\mathcal{N}\log\mathcal{N}$, is negligible.

The idea of approximating a Fourier transform by a discrete transform on sampled values had already been used by Whittaker (1948), who tested it on the first three odd Hermite functions and did not consider the problem of aliasing errors. By contrast, Sayre gave a lucid analysis of the sampling problems associated to this technique. If the periodic sampled map is written in the form of a weighted lattice distribution (as in Section 1.3.2.7.3) as

$$\rho^s = \sum_{\mathbf{m}\in\mathbb{Z}^3} \rho(\mathbf{N}^{-1}\mathbf{m})\delta_{(\mathbf{N}^{-1}\mathbf{m})},$$

then its discrete Fourier transform yields

$$F^s(\mathbf{h}) = \sum_{\boldsymbol{\eta}\in\mathbb{Z}^3} F(\mathbf{h}+\mathbf{N}^T\boldsymbol{\eta})$$

so that each correct value $F(\mathbf{h})$ is corrupted by its aliases $F(\mathbf{h}+\mathbf{N}^T\boldsymbol{\eta})$ for $\boldsymbol{\eta}\neq\mathbf{0}$.

To cure this aliasing problem, Sayre used 'hypothetical atoms' with form factors equal to those of standard atoms within the resolution range of interest, but set to zero outside that range. This amounts to using atomic densities with built-in series-termination errors, which has the detrimental effect of introducing slowly decaying ripples around the atom which require incrementing sample densities at many more grid points per atom.

Sayre considered another cure in the form of an artificial temperature factor B (Bragg & West, 1930) applied to all atoms. This spreads each atom on more grid points in real space but speeds up the decay of its transform in reciprocal space, thus allowing the use of a coarser sampling grid in real space. He discounted it as spoiling the agreement with observed data, but Ten Eyck (1977) pointed out that this agreement could be restored by applying the negative of the artificial temperature factor to the results. This idea cannot be carried to extremes: if B is chosen too large, the atoms will be so spread out in real space as each to occupy a sizeable fraction of the unit cell and the advantage of atom localization will be lost; furthermore, the form factors will fall off so rapidly that round-off error amplification will occur when the results are sharpened back. Clearly, there exists an optimal combination of B and sampling rate yielding the most economical computation for a given accuracy at a given resolution, and a formula will now be given to calculate it.

Let us make the simplifying assumption that all atoms are roughly equal and that their common form factor can be represented by an equivalent temperature factor $B_{\rm eq}$. Let $\Delta = 1/d^*_{\rm max}$ be the resolution to which structure factors are wanted. The Shannon sampling interval is $\Delta/2 = 1/2d^*_{\rm max}$. Let σ be the oversampling rate, so that the actual sampling interval in the map is $\Delta/2\sigma = 1/2\sigma d^*_{\rm max}$: then consecutive copies of the transform are separated by a distance $2\sigma d^*_{\rm max}$ in reciprocal space. Let the artificial temperature factor $B_{\rm extra}$ be added, and let

$$B = B_{\rm eq} + B_{\rm extra}.$$

The worst aliasing occurs at the outer resolution limit $d^*_{\rm max}$, where the 'signal' due to an atom is proportional to

$$\exp[(-B/4)(d^*_{\rm max})^2],$$

while the 'noise' due to the closest alias is proportional to

$$\exp\{(-B/4)[(2\sigma-1)d^*_{\rm max}]^2\}.$$

Thus the signal-to-noise ratio, or quality factor, Q is

$$\exp[B\sigma(\sigma-1)(d^*_{\rm max})^2].$$

If a certain value of Q is desired (*e.g.* $Q = 100$ for 1% accuracy), then the equation

$$B = \frac{\log Q}{\sigma(\sigma-1)(d^*_{\rm max})^2}$$

defines B in terms of $\sigma, d^*_{\rm max}$ and Q.

The overall cost of the structure-factor calculation from N atoms is then

(i) $C_1 \times B^{2/3} \times N$ for density generation,

(ii) $C_2 \times (2\sigma d^*_{\rm max})^3 \times \log[(2\sigma d^*_{\rm max})^3]$ for Fourier analysis,

where C_1 and C_2 are constant depending on the speed of the computer used. This overall cost may be minimized with respect to σ for given $d^*_{\rm max}$ and Q, determining the optimal B (and hence $B_{\rm extra}$) in passing by the above relation.

Sayre (1951) did observe that applying an artificial temperature factor in real space would not create series-termination ripples: the resulting atoms would have a smaller effective radius than his hypothetical atoms, so that step (i) would be faster. This optimality of Gaussian smearing is ultimately a consequence of Hardy's theorem (Section 1.3.2.4.4.3).

1.3.4.4.6. *Derivatives for variational phasing techniques*

Some methods of phase determination rely on maximizing a certain global criterion $S[\rho]$ involving the electron density, of the form $\int_{\mathbb{R}^3/\mathbb{Z}^3} K[\rho(\mathbf{x})]\ {\rm d}^3\mathbf{x}$, under constraint of agreement with the observed structure-factor amplitudes, typically measured by a χ^2 residual C. Several recently proposed methods use for $S[\rho]$ various measures of entropy defined by taking $K(\rho) = -\rho\log(\rho/\mu)$ or $K(\rho) = \log\rho$ (Bricogne, 1982; Britten & Collins, 1982; Narayan & Nityananda, 1982; Bryan *et al.*, 1983; Wilkins *et al.*, 1983; Bricogne, 1984; Navaza, 1985; Livesey & Skilling, 1985). Sayre's use of the squaring method to improve protein phases (Sayre, 1974) also belongs to this category, and is amenable to the same computational strategies (Sayre, 1980).

These methods differ from the density-modification procedures of Section 1.3.4.4.3.2 in that they seek an optimal solution by moving electron densities (or structure factors) *jointly* rather than pointwise, *i.e.* by moving along suitably chosen *search directions* $v_i(\mathbf{x})$ [or $V_i(\mathbf{h})$].

For computational purposes, these search directions may be handled either as column vectors of sample values $\{v_i(\mathbf{N}^{-1}\mathbf{m})\}_{\mathbf{m}\in\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3}$ on a grid in real space, or as column vectors of Fourier coefficients $\{V_i(\mathbf{h})\}_{\mathbf{h}\in\mathbb{Z}^3/\mathbf{N}^T\mathbb{Z}^3}$ in reciprocal space. These column vectors are the coordinates of the same vector $\mathbf{V}_i$ in an abstract vector space $\mathcal{V} \cong L(\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3)$ of dimension $\mathcal{N} = |\det\mathbf{N}|$ over $\mathbb{R}$, but referred to two different bases which are related by the DFT and its inverse (Section 1.3.2.7.3).

The problem of finding the optimum of S for a given value of C amounts to achieving collinearity between the gradients ∇S and ∇C of S and of C in $\mathcal{V}$, the scalar ratio between them being a Lagrange multiplier. In order to move towards such a solution from a trial position, the dependence of ∇S and ∇C on position in $\mathcal{V}$ must be represented. This involves the $\mathcal{N}\times\mathcal{N}$ Hessian matrices $\mathbf{H}(S)$ and $\mathbf{H}(C)$, whose size precludes their use in the whole of $\mathcal{V}$. Restricting the search to a smaller *search subspace* of dimension n spanned by $\{\mathbf{V}_i\}_{i=1,\ldots,n}$ we may build local quadratic models of S and C (Bryan & Skilling, 1980; Burch *et al.*, 1983) with respect to n coordinates $\mathbf{X}$ in that subspace:

$$\begin{aligned} S(\mathbf{X}) &= S(\mathbf{X}_0) + \mathbf{S}_0^T(\mathbf{X}-\mathbf{X}_0) \\ &\quad + \tfrac{1}{2}(\mathbf{X}-\mathbf{X}_0)^T\mathbf{H}_0(S)(\mathbf{X}-\mathbf{X}_0) \\ C(\mathbf{X}) &= C(\mathbf{X}_0) + \mathbf{C}_0^T(\mathbf{X}-\mathbf{X}_0) \\ &\quad + \tfrac{1}{2}(\mathbf{X}-\mathbf{X}_0)^T\mathbf{H}_0(C)(\mathbf{X}-\mathbf{X}_0). \end{aligned}$$

The coefficients of these linear models are given by scalar products:

$$[\mathbf{S}_0]_i = (\mathbf{V}_i, \nabla S)$$
$$[\mathbf{C}_0]_i = (\mathbf{V}_i, \nabla C)$$
$$[\mathbf{H}_0(S)]_{ij} = [\mathbf{V}_i, \mathbf{H}(S)\mathbf{V}_j]$$
$$[\mathbf{H}_0(C)]_{ij} = [\mathbf{V}_i, \mathbf{H}(C)\mathbf{V}_j]$$

which, by virtue of Parseval's theorem, may be evaluated either in real space or in reciprocal space (Bricogne, 1984). In doing so, special positions and reflections must be taken into account, as in Section 1.3.4.2.2.8. Scalar products involving S are best evaluated by real-space grid summation, because $\mathbf{H}(S)$ is *diagonal* in this representation; those involving C are best calculated by reciprocal-space summation, because $\mathbf{H}(C)$ is at worst 2×2 *block-diagonal* in this representation. Using these Hessian matrices in the wrong space would lead to prohibitively expensive convolutions instead of scalar (or at worst 2×2 matrix) multiplications.

1.3.4.4.7. *Derivatives for model refinement*

Since the origins of X-ray crystal structure analysis, the calculation of crystallographic Fourier series has been closely associated with the process of refinement. Fourier coefficients *with phases* were obtained for all or part of the measured reflections in the basis of some trial model for all or part of the structure, and Fourier syntheses were then used to complete and improve this initial model. This approach is clearly described in the classic paper by Bragg & West (1929), and was put into practice in the determination of the structures of topaz (Alston & West, 1929) and diopside (Warren & Bragg, 1929). Later, more systematic methods of arriving at a trial model were provided by the Patterson synthesis (Patterson, 1934, 1935*a*,*b*; Harker, 1936) and by isomorphous replacement (Robertson, 1935, 1936*c*). The role of Fourier syntheses, however, remained essentially unchanged [see Robertson (1937) for a review] until more systematic methods of structure refinement were introduced in the 1940s. A particularly good account of the processes of structure completion and refinement may be found in Chapters 15 and 16 of Stout & Jensen (1968).

It is beyond the scope of this section to review the vast topic of refinement methods: rather, it will give an account of those aspects of their development which have sought improved power by exploiting properties of the Fourier transformation. It is of more than historical interest that some recent advances in the crystallographic refinement of macromolecular structures had been anticipated by Cochran and Cruickshank in the early 1950s.

1.3.4.4.7.1. *The method of least squares*

Hughes (1941) was the first to use the already well established multivariate least-squares method (Whittaker & Robinson, 1944) to refine initial estimates of the parameters describing a model structure. The method gained general acceptance through the programming efforts of Friedlander *et al.* (1955), Sparks *et al.* (1956), Busing & Levy (1961), and others.

The Fourier relations between ρ and F (Section 1.3.4.2.2.6) are used to derive the 'observational equations' connecting the structure parameters $\{u_p\}_{p=1,\ldots,n}$ to the observations $\{|F_\mathbf{h}|^{\rm obs}, (\sigma^2_\mathbf{h})^{\rm obs}\}_{\mathbf{h}\in\mathscr{H}}$ comprising the amplitudes and their experimental variances for a set $\mathscr{H}$ of unique reflections.

The normal equations giving the corrections $\delta\mathbf{u}$ to the parameters are then

$$(\mathbf{A}^T\mathbf{W}\mathbf{A})\delta\mathbf{u} = -\mathbf{A}^T\mathbf{W}\Delta,$$

where

$$A_{\mathbf{h}p} = \frac{\partial|F^{\rm calc}_\mathbf{h}|}{\partial u_p}$$
$$\Delta_\mathbf{h} = |F^{\rm calc}_\mathbf{h}| - |F_\mathbf{h}|^{\rm obs}$$
$$\mathbf{W} = \text{diag } (W_\mathbf{h}) \quad \text{with} \quad W_\mathbf{h} = \frac{1}{(\sigma^2_\mathbf{h})^{\rm obs}}.$$

To calculate the elements of $\mathbf{A}$, write:

$$F = |F|\exp(i\varphi) = \alpha + i\beta;$$

hence

$$\frac{\partial|F|}{\partial u} = \frac{\partial\alpha}{\partial u}\cos\varphi + \frac{\partial\beta}{\partial u}\sin\varphi$$
$$= \mathscr{R}e\left[\frac{\partial F}{\partial u}\overline{\exp(i\varphi)}\right] = \mathscr{R}e\left[\overline{\frac{\partial F}{\partial u}}\exp(i\varphi)\right].$$

In the simple case of atoms with real-valued form factors and isotropic thermal agitation in space group $P1$,

$$F^{\rm calc}_\mathbf{h} = \sum_{j\in J} g_j(\mathbf{h})\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j),$$

where

$$g_j(\mathbf{h}) = Z_j f_j(\mathbf{h})\exp[-\tfrac{1}{4}B_j(d^*_\mathbf{h})^2],$$

Z_j being a fractional occupancy.

Positional derivatives with respect to $\mathbf{x}_j$ are given by

$$\frac{\partial F^{\rm calc}_\mathbf{h}}{\partial\mathbf{x}_j} = (2\pi i\mathbf{h})g_j(\mathbf{h})\exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j)$$
$$\frac{\partial|F^{\rm calc}_\mathbf{h}|}{\partial\mathbf{x}_j} = \mathscr{R}e[(-2\pi i\mathbf{h})g_j(\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j)\exp(i\varphi^{\rm calc}_\mathbf{h})]$$

so that the corresponding 3×1 subvector of the right-hand side of the normal equations reads:

$$-\sum_{\mathbf{h}\in\mathscr{H}} W_\mathbf{h}\frac{\partial|F^{\rm calc}_\mathbf{h}|}{\partial\mathbf{x}_j}(|F^{\rm calc}_\mathbf{h}| - |F_\mathbf{h}|^{\rm obs})$$
$$= -\mathscr{R}e\left[\sum_{\mathbf{h}\in\mathscr{H}} g_j(\mathbf{h})(-2\pi i\mathbf{h})W_\mathbf{h}(|F^{\rm calc}_\mathbf{h}| - |F_\mathbf{h}|^{\rm obs})\right.$$
$$\left.\times\exp(i\varphi^{\rm calc}_\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j)\right].$$

The setting up and solution of the normal equations lends itself well to computer programming and has the advantage of providing a thorough analysis of the accuracy of its results (Cruickshank, 1965*b*, 1970; Rollett, 1970). It is, however, an expensive task, of complexity $\propto n \times |\mathscr{H}|^2$, which is unaffordable for macromolecules.

1.3.4.4.7.2. *Booth's differential Fourier syntheses*

It was the use of Fourier syntheses in the completion of trial structures which provided the incentive to find methods for computing 2D and 3D syntheses efficiently, and led to the Beevers–Lipson strips. The limited accuracy of the latter caused the estimated positions of atoms (identified as peaks in the maps) to be somewhat in error. Methods were therefore sought to improve the accuracy with which the coordinates of the electron-density maxima could be determined. The naive method of peak-shape analysis from densities recalculated on a $3 \times 3 \times 3$ grid using high-accuracy trigonometric tables entailed 27 summations per atom.

Booth (1946*a*) suggested examining the rapidly varying derivatives of the electron density rather than its slowly varying values. If

$$\rho(\mathbf{x}) = \sum_{\mathbf{h}} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h}\cdot\mathbf{x})$$

then the gradient vector $\nabla_{\mathbf{x}}\rho$ of ρ at $\mathbf{x}^0$

$$(\nabla_{\mathbf{x}}\rho)(\mathbf{x}^0) = \sum_{\mathbf{h}} F_{\mathbf{h}}(-2\pi i\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}^0)$$

can be calculated by means of three Fourier summations from the 3×1 vector of Fourier coefficients

$$(-2\pi i\mathbf{h})F_{\mathbf{h}}.$$

Similarly, the Hessian matrix of ρ at $\mathbf{x}^0$

$$[(\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^T)\rho](\mathbf{x}^0) = \sum_{\mathbf{h}} F_{\mathbf{h}}(-4\pi^2\mathbf{h}\mathbf{h}^T) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x}^0)$$

can be calculated by six Fourier summations from the unique elements of the symmetric matrix of Fourier coefficients:

$$-4\pi^2 \begin{pmatrix} h^2 & hk & hl \\ hk & k^2 & kl \\ hl & kl & l^2 \end{pmatrix} F_{\mathbf{h}}.$$

The scalar maps giving the components of the gradient and Hessian matrix of ρ will be called differential syntheses of 1st order and 2nd order respectively. If $\mathbf{x}^0$ is approximately but not exactly a maximum of ρ, then the Newton–Raphson estimate of the true maximum $\mathbf{x}^*$ is given by:

$$\mathbf{x}^* = \mathbf{x}^0 - [[(\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^T)\rho](\mathbf{x}^0)]^{-1}[\nabla_{\mathbf{x}}\rho(\mathbf{x}^0)].$$

This calculation requires only nine accurate Fourier summations (instead of 27), and this number is further reduced to four if the peak is assumed to be spherically symmetrical.

The resulting positions are affected by series-termination errors in the differential syntheses. Booth (1945*c*, 1946*c*) proposed a 'back-shift correction' to eliminate them, and extended this treatment to the acentric case (Booth, 1946*b*). He cautioned against the use of an artificial temperature factor to fight series-termination errors (Brill *et al.*, 1939), as this could be shown to introduce coordinate errors by causing overlap between atoms (Booth, 1946*c*, 1947*a*,*b*).

Cruickshank was able to derive estimates for the standard uncertainties of the atomic coordinates obtained in this way (Cox & Cruickshank, 1948; Cruickshank, 1949*a*,*b*) and to show that they agreed with those provided by the least-squares method.

The calculation of differential Fourier syntheses was incorporated into the crystallographic programs of Ahmed & Cruickshank (1953*b*) and of Sparks *et al.* (1956).

1.3.4.4.7.3. *Booth's method of steepest descents*

Having defined the now universally adopted R factors (Booth, 1945*b*) as criteria of agreement between observed and calculated amplitudes or intensities, Booth proposed that R should be minimized with respect to the set of atomic coordinates $\{\mathbf{x}_j\}_{j\in J}$ by descending along the gradient of R in parameter space (Booth, 1947*c*,*d*). This 'steepest descents' procedure was compared with Patterson methods by Cochran (1948*d*).

When calculating the necessary derivatives, Booth (1948*a*, 1949) used the formulae given above in connection with least squares. This method was implemented by Qurashi (1949) and by Vand (1948, 1951) with parameter-rescaling modifications which made it very close to the least-squares method (Cruickshank, 1950; Qurashi & Vand, 1953; Qurashi, 1953).

1.3.4.4.7.4. *Cochran's Fourier method*

Cochran (1948*b*,*c*, 1951*a*) undertook to exploit an algebraic similarity between the right-hand side of the normal equations in the least-squares method on the one hand, and the expression for the coefficients used in Booth's differential syntheses on the other hand (see also Booth, 1948*a*). In doing so he initiated a remarkable sequence of formal and computational developments which are still actively pursued today.

Let $\rho_C(\mathbf{x})$ be the electron-density map corresponding to the current atomic model, with structure factors $|F_{\mathbf{h}}^{\rm calc}| \exp(i\varphi_{\mathbf{h}}^{\rm calc})$; and let $\rho_O(\mathbf{x})$ be the map calculated from observed moduli and calculated phases, *i.e.* with coefficients $\{|F_{\mathbf{h}}|^{\rm obs} \exp(i\varphi_{\mathbf{h}}^{\rm calc})\}_{\mathbf{h}\in\mathscr{H}}$. If there are enough data for ρ_C to have a resolved peak at each model atomic position $\mathbf{x}_j$, then

$$(\nabla_{\mathbf{x}}\rho_C)(\mathbf{x}_j) = \mathbf{0} \quad \text{for each } j \in J;$$

while if the calculated phases $\varphi_{\mathbf{h}}^{\rm calc}$ are good enough, ρ_O will also have peaks at each $\mathbf{x}_j$:

$$(\nabla_{\mathbf{x}}\rho_O)(\mathbf{x}_j) = \mathbf{0} \quad \text{for each } j \in J.$$

It follows that

$$\begin{aligned}[\nabla_{\mathbf{x}}(\rho_C - \rho_O)](\mathbf{x}_j) &= \sum_{\mathbf{h}}(-2\pi i\mathbf{h})[(|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs}) \exp(i\varphi_{\mathbf{h}}^{\rm calc})] \\ &\quad \times \exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j) \\ &= \mathbf{0} \text{ for each } j \in J,\end{aligned}$$

where the summation is over all reflections in $\mathscr{H}$ or related to $\mathscr{H}$ by space-group and Friedel symmetry (overlooking multiplicity factors!). This relation is less sensitive to series-termination errors than either of the previous two, since the spectrum of ρ_O could have been extrapolated beyond the data in $\mathscr{H}$ by using that of ρ_C [as in van Reijen (1942)] without changing its right-hand side.

Cochran then used the identity

$$\frac{\partial F_{\mathbf{h}}^{\rm calc}}{\partial \mathbf{x}_j} = (2\pi i\mathbf{h}) g_j(\mathbf{h}) \exp(2\pi i\mathbf{h}\cdot\mathbf{x}_j)$$

in the form

$$(-2\pi i\mathbf{h}) \exp(-2\pi i\mathbf{h}\cdot\mathbf{x}_j) = \frac{1}{g_j(\mathbf{h})} \overline{\frac{\partial F_{\mathbf{h}}^{\rm calc}}{\partial \mathbf{x}_j}}$$

to rewrite the previous relation as

$$\begin{aligned}&[\nabla_{\mathbf{x}}(\rho_C - \rho_O)](\mathbf{x}_j) \\ &\quad = \sum_{\mathbf{h}} \frac{1}{g_j(\mathbf{h})}(|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs})\mathscr{R}e\left[\overline{\frac{\partial F_{\mathbf{h}}^{\rm calc}}{\partial \mathbf{x}_j}} \exp(i\varphi_{\mathbf{h}}^{\rm calc})\right] \\ &\quad = \sum_{\mathbf{h}} \frac{1}{g_j(\mathbf{h})}(|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs}) \frac{\partial |F_{\mathbf{h}}^{\rm calc}|}{\partial \mathbf{x}_j} \\ &\quad = \mathbf{0} \quad \text{for each } j \in J\end{aligned}$$

(the operation $\mathscr{R}e$ [] on the first line being neutral because of Friedel symmetry). This is equivalent to the vanishing of the 3×1 subvector of the right-hand side of the normal equations associated to a least-squares refinement in which the weights would be

$$W_{\mathbf{h}} = \frac{1}{g_j(\mathbf{h})}.$$

Cochran concluded that, for equal-atom structures with $g_j(\mathbf{h}) = g(\mathbf{h})$ for all j, the positions $\mathbf{x}_j$ obtained by Booth's method applied to the difference map $\rho_O - \rho_C$ are such that they minimize the residual

$$\frac{1}{2}\sum_{\mathbf{h}} \frac{1}{g(\mathbf{h})}(|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs})^2$$

with respect to the atomic positions. If it is desired to minimize the residual of the ordinary least-squares method, then the differential

synthesis method should be applied to the *weighted* difference map

$$\sum_{\mathbf{h}} \frac{W_{\mathbf{h}}}{g(\mathbf{h})} (|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs}) \exp(i\varphi_{\mathbf{h}}^{\rm calc}).$$

He went on to show (Cochran, 1951*b*) that the refinement of temperature factors could also be carried out by inspecting appropriate derivatives of the weighted difference map.

This Fourier method was used by Freer *et al.* (1976) in conjunction with a stereochemical regularization procedure to refine protein structures.

1.3.4.4.7.5. *Cruickshank's modified Fourier method*

Cruickshank consolidated and extended Cochran's derivations in a series of classic papers (Cruickshank, 1949*b* , 1950, 1952, 1956). He was able to show that all the coefficients involved in the right-hand side and normal matrix of the least-squares method could be calculated by means of suitable differential Fourier syntheses *even when the atoms overlap*. This remarkable achievement lay essentially dormant until its independent rediscovery by Agarwal in 1978 (Section 1.3.4.4.7.6).

To ensure rigorous equivalence between the summations over $\mathbf{h} \in \mathscr{H}$ (in the expressions of least-squares right-hand side and normal matrix elements) and genuine Fourier summations, multiplicity-corrected weights were introduced by:

$$w_{\mathbf{h}} = \frac{1}{|G_{\mathbf{h}}|} W_{\mathbf{h}} \quad \text{if } \mathbf{h} \in G\mathbf{h} \quad \text{with } \mathbf{h} \in \mathscr{H},$$

$$w_{\mathbf{h}} = 0 \qquad \text{otherwise,}$$

where $G\mathbf{h}$ denotes the orbit of $\mathbf{h}$ and $G_{\mathbf{h}}$ its isotropy subgroup (Section 1.3.4.2.2.5). Similarly, derivatives with respect to parameters of symmetry-unique atoms were expressed, *via* the chain rule, as sums over the orbits of these atoms.

Let $p = 1, \ldots, n$ be the label of a parameter u_p belonging to atoms with label j. Then Cruickshank showed that the pth element of the right-hand side of the normal equations can be obtained as $D_{p,j}(\mathbf{x}_j)$, where $D_{p,j}$ is a differential synthesis of the form

$$D_{p,j}(\mathbf{x}) = \sum_{\mathbf{h}} P_p(\mathbf{h}) g_j(\mathbf{h}) w_{\mathbf{h}} (|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs}) \times \exp(i\varphi_{\mathbf{h}}^{\rm calc}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

with $P_p(\mathbf{h})$ a polynomial in (h, k, l) depending on the type of parameter p. The correspondence between parameter type and the associated polynomial extends Booth's original range of differential syntheses, and is recapitulated in the following table.

Parameter type	$P(h,k,l)$
x coordinate	$-2\pi ih$
y coordinate	$-2\pi ik$
z coordinate	$-2\pi il$
B isotropic	$-\frac{1}{4}(d_{\mathbf{h}}^*)^2$
B^{11} anisotropic	$-h^2$
B^{12} anisotropic	$-hk$
B^{13} anisotropic	$-hl$
B^{22} anisotropic	$-k^2$
B^{23} anisotropic	$-kl$
B^{33} anisotropic	$-l^2$.

Unlike Cochran's original heuristic argument, this result does not depend on the atoms being resolved.

Cruickshank (1952) also considered the elements of the normal matrix, of the form

$$\sum_{\mathbf{h}} w_{\mathbf{h}} \frac{\partial |F_{\mathbf{h}}^{\rm calc}|}{\partial u_p} \frac{\partial |F_{\mathbf{h}}^{\rm calc}|}{\partial u_q}$$

associated with positional parameters. The 3×3 block for parameters $\mathbf{x}_j$ and $\mathbf{x}_k$ may be written

$$\sum_{\mathbf{h}} w_{\mathbf{h}} (\mathbf{h}\mathbf{h}^T) \mathscr{R}e[(-2\pi i) g_j(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_j) \exp(i\varphi_{\mathbf{h}}^{\rm calc})] \times \mathscr{R}e[(-2\pi i) g_k(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_k) \exp(i\varphi_{\mathbf{h}}^{\rm calc})]$$

which, using the identity

$$\mathscr{R}e(z_1)\mathscr{R}e(z_2) = \tfrac{1}{2}[\mathscr{R}e(z_1 z_2) + \mathscr{R}e(z_1 \overline{z_2})],$$

becomes

$$2\pi^2 \sum_{\mathbf{h}} w_{\mathbf{h}} (\mathbf{h}\mathbf{h}^T) g_j(\mathbf{h}) g_k(\mathbf{h}) \times \{\exp[-2\pi i \mathbf{h} \cdot (\mathbf{x}_j - \mathbf{x}_k)] - \exp(2i\varphi_{\mathbf{h}}^{\rm calc}) \exp[-2\pi i \mathbf{h} \cdot (\mathbf{x}_j + \mathbf{x}_k)]\}$$

(Friedel's symmetry makes $\mathscr{R}e$ redundant on the last line). Cruickshank argued that the first term would give a good approximation to the diagonal blocks of the normal matrix and to those off-diagonal blocks for which $\mathbf{x}_j$ and $\mathbf{x}_k$ are close. On this basis he was able to justify the 'n-shift rule' of Shoemaker *et al.* (1950). Cruickshank gave this derivation in a general space group, but using a very terse notation which somewhat obscures it. Using the symmetrized trigonometric structure-factor kernel Ξ^- of Section 1.3.4.2.2.9 and its multiplication formula, the above expression is seen to involve the values of a Fourier synthesis at points of the form $\mathbf{x}_j \pm S_g(\mathbf{x}_k)$.

Cruickshank (1956) showed that this analysis could also be applied to the refinement of temperature factors.

These two results made it possible to obtain all coefficients involved in the normal equations by looking up the values of certain differential Fourier syntheses at $\mathbf{x}_j$ or at $\mathbf{x}_j \pm S_g(\mathbf{x}_k)$. At the time this did not confer any superiority over the standard form of the least-squares procedure, because the accurate computation of Fourier syntheses was an expensive operation. The modified Fourier method was used by Truter (1954) and by Ahmed & Cruickshank (1953*a*), and was incorporated into the program system described by Cruickshank *et al.* (1961). A more recent comparison with the least-squares method was made by Dietrich (1972).

There persisted, however, some confusion about the nature of the relationship between Fourier and least-squares methods, caused by the extra factors $g_j(\mathbf{h})$ which make it necessary to compute a differential synthesis for each type of atom. This led Cruickshank to conclude that 'in spite of their remarkable similarities the least-squares and modified-Fourier methods are fundamentally distinct'.

1.3.4.4.7.6. *Agarwal's FFT implementation of the Fourier method*

Agarwal (1978) rederived and completed Cruickshank's results at a time when the availability of the FFT algorithm made the Fourier method of calculating the coefficients of the normal equations much more economical than the standard method, especially for macromolecules.

As obtained by Cruickshank, the modified Fourier method required a full 3D Fourier synthesis

– for each type of parameter, since this determines [*via* the polynomial $P_p(\mathbf{h})$] the type of differential synthesis to be computed;

– for each type of atom $j \in J$, since the coefficients of the differential synthesis must be multiplied by $g_j(\mathbf{h})$.

Agarwal disposed of the latter dependence by pointing out that the multiplication involved is equivalent to a real-space *convolution* between the differential synthesis and $\sigma_j(\mathbf{x})$, the standard electron density ρ_j for atom type j (Section 1.3.4.2.1.2) smeared by the isotropic thermal agitation of that atom. Since σ_j is *localized*, this convolution involves only a small number of grid points. The requirement of a distinct differential synthesis for each parameter type, however, continued to hold, and created some difficulties at the FFT level because the symmetries of differential syntheses are more complex than ordinary space-group symmetries. Jack & Levitt (1978) sought to avoid the calculation of difference syntheses by using instead finite differences calculated from ordinary Fourier or difference Fourier maps.

In spite of its complication, this return to the Fourier implementation of the least-squares method led to spectacular increases in speed (Isaacs & Agarwal, 1978; Agarwal, 1980; Baker & Dodson, 1980) and quickly gained general acceptance (Dodson, 1981; Isaacs, 1982*a*,*b*, 1984).

1.3.4.4.7.7. *Lifchitz's reformulation*

Lifchitz [see Agarwal *et al.* (1981), Agarwal (1981)] proposed that the idea of treating certain multipliers in Cruickshank's modified differential Fourier syntheses by means of a convolution in real space should be applied not only to $g_j(\mathbf{h})$, but also to the polynomials $P_p(\mathbf{h})$ which determine the type of differential synthesis being calculated. This leads to convoluting $\partial\sigma_j/\partial u_p$ with the *same* ordinary weighted difference Fourier synthesis, rather than σ_j with the differential synthesis of type p. In this way, a *single* Fourier synthesis, with *ordinary* (scalar) symmetry properties, needs be computed; the parameter type and atom type both intervene through the function $\partial\sigma_j/\partial u_p$ with which it is convoluted. This approach has been used as the basis of an efficient general-purpose least-squares refinement program for macromolecular structures (Tronrud *et al.*, 1987).

This rearrangement amounts to using the fact (Section 1.3.2.3.9.7) that convolution commutes with differentiation. Let

$$D(\mathbf{x}) = \sum_{\mathbf{h}} w_{\mathbf{h}}(|F_{\mathbf{h}}^{\rm calc}| - |F_{\mathbf{h}}|^{\rm obs}) \exp(i\varphi_{\mathbf{h}}^{\rm calc}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x})$$

be the inverse-variance weighted difference map, and let us assume that parameter u_p belongs to atom j. Then the Agarwal form for the pth component of the right-hand side of the normal equations is

$$\left(\frac{\partial D}{\partial u_p} * \sigma_j\right)(x_j),$$

while the Lifchitz form is

$$\left(D * \frac{\partial\sigma_j}{\partial u_p}\right)(\mathbf{x}_j).$$

1.3.4.4.7.8. *A simplified derivation*

A very simple derivation of the previous results will now be given, which suggests the possibility of many generalizations.

The weighted difference map $D(\mathbf{x})$ has coefficients $D_{\mathbf{h}}$ which are the gradients of the global residual with respect to each $F_{\mathbf{h}}^{\rm calc}$:

$$D_{\mathbf{h}} = \frac{\partial R}{\partial A_{\mathbf{h}}^{\rm calc}} + i\frac{\partial R}{\partial B_{\mathbf{h}}^{\rm calc}}.$$

By the chain rule, a variation of each $F_{\mathbf{h}}^{\rm calc}$ by $\delta F_{\mathbf{h}}^{\rm calc}$ will result in a variation of R by δR with

$$\delta R = \sum_{\mathbf{h}}\left[\frac{\partial R}{\partial A_{\mathbf{h}}^{\rm calc}}\delta A_{\mathbf{h}}^{\rm calc} + \frac{\partial R}{\partial B_{\mathbf{h}}^{\rm calc}}\delta B_{\mathbf{h}}^{\rm calc}\right] = \mathscr{R}e\sum_{\mathbf{h}}[\overline{D_{\mathbf{h}}}\delta F_{\mathbf{h}}^{\rm calc}].$$

The $\mathscr{R}e$ operation is superfluous because of Friedel symmetry, so that δR may be simply written in terms of the Hermitian scalar product in $\ell^2(\mathbb{Z}^3)$:

$$\delta R = (\mathbf{D}, \delta\mathbf{F}^{\rm calc}).$$

If $\rho^{\rm calc}$ is the transform of $\delta\mathbf{F}^{\rm calc}$, we have also by Parseval's theorem

$$\delta R = (D, \delta\rho^{\rm calc}).$$

We may therefore write

$$D(\mathbf{x}) = \frac{\partial R}{\partial\rho^{\rm calc}(\mathbf{x})},$$

which states that $D(\mathbf{x})$ is the functional derivative of R with respect to $\rho^{\rm calc}$.

The right-hand side of the normal equations has $\partial R/\partial u_p$ for its pth element, and this may be written

$$\frac{\partial R}{\partial u_p} = \int_{\mathbb{R}^3/\mathbb{Z}^3}\frac{\partial R}{\partial\rho^{\rm calc}(\mathbf{x})}\frac{\partial\rho^{\rm calc}(\mathbf{x})}{\partial u_p}{\rm d}^2\mathbf{x} = \left(D, \frac{\partial\rho^{\rm calc}}{\partial u_p}\right).$$

If u_p belongs to atom j, then

$$\frac{\partial\rho^{\rm calc}}{\partial u_p} = \frac{\partial(\tau_{\mathbf{x}_j}\sigma_j)}{\partial u_p} = \tau_{\mathbf{x}_j}\left(\frac{\partial\sigma_j}{\partial u_p}\right);$$

hence

$$\frac{\partial R}{\partial u_p} = \left(D, \tau_{\mathbf{x}_j}\left(\frac{\partial\sigma_j}{\partial u_p}\right)\right).$$

By the identity of Section 1.3.2.4.3.5, this is identical to Lifchitz's expression $(D * \partial\sigma_j/\partial u_p)(\mathbf{x}_j)$. The present derivation in terms of scalar products [see Brünger (1989) for another presentation of it] is conceptually simpler, since it invokes only the chain rule [other uses of which have been reviewed by Lunin (1985)] and Parseval's theorem; economy of computation is obviously related to the good localization of $\partial\rho^{\rm calc}/\partial u_p$ compared to $\partial F^{\rm calc}/\partial u_p$. Convolutions, whose meaning is less clear, are no longer involved; they were a legacy of having first gone over to reciprocal space *via* differential syntheses in the 1940s.

Cast in this form, the calculation of derivatives by FFT methods appears as a particular instance of the procedure described in connection with variational techniques (Section 1.3.4.4.6) to calculate the coefficients of local quadratic models in a search subspace; this is far from surprising since varying the electron density through a variation of the parameters of an atomic model is a particular case of the 'free' variations considered by the variational approach. The latter procedure would accommodate in a very natural fashion the joint consideration of an energetic (Jack & Levitt, 1978; Brünger *et al.*, 1987; Brünger, 1988; Brünger *et al.*, 1989; Kuriyan *et al.*, 1989) or stereochemical (Konnert, 1976; Sussman *et al.*, 1977; Konnert & Hendrickson, 1980; Hendrickson & Konnert, 1980; Tronrud *et al.*, 1987) restraint function (which would play the role of S) and of the crystallographic residual (which would be C). It would even have over the latter the superiority of affording a genuine *second-order* approximation, albeit only in a subspace, hence the ability of detecting negative curvature and the resulting bifurcation behaviour (Bricogne, 1984). Current methods are unable to do this because they use only *first-order* models, and this is known to degrade severely the overall efficiency of the refinement process.

1.3.4.4.7.9. *Discussion of macromolecular refinement techniques*

The impossibility of carrying out a full-matrix least-squares refinement of a macromolecular crystal structure, caused by excessive computational cost and by the paucity of observations, led Diamond (1971) to propose a real-space refinement method in which stereochemical knowledge was used to keep the number of free parameters to a minimum. Refinement took place by a least-squares fit between the 'observed' electron-density map and a model density consisting of Gaussian atoms. This procedure, coupled to iterative recalculation of the phases, led to the first highly refined protein structures obtained without using full-matrix least squares (Huber *et al.*, 1974; Bode & Schwager, 1975; Deisenhofer & Steigemann, 1975; Takano, 1977*a*,*b*).

Real-space refinement takes advantage of the localization of atoms (each parameter interacts only with the density near the atom to which it belongs) and gives the most immediate description of stereochemical constraints. A disadvantage is that fitting the 'observed' electron density amounts to treating the phases of the structure factors as observed quantities, and to ignoring the experimental error estimates on their moduli. The method is also much more vulnerable to series-termination errors and accidentally missing data than the least-squares method. These objections led to the progressive disuse of Diamond's method, and to a switch towards reciprocal-space least squares following Agarwal's work.

The connection established above between the Cruickshank–Agarwal modified Fourier method and the simple use of the chain rule affords a partial refutation to both the premises of Diamond's method and to the objections made against it:

(i) it shows that refinement *can* be performed through localized computations in real space *without* having to treat the phases as observed quantities;

(ii) at the same time, it shows that measurement errors on the moduli *can* be fully utilized in real space, *via* the Fourier synthesis of the functional derivative $\partial R/\partial \rho^{\rm calc}(\mathbf{x})$ or by means of the coefficients of a quadratic model of R in a search subspace.

1.3.4.4.7.10. *Sampling considerations*

The calculation of the inner products $(D, \partial\rho^{\rm calc}/\partial u_p)$ from a *sampled* gradient map D requires even more caution than that of structure factors *via* electron-density maps described in Section 1.3.4.4.5, because the functions $\partial\sigma_j/\partial u_p$ have transforms which extend even further in reciprocal space than the σ_j themselves. Analytically, if the σ_j are Gaussians, the $\partial\sigma_j/\partial u_p$ are finite sums of multivariate Hermite functions (Section 1.3.2.4.4.2) and hence the same is true of their transforms. The difference map D must therefore be finely sampled and the relation between error and sampling rate may be investigated as in Section 1.3.4.4.5. An examination of the sampling rates commonly used (*e.g.* one third of the resolution) shows that they are insufficient. Tronrud *et al.* (1987) propose to relax this requirement by applying an artificial temperature factor to σ_j (*cf.* Section 1.3.4.4.5) and the negative of that temperature factor to D, a procedure of questionable validity because the latter 'sharpening' operation is ill defined [the function $\exp(\|\mathbf{x}\|^2)$ does not define a tempered distribution, so the associativity properties of convolution may be lost]. A more robust procedure would be to compute the scalar product by means of a more sophisticated numerical quadrature formula than a mere grid sum.

1.3.4.4.8. *Miscellaneous correlation functions*

Certain correlation functions can be useful to detect the presence of multiple copies of the same molecule (known or unknown) in the asymmetric unit of a crystal of unknown structure.

Suppose that a crystal contains one or several copies of a molecule $\mathcal{M}$ in its asymmetric unit. If $\mu(\mathbf{x})$ is the electron density of that molecule in some reference position and orientation, then

$$\rho^0 = \sum_{j\in J}\left[\sum_{g\in G} S_g^{\#}(T_j^{\#}\mu)\right],$$

where $T_j : \mathbf{x} \longmapsto \mathbf{C}_j\mathbf{x} + \mathbf{d}_j$ describes the placement of the jth copy of the molecule with respect to the reference copy. It is assumed that each such copy is in a general position, so that there is no isotropy subgroup.

The methods of Section 1.3.4.2.2.9 (with ρ_j replaced by $C_j^{\#}\mu$, and $\mathbf{x}_j$ by $\mathbf{d}_j$) lead to the following expression for the autocorrelation of ρ^0:

$$\check{\rho}^0 * \rho^0 = \sum_{j_1}\sum_{j_2}\sum_{g_1}\sum_{g_2} \boldsymbol{\tau}_{S_{g_2}(\mathbf{d}_{j_2})-S_{g_1}(\mathbf{d}_{j_1})} \times [(R_{g_1}^{\#}C_{j_1}^{\#}\check{\mu}) * (R_{g_2}^{\#}C_{j_2}^{\#}\mu)].$$

If μ is unknown, consider the subfamily σ of terms with $j_1 = j_2 = j$ and $g_1 = g_2 = g$:

$$\sigma = \sum_j\sum_g R_g^{\#}C_j^{\#}(\check{\mu} * \mu).$$

The scalar product $(\sigma, R^{\#}\sigma)$ in which R is a variable rotation will have a peak whenever

$$R = (R_{g_1}C_{j_1})^{-1}(R_{g_2}C_{j_2})$$

since two copies of the 'self-Patterson' $\check{\mu} * \mu$ of the molecule will be brought into coincidence. If the interference from terms in the Patterson $\pi = r * \check{\rho}^0 * \rho^0$ other than those present in σ is not too serious, the 'self-rotation function' $(\pi, R^{\#}\pi)$ (Rossmann & Blow, 1962; Crowther, 1972) will show the same peaks, from which the rotations $\{C_j\}_{j\in J}$ may be determined, either individually or jointly if for instance they form a group.

If μ is known, then its self-Patterson $\check{\mu} * \mu$ may be calculated, and the C_j may be found by examining the 'cross-rotation function' $[\pi, R^{\#}(\check{\mu} * \mu)]$ which will have peaks at $R = R_gC_j, g \in G, j \in J$. Once the C_j are known, then the various copies $C_j^{\#}\mu$ of $\mathcal{M}$ may be Fourier-analysed into structure factors:

$$M_j(\mathbf{h}) = \bar{\mathcal{F}}[C_j^{\#}\mu](\mathbf{h}).$$

The cross terms with $j_1 \neq j_2, g_1 \neq g_2$ in $\check{\rho}^0 * \rho^0$ then contain 'motifs'

$$(R_{g_1}^{\#}C_{j_1}^{\#}\check{\mu}) * (R_{g_2}^{\#}C_{j_2}^{\#}\mu),$$

with Fourier coefficients

$$\overline{M_{j_1}(\mathbf{R}_{g_1}^T\mathbf{h})} \times M_{j_2}(\mathbf{R}_{g_2}^T\mathbf{h}),$$

translated by $S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})$. Therefore the 'translation functions' (Crowther & Blow, 1967)

$$\mathcal{T}_{j_1g_1,j_2g_2}(\mathbf{s}) = \sum_{\mathbf{h}}|F_{\mathbf{h}}|^2\overline{M_{j_1}(\mathbf{R}_{g_1}^T\mathbf{h})} \times M_{j_2}(\mathbf{R}_{g_2}^T\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{s})$$

will have peaks at $\mathbf{s} = S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})$ corresponding to the detection of these motifs.

1.3.4.5. *Related applications*

1.3.4.5.1. *Helical diffraction*

The theory of diffraction by helical structures (Cochran *et al.*, 1952; Klug *et al.*, 1958) has played an important part in the study of polypeptides, of nucleic acids and of tobacco mosaic virus.

1.3.4.5.1.1. *Circular harmonic expansions in polar coordinates*

Let $f = f(x, y)$ be a reasonably regular function in two-dimensional real space. Going over to polar coordinates

$$x = r\cos\varphi \quad y = r\sin\varphi$$

and writing, by slight misuse of notation, $f(r, \varphi)$ for $f(r\cos\varphi, r\sin\varphi)$ we may use the periodicity of f with respect to φ to expand it as a Fourier series (Byerly, 1893):

$$f(r,\varphi) = \sum_{n\in\mathbb{Z}} f_n(r)\exp(in\varphi)$$

with

$$f_n(r) = \frac{1}{2\pi}\int_0^{2\pi} f(r,\varphi)\exp(-in\varphi)\ \mathrm{d}\varphi.$$

Similarly, in reciprocal space, if $F = F(\xi, \eta)$ and if

$$\xi = R\cos\psi \quad \eta = R\sin\psi$$

then

$$F(R,\psi) = \sum_{n\in\mathbb{Z}} i^n F_n(R)\exp(in\psi)$$

with

$$F_n(R) = \frac{1}{2\pi i^n}\int_0^{2\pi} F(R,\psi)\exp(-in\psi)\ \mathrm{d}\psi,$$

where the phase factor i^n has been introduced for convenience in the forthcoming step.

1.3.4.5.1.2. *The Fourier transform in polar coordinates*

The Fourier transform relation between f and F may then be written in terms of f_n's and F_n's. Observing that $\xi x + \eta y = Rr\cos(\varphi - \psi)$, and that (Watson, 1944)

$$\int_0^{2\pi} \exp(iX\cos\theta + in\theta)\ \mathrm{d}\theta = 2\pi i^n J_n(X),$$

we obtain:

$$\begin{aligned} F(R,\psi) &= \int_0^{\infty}\int_0^{2\pi}\left[\sum_{n\in\mathbb{Z}} f_n(r)\exp(in\varphi)\right] \\ &\quad\times \exp[2\pi iRr\cos(\varphi-\psi)]r\ \mathrm{d}r\ \mathrm{d}\varphi \\ &= \sum_{n\in\mathbb{Z}} i^n\left[\int_0^{\infty} f_n(r)J_n(2\pi Rr)2\pi r\ \mathrm{d}r\right]\exp(in\psi); \end{aligned}$$

hence, by the uniqueness of the Fourier expansion of F:

$$F_n(R) = \int_0^{\infty} f_n(r)J_n(2\pi Rr)2\pi r\ \mathrm{d}r.$$

The inverse Fourier relationship leads to

$$f_n(r) = \int_0^{\infty} F_n(R)J_n(2\pi rR)2\pi R\ \mathrm{d}R.$$

The integral transform involved in the previous two equations is called the *Hankel transform* (see *e.g.* Titchmarsh, 1922; Sneddon, 1972) of order n.

1.3.4.5.1.3. *The transform of an axially periodic fibre*

Let ρ be the electron-density distribution in a fibre, which is assumed to have translational periodicity with period 1 along z, and to have compact support with respect to the (x, y) coordinates. Thus ρ may be written

$$\rho = \left[\delta_x \otimes \delta_y \otimes \left(\sum_{k\in\mathbb{Z}}\delta_{(k)}\right)_z\right] * \rho^0,$$

where $\rho^0 = \rho^0(x, y, z)$ is the motif.

By the tensor product property, the inverse Fourier transform $F = \bar{\mathscr{F}}_{xyz}[\rho]$ may be written

$$F = \left[1_\xi \otimes 1_\eta \otimes \left(\sum_{l\in\mathbb{Z}}\delta_{(l)}\right)_\zeta\right] \times \bar{\mathscr{F}}[\rho^0]$$

and hence consists of 'layers' labelled by l:

$$F = \sum_{l\in\mathbb{Z}} F(\xi,\eta,l)(\delta_{(l)})_\zeta$$

with

$$F(\xi,\eta,l) = \int_0^1 \bar{\mathscr{F}}_{xy}[\rho^0](\xi,\eta,z)\exp(2\pi ilz)\ \mathrm{d}z.$$

Changing to polar coordinates in the (x, y) and (ξ, η) planes decomposes the calculation of F from ρ into the following steps:

$$\begin{aligned} g_{nl}(r) &= \frac{1}{2\pi}\int_0^{2\pi}\int_0^1 \rho(r,\varphi,z)\exp[i(-n\varphi + 2\pi lz)]\ \mathrm{d}\varphi\ \mathrm{d}z \\ G_{nl}(R) &= \int_0^{\infty} g_{nl}(r)J_n(2\pi Rr)2\pi r\ \mathrm{d}r \\ F(R,\psi,l) &= \sum_{n\in\mathbb{Z}} i^n G_{nl}(R)\exp(in\psi) \end{aligned}$$

and the calculation of ρ from F into:

$$\begin{aligned} G_{nl}(R) &= \frac{1}{2\pi i^n}\int_0^{2\pi} F(R,\psi,l)\exp(-in\psi)\ \mathrm{d}\psi \\ g_{nl}(r) &= \int_0^{\infty} G_{nl}(R)J_n(2\pi rR)2\pi R\ \mathrm{d}R \\ \rho(r,\varphi,z) &= \sum_{n\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} g_{nl}(r)\exp[i(n\varphi - 2\pi lz)]. \end{aligned}$$

These formulae are seen to involve a 2D Fourier series with respect to the two periodic coordinates φ and z, and Hankel transforms along the radial coordinates. The two periodicities in φ and z are independent, so that all combinations of indices (n, l) occur in the Fourier summations.

1.3.4.5.1.4. *Helical symmetry and associated selection rules*

Helical symmetry involves a 'clutching' between the two (hitherto independent) periodicities in φ (period 2π) and z (period 1) which causes a subdivision of the period lattice and hence a decimation (governed by 'selection rules') of the Fourier coefficients.

Let $\mathbf{i}$ and $\mathbf{j}$ be the basis vectors along $\varphi/2\pi$ and z. The integer lattice with basis $(\mathbf{i}, \mathbf{j})$ is a period lattice for the (φ, z) dependence of the electron density ρ of an axially periodic fibre considered in Section 1.3.4.5.1.3:

$$\rho(r, \varphi + 2\pi k_1, z + k_2) = \rho(r, \varphi, z).$$

Suppose the fibre now has *helical* symmetry, with u copies of the same molecule in t turns, where g.c.d. $(u, t) = 1$. Using the Euclidean algorithm, write $u = \lambda t + \mu$ with λ and μ positive integers and $\mu < t$. The period lattice for the (φ, z) dependence of ρ may be defined in terms of the new basis vectors:

I, joining subunit 0 to subunit l in the same turn;

J, joining subunit 0 to subunit λ after wrapping around.

In terms of the original basis

$$\mathbf{I} = \frac{t}{u}\mathbf{i} + \frac{1}{u}\mathbf{j}, \quad \mathbf{J} = \frac{-\mu}{u}\mathbf{i} + \frac{\lambda}{u}\mathbf{j}.$$

If α and β are coordinates along **I** and **J**, respectively,

$$\begin{pmatrix} \varphi/2\pi \\ z \end{pmatrix} = \frac{1}{u}\begin{pmatrix} t & -\mu \\ 1 & \lambda \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

or equivalently

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \lambda & \mu \\ -1 & t \end{pmatrix}\begin{pmatrix} \varphi/2\pi \\ z \end{pmatrix}.$$

By Fourier transformation,

$$\left(\frac{\varphi}{2\pi}, z\right) \Leftrightarrow (-n, l)$$
$$(\alpha, \beta) \Leftrightarrow (m, p)$$

with the transformations between indices given by the contragredients of those between coordinates, *i.e.*

$$\begin{pmatrix} n \\ l \end{pmatrix} = \begin{pmatrix} -\lambda & 1 \\ -\mu & t \end{pmatrix}\begin{pmatrix} m \\ p \end{pmatrix}$$

and

$$\begin{pmatrix} m \\ p \end{pmatrix} = \frac{1}{u}\begin{pmatrix} -t & 1 \\ \mu & \lambda \end{pmatrix}\begin{pmatrix} n \\ l \end{pmatrix}.$$

It follows that

$$l = tn + um,$$

or alternatively that

$$\mu n = up - \lambda l,$$

which are two equivalent expressions of the *selection rules* describing the decimation of the transform. These rules imply that only certain orders n contribute to a given layer l.

The 2D Fourier analysis may now be performed by analysing a *single* subunit referred to coordinates α and β to obtain

$$h_{m,p}(r) = \int_0^1\int_0^1 \rho(r, \alpha, \beta)\exp[2\pi i(m\alpha + p\beta)]\ \mathrm{d}\alpha\ \mathrm{d}\beta$$

and then reindexing to get only the allowed g_{nl}'s by

$$g_{nl}(r) = uh_{-\lambda m+p,\ \mu m+tp}(r).$$

This is u times faster than analysing u subunits with respect to the (φ, z) coordinates.

1.3.4.5.2. *Application to probability theory and direct methods*

The Fourier transformation plays a central role in the branch of probability theory concerned with the limiting behaviour of sums of large numbers of independent and identically distributed random variables or random vectors. This privileged role is a consequence of the convolution theorem and of the 'moment-generating' properties which follow from the exchange between differentiation and multiplication by monomials. When the limit theorems are applied to the calculation of joint probability distributions of structure factors, which are themselves closely related to the Fourier transformation, a remarkable phenomenon occurs, which leads to the saddlepoint approximation and to the maximum-entropy method.

1.3.4.5.2.1. *Analytical methods of probability theory*

The material in this section is not intended as an introduction to probability theory [for which the reader is referred to Cramér (1946), Petrov (1975) or Bhattacharya & Rao (1976)], but only as an illustration of the role played by the Fourier transformation in certain specific areas which are used in formulating and implementing direct methods of phase determination.

(*a*) *Convolution of probability densities*

The addition of independent random variables or vectors leads to the convolution of their probability distributions: if $\mathbf{X}_1$ and $\mathbf{X}_2$ are two n-dimensional random vectors independently distributed with probability densities P_1 and P_2, respectively, then their sum $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ has probability density $\mathscr{P}$ given by

$$\begin{aligned}\mathscr{P}(\mathbf{X}) &= \int_{\mathbf{R}^n} P_1(\mathbf{X}_1)P_2(\mathbf{X} - \mathbf{X}_1)\ \mathrm{d}^n\mathbf{X}_1 \\ &= \int_{\mathbf{R}^n} P_1(\mathbf{X} - \mathbf{X}_2)P_2(\mathbf{X}_2)\ \mathrm{d}^n\mathbf{X}_2\end{aligned}$$

i.e.

$$\mathscr{P} = P_1 * P_2.$$

This result can be extended to the case where P_1 and P_2 are singular measures (distributions of order zero, Section 1.3.2.3.4) and do not have a density with respect to the Lebesgue measure in $\mathbb{R}^n$.

(*b*) *Characteristic functions*

This convolution can be turned into a simple multiplication by considering the Fourier transforms (called the *characteristic functions*) of P_1, P_2 and $\mathscr{P}$, defined with a slightly different normalization in that there is no factor of 2π in the exponent (see Section 1.3.2.4.5), *e.g.*

$$C(\mathbf{t}) = \int_{\mathbf{R}^n} P(\mathbf{X})\exp(i\mathbf{t}\cdot\mathbf{X})\ \mathrm{d}^n\mathbf{X}.$$

Then by the convolution theorem

$$\mathscr{C}(\mathbf{t}) = C_1(\mathbf{t}) \times C_2(\mathbf{t}),$$

so that $\mathscr{P}(\mathbf{X})$ may be evaluated by Fourier inversion of its characteristic function as

$$\mathscr{P}(\mathbf{X}) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} C_1(\mathbf{t})C_2(\mathbf{t})\exp(-i\mathbf{t}\cdot\mathbf{X})\ \mathrm{d}^n\mathbf{t}$$

(see Section 1.3.2.4.5 for the normalization factors).

It follows from the differentiation theorem that the partial derivatives of the characteristic function $C(\mathbf{t})$ at $\mathbf{t} = \mathbf{0}$ are related to the moments of a distribution $\boldsymbol{P}$ by the identities

$$\begin{aligned}\mu_{r_1 r_2 \ldots r_n} &\equiv \int_D P(\mathbf{X})X_1^{r_1}X_2^{r_2}\ldots X_n^{r_n}\ \mathrm{d}^n\mathbf{X} \\ &= i^{-(r_1+\ldots+r_n)}\frac{\partial^{r_1+\ldots+r_n}C}{\partial t_1^{r_1}\ldots\partial t_n^{r_n}}\bigg|_{\mathbf{t}=\mathbf{0}}\end{aligned}$$

for any n-tuple of non-negative integers $(r_1, r_2, \ldots, r_n)$.

(*c*) *Moment-generating functions*

The above relation can be freed from powers of i by defining (at least formally) the *moment-generating function*:

$$M(\mathbf{t}) = \int_{\mathbb{R}^n} P(\mathbf{X}) \exp(\mathbf{t}\cdot\mathbf{X})\ \mathrm{d}^n\mathbf{X}$$

which is related to $C(\mathbf{t})$ by $C(\mathbf{t}) = \boldsymbol{M}(i\mathbf{t})$ so that the inversion formula reads

$$\mathscr{P}(\mathbf{X}) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} M_1(i\mathbf{t})M_2(i\mathbf{t})\exp(-i\mathbf{t}\cdot\mathbf{X})\ \mathrm{d}^n\mathbf{t}.$$

The moment-generating function is well defined, in particular, for any probability distribution with compact support, in which case it may be continued analytically from a function over $\mathbb{R}^n$ into an entire function of n *complex* variables by virtue of the Paley–Wiener theorem (Section 1.3.2.4.2.10). Its moment-generating properties are summed up in the following relations:

$$\mu_{r_1 r_2 \ldots r_n} = \left.\frac{\partial^{r_1+\ldots+r_n} M}{\partial t_1^{r_1}\ldots\partial t_n^{r_n}}\right|_{\mathbf{t}=\mathbf{0}}.$$

(*d*) *Cumulant-generating functions*

The multiplication of moment-generating functions may be further simplified into the addition of their logarithms:

$$\log \mathscr{M} = \log M_1 + \log M_2,$$

or equivalently of the coefficients of their Taylor series at $\mathbf{t} = \mathbf{0}$, *viz*:

$$\kappa_{r_1 r_2 \ldots r_n} = \left.\frac{\partial^{r_1+\ldots+r_n}(\log M)}{\partial t_1^{r_1}\ldots\partial t_n^{r_n}}\right|_{\mathbf{t}=\mathbf{0}}.$$

These coefficients are called *cumulants*, since they add when the independent random vectors to which they belong are added, and $\log M$ is called the *cumulant-generating function.* The inversion formula for $\mathscr{P}$ then reads

$$\mathscr{P}(\mathbf{X}) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \exp[\log M_1(i\mathbf{t}) + \log M_2(i\mathbf{t}) - i\mathbf{t}\cdot\mathbf{X}]\ \mathrm{d}^n\mathbf{t}.$$

(*e*) *Asymptotic expansions and limit theorems*

Consider an n-dimensional random vector $\mathbf{X}$ of the form

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_N,$$

where the N summands are independent n-dimensional random vectors identically distributed with probability density P. Then the distribution $\mathscr{P}$ of $\mathbf{X}$ may be written in closed form as a Fourier transform:

$$\begin{aligned}\mathscr{P}(\mathbf{X}) &= \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} M^N(i\mathbf{t})\exp(-i\mathbf{t}\cdot\mathbf{X})\ \mathrm{d}^n\mathbf{t}\\ &= \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \exp[N\log M(i\mathbf{t}) - i\mathbf{t}\cdot\mathbf{X}]\ \mathrm{d}^n\mathbf{t},\end{aligned}$$

where

$$M(\mathbf{t}) = \int_{\mathbb{R}^n} P(\mathbf{Y})\exp(\mathbf{t}\cdot\mathbf{Y})\ \mathrm{d}^n\mathbf{Y}$$

is the moment-generating function common to all the summands.

This an exact expression for $\mathscr{P}$, which may be exploited analytically or numerically in certain favourable cases. Supposing for instance that P has compact support, then its characteristic function $M(i\mathbf{t})$ can be sampled finely enough to accommodate the bandwidth of the support of $\mathscr{P} = P^{*N}$ (this sampling rate clearly depends on n) so that the above expression for $\mathscr{P}$ can be used for its numerical evaluation as the *discrete* Fourier transform of $M^N(i\mathbf{t})$. This exact method is practical only for small values of the dimension n.

In all other cases some form of approximation must be used in the Fourier inversion of $M^N(i\mathbf{t})$. For this purpose it is customary (Cramér, 1946) to expand the cumulant-generating function around $\mathbf{t} = \mathbf{0}$ with respect to the carrying variables $\mathbf{t}$:

$$\log[M^N(i\mathbf{t})] = \sum_{\mathbf{r}\in\mathbb{N}^n}\frac{N\kappa_{\mathbf{r}}}{\mathbf{r}!}(i\mathbf{t})^{\mathbf{r}},$$

where $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ is a multi-index (Section 1.3.2.2.3). The first-order terms may be eliminated by recentring $\mathscr{P}$ around its vector of first-order cumulants

$$\langle\mathbf{X}\rangle = \sum_{j=1}^{N}\langle\mathbf{X}_j\rangle,$$

where $\langle\cdot\rangle$ denotes the mathematical expectation of a random vector. The second-order terms may be grouped separately from the terms of third or higher order to give

$$\begin{aligned}M^N(i\mathbf{t}) &= \exp(-\tfrac{1}{2}N\mathbf{t}^U\mathbf{Q}\mathbf{t})\\ &\quad\times\exp\left\{\sum_{|\mathbf{r}|\geq 3}\frac{N\kappa_{\mathbf{r}}}{\mathbf{r}!}(i\mathbf{t})^{\mathbf{r}}\right\},\end{aligned}$$

where $\mathbf{Q} = \nabla\nabla^T(\log M)$ is the covariance matrix of the multivariate distribution P. Expanding the exponential gives rise to a series of terms of the form

$$\exp(-\tfrac{1}{2}N\mathbf{t}^T\mathbf{Q}\mathbf{t})\times\text{monomial in } t_1, t_2, \ldots, t_n,$$

each of which may now be subjected to a Fourier transformation to yield a Hermite function of $\mathbf{t}$ (Section 1.3.2.4.4.2) with coefficients involving the cumulants κ of P. Taking the transformed terms in natural order gives an asymptotic expansion of P for large N called the *Gram–Charlier series* of $\mathscr{P}$, while grouping the terms according to increasing powers of $1/\sqrt{N}$ gives another asymptotic expansion called the *Edgeworth series* of $\mathscr{P}$. Both expansions comprise a leading Gaussian term which embodies the *central-limit theorem*:

$$\mathscr{P}(\mathbf{E}) = \frac{1}{\sqrt{\det\,(2\pi\mathbf{Q})}}\exp(-\tfrac{1}{2}\mathbf{E}^T\mathbf{Q}^{-1}\mathbf{E}),\quad\text{where } \mathbf{E} = \frac{\mathbf{X}-\langle\mathbf{X}\rangle}{\sqrt{N}}.$$

(*f*) *The saddlepoint approximation*

A limitation of the Edgeworth series is that it gives an accurate estimate of $\mathscr{P}(\mathbf{X})$ only in the vicinity of $\mathbf{X} = \langle\mathbf{X}\rangle$, *i.e.* for small values of $\mathbf{E}$. These convergence difficulties are easily understood: one is substituting a *local* approximation to $\log M$ (*viz* a Taylor-series expansion valid near $\mathbf{t} = \mathbf{0}$) into an integral, whereas integration is a *global* process which consults values of $\log M$ far from $\mathbf{t} = \mathbf{0}$.

It is possible, however, to let the point $\mathbf{t}$ where $\log M$ is expanded as a Taylor series depend on the particular value $\mathbf{X}^*$ of $\mathbf{X}$ for which an accurate evaluation of $\mathscr{P}(\mathbf{X})$ is desired. This is the essence of the *saddlepoint method* (Fowler, 1936; Khinchin 1949; Daniels, 1954; de Bruijn, 1970; Bleistein & Handelsman, 1986), which uses an analytical continuation of $M(\mathbf{t})$ from a function over $\mathbb{R}^n$ to a function over $\mathbb{C}^n$ (see Section 1.3.2.4.2.10). Putting then $\mathbf{t} = \mathbf{s} - i\tau$, the $\mathbb{C}^n$ version of Cauchy's theorem (Hörmander, 1973) gives rise to the identity

$$\mathscr{P}(\mathbf{X}^*) = \frac{\exp(-\boldsymbol{\tau}\cdot\mathbf{X}^*)}{(2\pi)^n} \times \int_{\mathbb{R}^n} \exp\left\{N\left[\log M(\boldsymbol{\tau}+i\mathbf{s}) - i\mathbf{s}\cdot\frac{\mathbf{X}^*}{N}\right]\right\} \mathrm{d}^n\mathbf{s}$$

for *any* $\boldsymbol{\tau}\in\mathbb{R}^n$. By a convexity argument involving the positive-definiteness of covariance matrix $\mathbf{Q}$, there is a unique value of $\boldsymbol{\tau}$ such that

$$\nabla(\log M)|_{\mathbf{t}=\mathbf{0}-i\boldsymbol{\tau}} = \frac{\mathbf{X}^*}{N}.$$

At the *saddlepoint* $\mathbf{t}^* = \mathbf{0} - i\boldsymbol{\tau}$, the modulus of the integrand above is a maximum and its phase is stationary with respect to the integration variable $\mathbf{s}$: as N tends to infinity, all contributions to the integral cancel because of rapid oscillation, except those coming from the immediate vicinity of $\mathbf{t}^*$ where there is no oscillation. A Taylor expansion of $\log M^N$ to second order with respect to $\mathbf{s}$ at $\mathbf{t}^*$ then gives

$$\log M^N(\boldsymbol{\tau}+i\mathbf{s}) \approx \log M^N(\boldsymbol{\tau}) + i\mathbf{s}\cdot\mathbf{X}^* - \frac{N}{2}[\mathbf{s}^T\mathbf{Q}\mathbf{s}]$$

and hence

$$\mathscr{P}(\mathbf{X}^*) \approx \exp[\log M^N(\boldsymbol{\tau}) - \boldsymbol{\tau}\cdot\mathbf{X}^*]\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\exp(-\tfrac{1}{2}\mathbf{s}^T\mathscr{Q}\mathbf{s})\ \mathrm{d}^n\mathbf{s}.$$

The last integral is elementary and gives the 'saddlepoint approximation':

$$\mathscr{P}^{\mathrm{SP}}(\mathbf{X}^*) = \frac{\exp(\mathsf{S})}{\sqrt{\det\,(2\pi\mathscr{Q})}},$$

where

$$\mathsf{S} = \log M^N(\boldsymbol{\tau}) - \boldsymbol{\tau}\cdot\mathbf{X}^*$$

and where

$$\mathscr{Q} = \nabla\nabla^T(\log M^N) = N\mathbf{Q}.$$

This approximation scheme amounts to using the 'conjugate distribution' (Khinchin, 1949)

$$P_{\boldsymbol{\tau}}(\mathbf{X}_j) = P(\mathbf{X}_j)\frac{\exp(\boldsymbol{\tau}\cdot\mathbf{X}_j)}{M(\boldsymbol{\tau})}$$

instead of the original distribution $\boldsymbol{P}(\mathbf{X}_j) = \boldsymbol{P}_{\mathbf{0}}(\mathbf{X}_j)$ for the common distribution of all $\boldsymbol{N}$ random vectors $\mathbf{X}_j$. The exponential modulation results from the analytic continuation of the characteristic (or moment-generating) function into $\mathbb{C}^n$, as in Section 1.3.2.4.2.10. The saddlepoint approximation $\mathscr{P}^{\mathrm{SP}}$ is only the leading term of an asymptotic expansion (called the *saddlepoint expansion*) for $\mathscr{P}$, which is actually the Edgeworth expansion associated with $P^{*N}_{\boldsymbol{\tau}}$.

1.3.4.5.2.2. *The statistical theory of phase determination*

The methods of probability theory just surveyed were applied to various problems formally similar to the crystallographic phase problem [*e.g.* the 'problem of the random walk' of Pearson (1905)] by Rayleigh (1880, 1899, 1905, 1918, 1919) and Kluyver (1906). They became the basis of the statistical theory of communication with the classic papers of Rice (1944, 1945).

The Gram–Charlier and Edgeworth series were introduced into crystallography by Bertaut (1955*a*,*b*,*c*, 1956*a*) and by Klug (1958), respectively, who showed them to constitute the mathematical basis of numerous formulae derived by Hauptman & Karle (1953). The saddlepoint approximation was introduced by Bricogne (1984) and was shown to be related to variational methods involving the maximization of certain entropy criteria. This connection exhibits most of the properties of the Fourier transform at play simultaneously, and will now be described as a final illustration.

(*a*) *Definitions and conventions*

Let H be a set of unique non-origin reflections $\mathbf{h}$ for a crystal with lattice Λ and space group G. Let H contain n_a acentric and n_c centric reflections. Structure-factor values attached to all reflections in H will comprise $n = 2n_a + n_c$ real numbers. For $\mathbf{h}$ acentric, $\alpha_{\mathbf{h}}$ and $\beta_{\mathbf{h}}$ will be the real and imaginary parts of the complex structure factor; for $\mathbf{h}$ centric, $\gamma_{\mathbf{h}}$ will be the real coordinate of the (possibly complex) structure factor measured along a real axis rotated by one of the two angles $\theta_{\mathbf{h}}$, π apart, to which the phase is restricted modulo 2π (Section 1.3.4.2.2.5). These n real coordinates will be arranged as a column vector containing the acentric then the centric data, *i.e.* in the order

$$\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_{n_a}, \beta_{n_a}, \gamma_1, \gamma_2, \ldots, \gamma_{n_c}.$$

(*b*) *Vectors of trigonometric structure-factor expressions*

Let $\boldsymbol{\xi}(\mathbf{x})$ denote the vector of trigonometric structure-factor expressions associated with $\mathbf{x}\in D$, where D denotes the asymmetric unit. These are defined as follows:

$$\begin{aligned}\alpha_{\mathbf{h}}(\mathbf{x}) + i\beta_{\mathbf{h}}(\mathbf{x}) &= \Xi(\mathbf{h},\mathbf{x}) \quad \text{for } \mathbf{h} \text{ acentric}\\ \gamma_{\mathbf{h}}(\mathbf{x}) &= \exp(-i\theta_{\mathbf{h}})\Xi(\mathbf{h},\mathbf{x}) \quad \text{for } \mathbf{h} \text{ centric,}\end{aligned}$$

where

$$\Xi(\mathbf{h},\mathbf{x}) = \frac{1}{|G_{\mathbf{x}}|}\sum_{g\in G}\exp\{2\pi i\mathbf{h}\cdot[S_g(\mathbf{x})]\}.$$

According to the convention above, the coordinates of $\boldsymbol{\xi}(\mathbf{x})$ in $\mathbb{R}^n$ will be arranged in a column vector as follows:

$$\begin{aligned}\boldsymbol{\xi}_{2r-1}(\mathbf{x}) &= \alpha_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = 1, \ldots, n_a,\\ \boldsymbol{\xi}_{2r}(\mathbf{x}) &= \beta_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = 1, \ldots, n_a,\\ \boldsymbol{\xi}_{n_a+r}(\mathbf{x}) &= \gamma_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = n_a+1, \ldots, n_a+n_c.\end{aligned}$$

(*c*) *Distributions of random atoms and moment-generating functions*

Let position $\mathbf{x}$ in D now become a random vector with probability density $m(\mathbf{x})$. Then $\boldsymbol{\xi}(\mathbf{x})$ becomes itself a random vector in $\mathbb{R}^n$, whose distribution $p(\boldsymbol{\xi})$ is the image of distribution $m(\mathbf{x})$ through the mapping $\mathbf{x}\rightarrow\boldsymbol{\xi}(\mathbf{x})$ just defined. The locus of $\boldsymbol{\xi}(\mathbf{x})$ in $\mathbb{R}^n$ is a compact algebraic manifold $\mathscr{L}$ (the multidimensional analogue of a Lissajous curve), so that p is a singular measure (a distribution of order 0, Section 1.3.2.3.4, concentrated on that manifold) with compact support. The average with respect to p of any function Ω over $\mathbb{R}^n$ which is infinitely differentiable in a neighbourhood of $\mathscr{L}$ may be calculated as an average with respect to m over D by the 'induction formula':

$$\langle p,\Omega\rangle = \int_D m(\mathbf{x})\Omega[\boldsymbol{\xi}(\mathbf{x})]\ \mathrm{d}^3\mathbf{x}.$$

In particular, one can calculate the moment-generating function M for distribution p as

$$M(\mathbf{t}) \equiv \langle p_{\boldsymbol{\xi}},\exp(\mathbf{t}\cdot\boldsymbol{\xi})\rangle = \int_D m(\mathbf{x})\exp[\mathbf{t}\cdot\boldsymbol{\xi}(\mathbf{x})]\ \mathrm{d}^3\mathbf{x}$$

and hence calculate the moments μ (respectively cumulants κ) of p by differentiation of M (respectively $\log M$) at $\mathbf{t} = \mathbf{0}$:

$$\mu_{r_1 r_2 \ldots r_n} \equiv \int_D m(\mathbf{x}) \boldsymbol{\xi}_1^{r_1}(\mathbf{x}) \boldsymbol{\xi}_2^{r_2}(\mathbf{x}) \ldots \boldsymbol{\xi}_n^{r_n}(\mathbf{x}) \, \mathrm{d}^3\mathbf{x}$$
$$= \frac{\partial^{r_1+\ldots+r_n}(M)}{\partial t_1^{r_1} \ldots \partial t_n^{r_n}}$$
$$\kappa_{r_1 r_2 \ldots r_n} = \frac{\partial^{r_1+\ldots+r_n}(\log M)}{\partial t_1^{r_1} \ldots \partial t_n^{r_n}}.$$

The structure-factor algebra for group G (Section 1.3.4.2.2.9) then allows one to express products of $\boldsymbol{\xi}$'s as linear combinations of other $\boldsymbol{\xi}$'s, and hence to express all moments and cumulants of distribution $p(\boldsymbol{\xi})$ as linear combinations of real and imaginary parts of Fourier coefficients of the prior distribution of atoms $m(\mathbf{x})$. This plays a key role in the use of non-uniform distributions of atoms.

(*d*) *The joint probability distribution of structure factors*

In the random-atom model of an equal-atom structure, N atoms are placed randomly, independently of each other, in the asymmetric unit D of the crystal with probability density $m(\mathbf{x})$. For point atoms of unit weight, the vector $\mathbf{F}$ of structure-factor values for reflections $\mathbf{h} \in H$ may be written

$$\mathbf{F} = \sum_{I=1}^{N} \boldsymbol{\xi}^{[I]},$$

where the N copies $\boldsymbol{\xi}^{[I]}$ of random vector $\boldsymbol{\xi}$ are independent and have the same distribution $p(\boldsymbol{\xi})$.

The joint probability distribution $\mathscr{P}(\mathbf{F})$ is then [Section 1.3.4.5.2.1(*e*)]

$$\mathscr{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp[N \log M(i\mathbf{t}) - i\mathbf{t} \cdot \mathbf{X}] \, \mathrm{d}^n\mathbf{t}.$$

For low dimensionality n it is possible to carry out the Fourier transformation numerically after discretization, provided $M(i\mathbf{t})$ is sampled sufficiently finely that no aliasing results from taking its Nth power (Barakat, 1974). This exact approach can also accommodate heterogeneity, and has been used first in the field of intensity statistics (Shmueli *et al.*, 1984, 1985; Shmueli & Weiss, 1987, 1988), then in the study of the Σ_1 and Σ_2 relations in triclinic space groups (Shmueli & Weiss, 1985, 1986). Some of these applications are described in Chapter 2.1 of this volume. This method could be extended to the construction of any joint probability distribution (j.p.d.) in any space group by using the generic expression for the moment-generating function (m.g.f.) derived by Bricogne (1984). It is, however, limited to small values of n by the necessity to carry out n-dimensional FFTs on large arrays of sample values.

The asymptotic expansions of Gram–Charlier and Edgeworth have good convergence properties only if $F_{\mathbf{h}}$ lies in the vicinity of $\langle F_{\mathbf{h}} \rangle = N\widetilde{\mathscr{F}}[m](\mathbf{h})$ for all $\mathbf{h} \in H$. Previous work on the j.p.d. of structure factors has used for $m(\mathbf{x})$ a uniform distribution, so that $\langle \mathbf{F} \rangle = \mathbf{0}$; as a result, the corresponding expansions are accurate only if all moduli $|F_{\mathbf{h}}|$ are small, in which case the j.p.d. contains little phase information.

The saddlepoint method [Section 1.3.4.5.2.1(*f*)] constitutes the method of choice for evaluating the joint probability $\mathscr{P}(\mathbf{F}^*)$ of structure factors when some of the moduli in $\mathbf{F}^*$ are large. As shown previously, this approximation amounts to using the 'conjugate distribution'

$$p_{\boldsymbol{\tau}}(\boldsymbol{\xi}) = p(\boldsymbol{\xi}) \frac{\exp(\boldsymbol{\tau} \cdot \boldsymbol{\xi})}{M(\boldsymbol{\tau})}$$

instead of the original distribution $p(\boldsymbol{\xi}) = p_{\mathbf{0}}(\boldsymbol{\xi})$ for the distribution of random vector $\boldsymbol{\xi}$. This conjugate distribution $p_{\boldsymbol{\tau}}$ is induced from the modified distribution of atoms

$$q_{\boldsymbol{\tau}}(\mathbf{x}) = m(\mathbf{x}) \frac{\exp[\boldsymbol{\tau} \cdot \boldsymbol{\xi}(\mathbf{x})]}{M(\boldsymbol{\tau})}, \tag{SP1}$$

where, by the induction formula, $M(\boldsymbol{\tau})$ may be written as

$$M(\boldsymbol{\tau}) = \int_D m(\mathbf{x}) \exp[\boldsymbol{\tau} \cdot \boldsymbol{\xi}(\mathbf{x})] \, \mathrm{d}^3\mathbf{x} \tag{SP2}$$

and where $\boldsymbol{\tau}$ is the unique solution of the saddlepoint equation:

$$\nabla_{\boldsymbol{\tau}}(\log M^N) = \mathbf{F}^*. \tag{SP3}$$

The desired approximation is then

$$\mathscr{P}^{\mathrm{SP}}(\mathbf{F}^*) = \frac{\exp(\mathsf{S})}{\sqrt{\det\,(2\pi\mathscr{Q})}},$$

where

$$\mathsf{S} = \log M^N(\boldsymbol{\tau}) - \boldsymbol{\tau} \cdot \mathbf{F}^*$$

and where

$$\mathscr{Q} = \nabla\nabla^T(\log M^N) = \mathbf{NQ}.$$

Finally, the elements of the Hessian matrix $\mathbf{Q} = \nabla\nabla^T(\log M)$ are just the trigonometric second-order cumulants of distribution p, and hence can be calculated *via* structure-factor algebra from the Fourier coefficients of $q_{\boldsymbol{\tau}}(\mathbf{x})$. All the quantities involved in the expression for $\mathscr{P}^{\mathrm{SP}}(\mathbf{F}^*)$ are therefore effectively computable from the initial data $m(\mathbf{x})$ and $\mathbf{F}^*$.

(*e*) *Maximum-entropy distributions of atoms*

One of the main results in Bricogne (1984) is that the modified distribution $q_{\boldsymbol{\tau}}(\mathbf{x})$ in (SP1) is the unique distribution which has maximum entropy $\mathscr{S}_m(q)$ relative to $m(\mathbf{x})$, where

$$\mathscr{S}_m(q) = -\int_D q(\mathbf{x}) \log\left[\frac{q(\mathbf{x})}{m(\mathbf{x})}\right] \mathrm{d}^3\mathbf{x},$$

under the constraint that $\mathbf{F}^*$ be the centroid vector of the corresponding conjugate distribution $\mathscr{P}_{\boldsymbol{\tau}}(\mathbf{F})$. The traditional notation of maximum-entropy (ME) theory (Jaynes, 1957, 1968, 1983) is in this case (Bricogne, 1984)

$$q^{\mathrm{ME}}(\mathbf{x}) = m(\mathbf{x}) \frac{\exp[\boldsymbol{\lambda} \cdot \boldsymbol{\xi}(\mathbf{x})]}{Z(\boldsymbol{\lambda})} \tag{ME1}$$

$$Z(\boldsymbol{\lambda}) = \int_D m(\mathbf{x}) \exp[\boldsymbol{\lambda} \cdot \boldsymbol{\xi}(\mathbf{x})] \, \mathrm{d}^3\mathbf{x} \tag{ME2}$$

$$\nabla_{\lambda}(\log Z^N) = \mathbf{F}^* \tag{ME3}$$

so that Z is identical to the m.g.f. M, and the coordinates $\boldsymbol{\tau}$ of the saddlepoint are the Lagrange multipliers $\boldsymbol{\lambda}$ for the constraints $\mathbf{F}^*$.

Jaynes's ME theory also gives an estimate for $\mathscr{P}(\mathbf{F}^*)$:

$$\mathscr{P}^{\mathrm{ME}}(\mathbf{F}^*) \approx \exp(\mathscr{S}),$$

where

$$\mathscr{S} = \log Z^N - \boldsymbol{\lambda} \cdot \mathbf{F}^* = N\mathscr{S}_m(\mathbf{q}^{\mathrm{ME}})$$

is the total entropy and is the counterpart to S under the equivalence just established.

$\mathscr{P}^{\mathrm{ME}}$ is identical to $\mathscr{P}^{\mathrm{SP}}$, but lacks the denominator. The latter, which is the normalization factor of a multivariate Gaussian with covariance matrix $\mathscr{Q}$, may easily be seen to arise through Szegö's theorem (Sections 1.3.2.6.9.4, 1.3.4.2.1.10) from the extra logarithmic term in Stirling's formula

$$\log(q!) \approx q \log q - q + \tfrac{1}{2}\log(2\pi q)$$

(see, for instance, Reif, 1965) beyond the first two terms which serve to define entropy, since

$$\frac{1}{n}\log\det\,(2\pi\mathbf{Q}) \approx \int\limits_{\mathbb{R}^3/\mathbb{Z}^3} \log 2\pi q^{\rm ME}(\mathbf{x})\;{\rm d}^3\mathbf{x}.$$

The relative effect of this extra normalization factor depends on the ratio

$$\frac{n}{N} = \frac{\text{dimension of } \mathbf{F} \text{ over } \mathbb{R}}{\text{number of atoms}}.$$

The above relation between entropy maximization and the saddlepoint approximation is the basis of a Bayesian statistical approach to the phase problem (Bricogne, 1988) where the assumptions under which joint distributions of structure factors are sought incorporate many new ingredients (such as molecular boundaries, isomorphous substitutions, known fragments, noncrystallographic symmetries, multiple crystal forms) besides trial phase choices for basis reflections. The ME criterion intervenes in the construction of $q^{\rm ME}(\mathbf{x})$ under these assumptions, and the distribution $q^{\rm ME}(\mathbf{x})$ is a very useful computational intermediate in obtaining the approximate joint probability $\mathscr{P}^{\rm SP}(\mathbf{F}^*)$ and the associated conditional distributions and likelihood functions.

(*f*) *Role of the Fourier transformation*

The formal developments presented above make use of the following properties of the Fourier transformation:

(i) the convolution theorem, which turns the convolution of probability distributions into the multiplication of their characteristic functions;

(ii) the differentiation property, which confers moment-generating properties to characteristic functions;

(iii) the reciprocity theorem, which allows the retrieval of a probability distribution from its characteristic or moment-generating function;

(iv) the Paley–Wiener theorem, which allows the analytic continuation of characteristic functions associated to probability distributions with compact support, and thus gives rise to conjugate families of distributions;

(v) Bertaut's structure-factor algebra (a discrete symmetrized version of the convolution theorem), which allows the calculation of all necessary moments and cumulants when the dimension n is small;

(vi) Szegö's theorem, which provides an asymptotic approximation of the normalization factor when n is large.

This multi-faceted application seems an appropriate point at which to end this description of the Fourier transformation and of its use in crystallography.

Acknowledgements

Many aspects of the theory of discrete Fourier transform algorithms and of its extension to incorporate crystallographic symmetry have been the focus of a long-standing collaborative effort between Professor Louis Auslander, Professor Richard Tolimieri, their coworkers and the writer. I am most grateful to them for many years of mathematical stimulation and enjoyment, for introducing me to the 'big picture' of the discrete Fourier transform which they have elaborated over the past decade, and for letting me describe here some of their unpublished work. In particular, the crystallographic extensions of the Rader/Winograd algorithms presented in Section 1.3.4.3.4.3 were obtained by Richard Tolimieri, in a collaboration partially supported by NIH grant GM 32362 (to the writer).

I am indebted to the Editor for many useful and constructive suggestions of possible improvements to the text, only a few of which I have been able to implement. I hope to incorporate many more of them in the future. I also wish to thank Dr D. Sayre for many useful comments on an early draft of the manuscript.

This contribution was written during the tenure of a Visiting Fellowship at Trinity College, Cambridge, with partial financial support from Trinity College and the MRC Laboratory of Molecular Biology. I am most grateful to both institutions for providing ideal working conditions.

1.2 (*cont.*)

Materlik, G., Sparks, C. J. & Fischer, K. (1994). *Resonant anomalous X-ray scattering. Theory and applications.* Amsterdam: North-Holland.

Paturle, A. & Coppens, P. (1988). *Normalization factors for spherical harmonic density functions. Acta Cryst.* A**44**, 6–7.

Ruedenberg, K. (1962). *The nature of the chemical bond. Phys. Rev.* **34**, 326–376.

Scheringer, C. (1985*a*). *A general expression for the anharmonic temperature factor in the isolated-atom-potential approach. Acta Cryst.* A**41**, 73–79.

Scheringer, C. (1985*b*). *A deficiency of the cumulant expansion of the anharmonic temperature factor. Acta Cryst.* A**41**, 79–81.

Schomaker, V. & Trueblood, K. N. (1968). *On the rigid-body motion of molecules in crystals. Acta Cryst.* B**24**, 63–76.

Schwarzenbach, D. (1986). Private communication.

Squires, G. L. (1978). *Introduction to the theory of thermal neutron scattering*. Cambridge University Press.

Stewart, R. F. (1969*a*). *Generalized X-ray scattering factors. J. Chem. Phys.* **51**, 4569–4577.

Stewart, R. F. (1969*b*). *Small Gaussian expansions of atomic orbitals. J. Chem. Phys.* **50**, 2485–2495.

Stewart, R. F. (1970). *Small Gaussian expansions of Slater-type orbitals. J. Chem. Phys.* **52**, 431–438.

Stewart, R. F. (1980). *Electron and magnetization densities in molecules and solids*, edited by P. J. Becker, pp. 439–442. New York: Plenum.

Stewart, R. F. & Hehre, W. J. (1970). *Small Gaussian expansions of atomic orbitals: second-row atoms. J. Chem. Phys.* **52**, 5243–5247.

Su, Z. & Coppens, P. (1990). *Closed-form expressions for Fourier–Bessel transforms of Slater-type functions. J. Appl. Cryst.* **23**, 71–73.

Su, Z. & Coppens, P. (1994). *Normalization factors for Kubic harmonic density functions. Acta Cryst.* A**50**, 408–409.

Tanaka, K. & Marumo, F. (1983). *Willis formalism of anharmonic temperature factors for a general potential and its application in the least-squares method. Acta Cryst.* A**39**, 631–641.

Von der Lage, F. C. & Bethe, H. A. (1947). *A method for obtaining electronic functions and eigenvalues in solids with an application to sodium. Phys. Rev.* **71**, 612–622.

Waller, I. & Hartree, D. R. (1929). *Intensity of total scattering X-rays. Proc. R. Soc. London Ser. A*, **124**, 119–142.

Weiss, R. J. & Freeman, A. J. (1959). *X-ray and neutron scattering for electrons in a crystalline field and the determination of outer electron configurations in iron and nickel. J. Phys. Chem. Solids*, **10**, 147–161.

Willis, B. T. M. (1969). *Lattice vibrations and the accurate determination of structure factors for the elastic scattering of X-rays and neutrons. Acta Cryst.* A**25**, 277–300.

Zucker, U. H. & Schulz, H. (1982). *Statistical approaches for the treatment of anharmonic motion in crystals. I. A comparison of the most frequently used formalisms of anharmonic thermal vibrations. Acta Cryst.* A**38**, 563–568.

1.3

Agarwal, R. C. (1978). *A new least-squares refinement technique based on the fast Fourier transform algorithm. Acta Cryst.* A**34**, 791–809.

Agarwal, R. C. (1980). *The refinement of crystal structure by fast Fourier least-squares*. In *Computing in crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 18.01–18.13. Bangalore: The Indian Academy of Science.

Agarwal, R. C. (1981). *New results on fast Fourier least-squares refinement technique*. In *Refinement of protein structures*, compiled by P. A. Machin, J. W. Campbell & M. Elder (ref. DL/SCI/R16), pp. 24–28. Warrington: SERC Daresbury Laboratory.

Agarwal, R. C. & Cooley, J. W. (1986). *Fourier transform and convolution subroutines for the IBM 3090 Vector facility. IBM J. Res. Dev.* **30**, 145–162.

Agarwal, R. C. & Cooley, J. W. (1987). *Vectorized mixed radix discrete Fourier transform algorithms. Proc. IEEE*, **75**, 1283–1292.

Agarwal, R. C., Lifchitz, A. & Dodson, E. J. (1981). Appendix (pp. 36–38) to Dodson (1981).

Ahlfors, L. V. (1966). *Complex analysis*. New York: McGraw-Hill.

Ahmed, F. R. & Barnes, W. H. (1958). *Generalized programmes for crystallographic computations. Acta Cryst.* **11**, 669–671.

Ahmed, F. R. & Cruickshank, D. W. J. (1953*a*). *A refinement of the crystal structure analysis of oxalic acid dihydrate. Acta Cryst.* **6**, 385–392.

Ahmed, F. R. & Cruickshank, D. W. J. (1953*b*). *Crystallographic calculations on the Manchester University electronic digital computer (Mark II). Acta Cryst.* **6**, 765–769.

Akhiezer, N. I. (1965). *The classical moment problem*. Edinburgh and London: Oliver & Boyd.

Alston, N. A. & West, J. (1929). *The structure of topaz, $[Al(F,OH)]_2SiO_4$. Z. Kristallogr.* **69**, 149–167.

Apostol, T. M. (1976). *Introduction to analytic number theory.* New York: Springer-Verlag.

Arnold, H. (1995). *Transformations in crystallography*. In *International tables for crystallography*, Vol. A. *Space-group symmetry*, edited by Th. Hahn, pp. 69–79. Dordrecht: Kluwer Academic Publishers.

Artin, E. (1944). *Galois theory*. Notre Dame University Press.

Ascher, E. & Janner, A. (1965). *Algebraic aspects of crystallography. I. Space groups as extensions. Helv. Phys. Acta*, **38**, 551–572.

Ascher, E. & Janner, A. (1968). *Algebraic aspects of crystallography. II. Non-primitive translations in space groups. Commun. Math. Phys*. **11**, 138–167.

Ash, J. M. (1976). *Multiple trigonometric series*. In *Studies in harmonic analysis*, edited by J. M. Ash, pp. 76–96. *MAA studies in mathematics*, Vol. 13. The Mathematical Association of America.

Auslander, L. (1965). *An account of the theory of crystallographic groups. Proc. Am. Math. Soc*. **16**, 1230–1236.

Auslander, L., Feig, E. & Winograd, S. (1982). *New algorithms for evaluating the 3-dimensional discrete Fourier transform*. In *Computational crystallography*, edited by D. Sayre, pp. 451–461. New York: Oxford University Press.

Auslander, L., Feig, E. & Winograd, S. (1983). *New algorithms for the multidimensional discrete Fourier transform. IEEE Trans. Acoust. Speech Signal Process*. **31**, 388–403.

Auslander, L., Feig, E. & Winograd, S. (1984). *Abelian semi-simple algebras and algorithms for the discrete Fourier transform. Adv. Appl. Math*. **5**, 31–55.

Auslander, L., Johnson, R. W. & Vulis, M. (1988). *Evaluating finite Fourier transforms that respect group symmetries. Acta Cryst.* A**44**, 467–478.

Auslander, L. & Shenefelt, M. (1987). *Fourier transforms that respect crystallographic symmetries. IBM J. Res. Dev*. **31**, 213–223.

Auslander, L. & Tolimieri, R. (1979). *Is computing with the finite Fourier transform pure or applied mathematics? Bull. Am. Math. Soc*. **1**, 847–897.

Auslander, L. & Tolimieri, R. (1985). *Ring structure and the Fourier transform. Math. Intelligencer*, **7**, 49–54.

Auslander, L., Tolimieri, R. & Winograd, S. (1982). *Hecke's theorem in quadratic reciprocity, finite nilpotent groups and the Cooley–Tukey algorithm. Adv. Math.* **43**, 122–172.

Ayoub, R. (1963). *An introduction to the analytic theory of numbers*. Providence, RI: American Mathematical Society.

Baker, E. N. & Dodson, E. J. (1980). *Crystallographic refinement of the structure of actinidin at 1.7 Å resolution by fast Fourier least-squares methods. Acta Cryst.* A**36**, 559–572.

Bantz, D. & Zwick, M. (1974). *The use of symmetry with the fast Fourier algorithm. Acta Cryst.* A**30**, 257–260.

Barakat, R. (1974). *First-order statistics of combined random sinusoidal waves with applications to laser speckle patterns. Opt. Acta*, **21**, 903–921.

1.3 (*cont.*)

Barrett, A. N. & Zwick, M. (1971). *A method for the extension and refinement of crystallographic protein phases utilizing the fast Fourier transform. Acta Cryst.* A**27**, 6–11.

Beevers, C. A. (1952). *Fourier strips at a 3° interval. Acta Cryst.* **5**, 670–673.

Beevers, C. A. & Lipson, H. (1934). *A rapid method for the summation of a two-dimensional Fourier series. Philos. Mag*. **17**, 855–859.

Beevers, C. A. & Lipson, H. (1936). *A numerical method for two-dimensional Fourier synthesis. Nature (London)*, **137**, 825–826.

Beevers, C. A. & Lipson, H. (1952). *The use of Fourier strips for calculating structure factors. Acta Cryst.* **5**, 673–675.

Bellman, R. (1958). *Dynamic programming and stochastic control processes. Inf. Control*, **1**, 228–239.

Bennett, J. M. & Kendrew, J. C. (1952). *The computation of Fourier syntheses with a digital electronic calculating machine. Acta Cryst.* **5**, 109–116.

Berberian, S. K. (1962). *Measure and integration*. New York: Macmillan. [Reprinted by Chelsea, New York, 1965.]

Bertaut, E. F. (1952). *L'énergie électrostatique de réseaux ioniques. J. Phys. Radium*, **13**, 499–505.

Bertaut, E. F. (1955*a*). *La méthode statistique en cristallographie. I. Acta Cryst*. **8**, 537–543.

Bertaut, E. F. (1955*b*). *La méthode statistique en cristallographie. II. Quelques applications. Acta Cryst.* **8**, 544–548.

Bertaut, E. F. (1955*c*). *Fonction de répartition: application à l'approache directe des structures. Acta Cryst.* **8**, 823–832.

Bertaut, E. F. (1956*a*). *Les groupes de translation non primitifs et la méthode statistique. Acta Cryst*. **9**, 322.

Bertaut, E. F. (1956*b*). *Tables de linéarisation des produits et puissances des facteurs de structure. Acta Cryst.* **9**, 322–323.

Bertaut, E. F. (1956*c*). *Algèbre des facteurs de structure. Acta Cryst.* **9**, 769–770.

Bertaut, E. F. (1959*a*). *IV. Algèbre des facteurs de structure. Acta Cryst.* **12**, 541–549.

Bertaut, E. F. (1959*b*). *V. Algèbre des facteurs de structure. Acta Cryst.* **12**, 570–574.

Bertaut, E. F. & Waser, J. (1957). *Structure factor algebra. II. Acta Cryst.* **10**, 606–607.

Bhattacharya, R. N. & Rao, R. R. (1976). *Normal approximation and asymptotic expansions*. New York: John Wiley.

Bieberbach, L. (1911). *Über die Bewegungsgruppen der Euklidischen Raume I. Math. Ann.* **70**, 297–336.

Bieberbach, L. (1912). *Über die Bewegungsgruppen der Euklidischen Raume II. Math. Ann.* **72**, 400–412.

Bienenstock, A. & Ewald, P. P. (1962). *Symmetry of Fourier space. Acta Cryst.* **15**, 1253–1261.

Blahut, R. E. (1985). *Fast algorithms for digital signal processing*. Reading: Addison-Wesley.

Bleistein, N. & Handelsman, R. A. (1986). *Asymptotic expansions of integrals*. New York: Dover Publications.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Protein disk of tobacco mosaic virus at 2.8 Ångström resolution showing the interactions within and between subunits. Nature (London)*, **276**, 362–368.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Bochner, S. (1932). *Vorlesungen über Fouriersche Integrale*. Leipzig: Akademische Verlagsgesellschaft.

Bochner, S. (1959). *Lectures on Fourier integrals*. Translated from Bochner (1932) by M. Tenenbaum & H. Pollard. Princeton University Press.

Bode, W. & Schwager, P. (1975). *The refined crystal structure of bovine β-trypsin at 1.8Å resolution. II. Crystallographic refinement, calcium-binding site, benzamidine binding site and active site at pH 7.0. J. Mol. Biol.* **98**, 693–717.

Bondot, P. (1971). *Application de la transformée de Fourier performante aux calculs cristallographiques. Acta Cryst.* A**27**, 492–494.

Booth, A. D. (1945*a*). *Two new modifications of the Fourier method of X-ray structure analysis. Trans. Faraday Soc.* **41**, 434–438.

Booth, A. D. (1945*b*). *An expression for following the process of refinement in X-ray structure analysis using Fourier series. Philos. Mag.* **36**, 609–615.

Booth, A. D. (1945*c*). *Accuracy of atomic co-ordinates derived from Fourier synthesis. Nature (London)*, **156**, 51–52.

Booth, A. D. (1946*a*). *A differential Fourier method for refining atomic parameters in crystal structure analysis. Trans. Faraday Soc.* **42**, 444–448.

Booth, A. D. (1946*b*). *The simultaneous differential refinement of co-ordinates and phase angles in X-ray Fourier synthesis. Trans. Faraday Soc.* **42**, 617–619.

Booth, A. D. (1946*c*). *The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis. Proc. R. Soc. London Ser. A*, **188**, 77–92.

Booth, A. D. (1947*a*). *The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis. III. Proc. R. Soc. London Ser. A*, **190**, 482–489.

Booth, A. D. (1947*b*). *The accuracy of atomic co-ordinates derived from Fourier series in X-ray structure analysis. IV. The two-dimensional projection of oxalic acid. Proc. R. Soc. London Ser. A*, **190**, 490–496.

Booth, A. D. (1947*c*). *A new refinement technique for X-ray structure analysis. J. Chem. Phys.* **15**, 415–416.

Booth, A. D. (1947*d*). *Application of the method of steepest descents to X-ray structure analysis. Nature (London)*, **160**, 196.

Booth, A. D. (1948*a*). *A new Fourier refinement technique. Nature (London)*, **161**, 765–766.

Booth, A. D. (1948*b*). *Fourier technique in X-ray organic structure analysis*. Cambridge University Press.

Booth, A. D. (1949). *The refinement of atomic parameters by the technique known in X-ray crystallography as 'the method of steepest descents'. Proc. R. Soc. London Ser. A*, **197**, 336–355.

Born, M. & Huang, K. (1954). *Dynamical theory of crystal lattices*. Oxford University Press.

Bracewell, R. N. (1986). *The Fourier transform and its applications*, 2nd ed., revised. New York: McGraw-Hill.

Bragg, W. H. (1915). *X-rays and crystal structure*. (Bakerian Lecture.) *Philos. Trans. R. Soc. London Ser. A*, **215**, 253–274.

Bragg, W. L. (1929). *Determination of parameters in crystal structures by means of Fourier series. Proc. R. Soc. London Ser. A*, **123**, 537–559.

Bragg, W. L. (1975). *The development of X-ray analysis*, edited by D. C. Phillips & H. Lipson. London: Bell.

Bragg, W. L. & Lipson, H. (1936). *The employment of contoured graphs of structure-factor in crystal analysis. Z. Kristallogr.* **95**, 323–337.

Bragg, W. L. & West, J. (1929). *A technique for the X-ray examination of crystal structures with many parameters. Z. Kristallogr.* **69**, 118–148.

Bragg, W. L. & West, J. (1930). *A note on the representation of crystal structure by Fourier series. Philos. Mag.* **10**, 823–841.

Bremermann, H. (1965). *Distributions, complex variables, and Fourier transforms*. Reading: Addison-Wesley.

Bricogne, G. (1974). *Geometric sources of redundancy in intensity data and their use for phase determination. Acta Cryst.* A**30**, 395–405.

Bricogne, G. (1976). *Methods and programs for direct-space exploitation of geometric redundancies. Acta Cryst.* A**32**, 832–847.

Bricogne, G. (1982). *Generalised density modification methods*. In *Computational crystallography*, edited by D. Sayre, pp. 258–264. New York: Oxford University Press.

Bricogne, G. (1984). *Maximum entropy and the foundations of direct methods. Acta Cryst.* A**40**, 410–445.

Bricogne, G. (1988). *A Bayesian statistical theory of the phase problem. I. A multichannel maximum entropy formalism for constructing generalised joint probability distributions of structure factors. Acta Cryst.* A**44**, 517–545.

Bricogne, G. & Tolimieri, R. (1990). *Two-dimensional FFT algorithms on data admitting 90°-rotational symmetry*. In *Signal processing theory*, edited by L. Auslander, T. Kailath & S. Mitter, pp. 25–35. New York: Springer-Verlag.

1.3 (*cont.*)

Brigham, E. O. (1988). *The fast Fourier transform and its applications*. Englewood Cliffs: Prentice-Hall.

Brill, R., Grimm, H., Hermann, C. & Peters, C. (1939). *Anwendung der röntgenographischen Fourieranalyse auf Fragen den chemischen Bindung. Ann. Phys. (Leipzig)*, **34**, 393–445.

Britten, P. L. & Collins, D. M. (1982). *Information theory as a basis for the maximum determinant. Acta Cryst.* A**38**, 129–132.

Brown, H. (1969). *An algorithm for the determination of space groups. Math. Comput*. **23**, 499–514.

Brown, H., Bülow, R., Neubüser, J., Wondratschek, H. & Zassenhaus, H. (1978). *Crystallographic groups of four-dimensional space*. New York: John Wiley.

Bruijn, N. G. de (1970). *Asymptotic methods in analysis*, 3rd ed. Amsterdam: North-Holland.

Brünger, A. T. (1988). *Crystallographic refinement by simulated annealing*. In *Crystallographic computing 4: techniques and new technologies*, edited by N. W. Isaacs & M. R. Taylor, pp. 126–140. New York: Oxford University Press.

Brünger, A. T. (1989). *A memory-efficient fast Fourier transformation algorithm for crystallographic refinement on supercomputers. Acta Cryst.* A**45**, 42–50.

Brünger, A. T., Karplus, M. & Petsko, G. A. (1989). *Crystallographic refinement by simulated annealing: application to crambin. Acta Cryst.* A**45**, 50–61.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R factor refinement by molecular dynamics. Science*, **235**, 458–460.

Bryan, R. K., Bansal, M., Folkhard, W., Nave, C. & Marvin, D. A. (1983). *Maximum-entropy calculation of the electron density at 4Å resolution of Pf1 filamentous bacteriophage. Proc. Natl Acad. Sci. USA*, **80**, 4728–4731.

Bryan, R. K. & Skilling, J. (1980). *Deconvolution by maximum entropy, as illustrated by application to the jet of M87. Mon. Not. R. Astron. Soc.* **191**, 69–79.

Burch, S. F., Gull, S. F. & Skilling, J. (1983). *Image restoration by a powerful maximum entropy method. Comput. Vision, Graphics Image Process*. **23**, 113–128.

Burnett, R. M. & Nordman, C. E. (1974). *Optimization of the calculation of structure factors for large molecules. J. Appl. Cryst.* **7**, 625–627.

Burnside, W. (1911). *Theory of groups of finite order*, 2nd ed. Cambridge University Press.

Burrus, C. S. & Eschenbacher, P. W. (1981). *An in-place, in-order prime factor FFT algorithm. IEEE Trans. Acoust. Speech Signal Process*. **29**, 806–817.

Busing, W. R. & Levy, H. A. (1961). *Least squares refinement programs for the IBM 704*. In *Computing methods and the phase problem in X-ray crystal analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 146–149. Oxford: Pergamon Press.

Byerly, W. E. (1893). *An elementary treatise on Fourier's series and spherical, cylindrical and ellipsoidal harmonics*. Boston: Ginn & Co. [Reprinted by Dover Publications, New York, 1959.]

Campbell, G. A. & Foster, R. M. (1948). *Fourier integrals for practical applications*. Princeton: Van Nostrand.

Carathéodory, C. (1911). *Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Functionen. Rend. Circ. Mat. Palermo*, **32**, 193–217.

Carslaw, H. S. (1930). *An introduction to the theory of Fourier's series and integrals*. London: Macmillan. [Reprinted by Dover Publications, New York, 1950.]

Carslaw, H. S. & Jaeger, J. C. (1948). *Operational methods in applied mathematics*. Oxford University Press.

Cartan, H. (1961). *Théorie des fonctions analytiques*. Paris: Hermann.

Challifour, J. L. (1972). *Generalized functions and Fourier analysis*. Reading: Benjamin.

Champeney, D. C. (1973). *Fourier transforms and their physical applications*. New York: Academic Press.

Churchill, R. V. (1958). *Operational mathematics*, 2nd ed. New York: McGraw-Hill.

Cochran, W. (1948*a*). *A critical examination of the Beevers–Lipson method of Fourier series summation. Acta Cryst.* **1**, 54–56.

Cochran, W. (1948*b*). *The Fourier method of crystal structure analysis. Nature (London)*, **161**, 765.

Cochran, W. (1948*c*). *The Fourier method of crystal-structure analysis. Acta Cryst*. **1**, 138–142.

Cochran, W. (1948*d*). *X-ray analysis and the method of steepest descents. Acta Cryst*. **1**, 273.

Cochran, W. (1951*a*). *Some properties of the $F_o - F_c$-synthesis. Acta Cryst*. **4**, 408–411.

Cochran, W. (1951*b*). *The structures of pyrimidines and purines. V. The electron distribution in adenine hydrochloride. Acta Cryst.* **4**, 81–92.

Cochran, W., Crick, F. H. C. & Vand, V. (1952). *The structure of synthetic polypeptides. I. The transform of atoms on a helix. Acta Cryst*. **5**, 581–586.

Cochran, W. T., Cooley, J. W., Favin, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., Maling, G. C., Nelson, D. E., Rader, C. M. & Welch, P. D. (1967). *What is the fast Fourier transform? IEEE Trans. Audio*, **15**, 45–55.

Collins, D. M. (1975). *Efficiency in Fourier phase refinement for protein crystal structures. Acta Cryst*. A**31**, 388–389.

Collins, D. M., Brice, M. D., Lacour, T. F. M. & Legg, M. J. (1976). *Fourier phase refinement and extension by modification of electron-density maps*. In *Crystallographic computing techniques*, edited by F. R. Ahmed, pp. 330–335. Copenhagen: Munksgaard.

Colman, P. M. (1974). *Non-crystallographic symmetry and the sampling theorem. Z. Kristallogr.* **140**, 344–349.

Cooley, J. W. & Tukey, J. W. (1965). *An algorithm for the machine calculation of complex Fourier series. Math. Comput.* **19**, 297–301.

Cox, E. G. & Cruickshank, D. W. J. (1948). *The accuracy of electron-density maps in X-ray structure analysis. Acta Cryst*. **1**, 92–93.

Cox, E. G., Gross, L. & Jeffrey, G. A. (1947). *A Hollerith punched-card method for the evaluation of electron density in crystal structure analysis. Proc. Leeds Philos. Soc.* **5**, 1–13.

Cox, E. G., Gross, L. & Jeffrey, G. A. (1949). *A Hollerith technique for computing three-dimensional differential Fourier syntheses in X-ray crystal structure analysis. Acta Cryst.* **2**, 351–355.

Cox, E. G. & Jeffrey, G. A. (1949). *The use of Hollerith computing equipment in crystal structure analysis. Acta Cryst.* **2**, 341–343.

Coxeter, H. S. M. & Moser, W. O. J. (1972). *Generators and relations for discrete groups*, 3rd ed. Berlin: Springer-Verlag.

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton University Press.

Crowther, R. A. (1967). *A linear analysis of the non-crystallographic symmetry problem. Acta Cryst.* **22**, 758–764.

Crowther, R. A. (1969). *The use of non-crystallographic symmetry for phase determination. Acta Cryst.* B**25**, 2571–2580.

Crowther, R. A. (1972). *The fast rotation function*. In *The molecular replacement method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.

Crowther, R. A. & Blow, D. M. (1967). *A method of positioning a known molecule in an unknown crystal structure. Acta Cryst.* **23**, 544–548.

Cruickshank, D. W. J. (1949*a*). *The accuracy of electron-density maps in X-ray analysis with special reference to dibenzyl. Acta Cryst.* **2**, 65–82.

Cruickshank, D. W. J. (1949*b*). *The accuracy of atomic co-ordinates derived by least-squares or Fourier methods. Acta Cryst.* **2**, 154–157.

Cruickshank, D. W. J. (1950). *The convergence of the least-squares and Fourier refinement methods. Acta Cryst.* **3**, 10–13.

Cruickshank, D. W. J. (1952). *On the relations between Fourier and least-squares methods of structure determination. Acta Cryst.* **5**, 511–518.

Cruickshank, D. W. J. (1956). *The determination of the anisotropic thermal motion of atoms in crystals. Acta Cryst.* **9**, 747–753.

Cruickshank, D. W. J. (1965*a*). *Errors in Fourier series*. In *Computing methods in crystallography*, edited by J. S. Rollett, pp. 107–111. Oxford: Pergamon Press.

1.3 (*cont.*)

Cruickshank, D. W. J. (1965*b*). *Errors in least-squares methods*. In *Computing methods in crystallography*, edited by J. S. Rollett, pp. 112–116. Oxford: Pergamon Press.

Cruickshank, D. W. J. (1970). *Least-squares refinement of atomic parameters*. In *Crystallographic computing*, edited by F. R. Ahmed, pp. 187–197. Copenhagen: Munksgaard.

Cruickshank, D. W. J., Pilling, D. E., Bujosa, A., Lovell, F. M. & Truter, M. R. (1961). *Crystallographic calculations on the Ferranti Pegasus and mark I computers*. In *Computing methods and the phase problem in X-ray crystal analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 32–78. Oxford: Pergamon Press.

Cruickshank, D. W. J. & Rollett, J. S. (1953). *Electron-density errors at special positions. Acta Cryst*. **6**, 705–707.

Curtis, C. W. & Reiner, I. (1962). *Representation theory of finite groups and associative algebras*. New York: Wiley–Interscience.

Daniels, H. E. (1954). *Saddlepoint approximation in statistics. Ann. Math. Stat.* **25**, 631–650.

Deisenhofer, J. & Steigemann, W. (1975). *Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. Acta Cryst*. B**31**, 238–250.

Diamond, R. (1971). *A real-space refinement procedure for proteins. Acta Cryst*. A**27**, 436–452.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961*a*). *The phase problem and isomorphous replacement methods in protein structures*. In *Computing methods and the phase problem in X-ray crystal analysis*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 236–251. Oxford: Pergamon Press.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961*b*). *The crystal structure of myoglobin: phase determination to a resolution of 2 Å by the method of isomorphous replacement. Acta Cryst.* **14**, 1188–1195.

Dietrich, H. (1972). *A reconsideration of Fourier methods for the refinement of crystal structures. Acta Cryst.* B**28**, 2807–2814.

Dieudonné, J. (1969). *Foundations of modern analysis*. New York and London: Academic Press.

Dieudonné, J. (1970). *Treatise on analysis*, Vol. II. New York and London: Academic Press.

Dirac, P. A. M. (1958). *The principles of quantum mechanics*, 4th ed. Oxford: Clarendon Press.

Dodson, E. J. (1981). *Block diagonal least squares refinement using fast Fourier techniques*. In *Refinement of protein structures*, compiled by P. A. Machin J. W. Campbell & M. Elder (ref. DL/SCI/R16), pp. 29–39. Warrington: SERC Daresbury Laboratory.

Donohue, J. & Schomaker, V. (1949). *The use of punched cards in molecular structure determinations*. III. *Structure factor calculations of X-ray crystallography. Acta Cryst.* **2**, 344–347.

Duane, W. (1925). *The calculation of the X-ray diffracting power at points in a crystal. Proc. Natl Acad. Sci. USA*, **11**, 489–493.

Dym, H. & McKean, H. P. (1972). *Fourier series and integrals*. New York and London: Academic Press.

Eklundh, J. O. (1972). *A fast computer method for matrix transposing. IEEE Trans*. C-**21**, 801–803.

Engel, P. (1986). *Geometric crystallography*. Dordrecht: Kluwer Academic Publishers.

Erdélyi, A. (1954). *Tables of integral transforms*, Vol. I. New York: McGraw-Hill.

Erdélyi, A. (1962). *Operational calculus and generalized functions*. New York: Holt, Rinehart & Winston.

Ewald, P. P. (1921). *Die Berechnung optischer und electrostatischer Gitterpotentiale. Ann. Phys. Leipzig*, **64**, 253–287.

Ewald, P. P. (1940). *X-ray diffraction by finite and imperfect crystal lattices. Proc. Phys. Soc. London*, **52**, 167–174.

Ewald, P. P. (1962). *Fifty years of X-ray diffraction*. Dordrecht: Kluwer Academic Publishers.

Farkas, D. R. (1981). *Crystallographic groups and their mathematics. Rocky Mountain J. Math.* **11**, 511–551.

Fornberg, A. (1981). *A vector implementation of the fast Fourier transform algorithm. Math. Comput.* **36**, 189–191.

Forsyth, J. B. & Wells, M. (1959). *On an analytical approximation to the atomic scattering factor. Acta Cryst.* **12**, 412–415.

Fowler, R. H. (1936). *Statistical mechanics*, 2nd ed. Cambridge University Press.

Fowweather, F. (1955). *The use of general programmes for crystallographic calculations on the Manchester University electronic digital computer (Mark II). Acta Cryst.* **8**, 633–637.

Freer, S. T., Alden, R. A., Levens, S. A. & Kraut, J. (1976). *Refinement of five protein structures by constrained $F_o - F_c$ Fourier methods*. In *Crystallographic computing techniques*, edited by F. R. Ahmed, pp. 317–321. Copenhagen: Munksgaard.

Friedel, G. (1913). *Sur les symétries cristallines que peut révéler la diffraction des rayons Röntgen. C. R. Acad. Sci. Paris*, **157**, 1533–1536.

Friedlander, F. G. (1982). *Introduction to the theory of distributions*. Cambridge University Press.

Friedlander, P. H., Love, W. & Sayre, D. (1955). *Least-squares refinement at high speed. Acta Cryst.* **8**, 732.

Friedman, A. (1970). *Foundations of modern analysis*. New York: Holt, Rinehart & Winston. [Reprinted by Dover, New York, 1982.]

Frobenius, G. (1911). *Über die unzerlegbaren diskreten Bewegungsgruppen. Sitzungsber. Preuss. Akad. Wiss. Berlin*, **29**, 654–665.

Gassmann, J. (1976). *Improvement and extension of approximate phase sets in structure determination*. In *Crystallographic computing techniques*, edited by F. R. Ahmed, pp. 144–154. Copenhagen: Munksgaard.

Gassmann, J. & Zechmeister, K. (1972). *Limits of phase expansion in direct methods. Acta Cryst.* A**28**, 270–280.

Gel'fand, I. M. & Shilov, G. E. (1964). *Generalized functions*, Vol. I. New York and London: Academic Press.

Gentleman, W. M. & Sande, G. (1966). *Fast Fourier transforms – for fun and profit*. In *AFIPS Proc. 1966 Fall Joint Computer Conference*, pp. 563–578. Washington, DC: Spartan Books.

Gillis, J. (1948*a*). *Structure factor relations and phase determination. Acta Cryst.* **1**, 76–80.

Gillis, J. (1948*b*). *The application of the Harker–Kasper method of phase determination. Acta Cryst.* **1**, 174–179.

Goedkoop, J. A. (1950). *Remarks on the theory of phase-limiting inequalities and equalities. Acta Cryst.* **3**, 374–378.

Goldstine, H. H. (1977). *A history of numerical analysis from the 16th through the 19th century*. New York: Springer-Verlag.

Good, I. J. (1958). *The interaction algorithm and practical Fourier analysis. J. R. Stat. Soc.* B**20**, 361–372.

Good, I. J. (1960). *Addendum* [to Good (1958)]. *J. R. Stat. Soc.* B**22**, 372–375.

Good, I. J. (1971). *The relationship between two fast Fourier transforms. IEEE Trans.* C-**20**, 310–317.

Greenhalgh, D. M. S. & Jeffrey, G. A. (1950). *A new punched card method of Fourier synthesis. Acta Cryst.* **3**, 311–312.

Grems, M. D. & Kasper, J. S. (1949). *An improved punched-card method for crystal structure-factor calculations. Acta Cryst.* **2**, 347–351.

Grenander, U. (1952). *On Toeplitz forms and stationary processes. Ark. Math.* **1**, 555–571.

Grenander, U. & Szegö, G. (1958). *Toeplitz forms and their applications*. Berkeley: University of California Press.

Guessoum, A. & Mersereau, R. M. (1986). *Fast algorithms for the multidimensional discrete Fourier transform. IEEE Trans. Acoust. Speech Signal Process.* **34**, 937–943.

Hadamard, J. (1932). *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques*. Paris: Hermann.

Hadamard, J. (1952). *Lectures on Cauchy's problem in linear partial differential equations*. New York: Dover Publications.

Hall, M. (1959). *The theory of groups*. New York: Macmillan.

Hardy, G. H. (1933). *A theorem concerning Fourier transforms. J. London Math. Soc.* **8**, 227–231.

Harker, D. (1936). *The application of the three-dimensional Patterson method and the crystal structures of proustite Ag_3AsS_3, and pyrargyrite, Ag_3SbS_3. J. Chem. Phys.* **4**, 381–390.

Harker, D. & Kasper, J. S. (1948). *Phases of Fourier coefficients directly from crystal diffraction data. Acta Cryst.* **1**, 70–75.

1.3 (*cont.*)

Harris, D. B., McClellan, J. H., Chan, D. S. K. & Schuessler, H. W. (1977). *Vector radix fast Fourier transform. Rec. 1977 IEEE Internat. Conf. Acoust. Speech Signal Process*. pp. 548–551.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Tomato bushy stunt virus at 2.9 Ångström resolution. Nature (London)*, **276**, 368–373.

Hartman, P. & Wintner, A. (1950). *On the spectra of Toeplitz's matrices. Am. J. Math.* **72**, 359–366.

Hartman, P. & Wintner, A. (1954). *The spectra of Toeplitz's matrices. Am. J. Math.* **76**, 867–882.

Hauptman, H. & Karle, J. (1953). *Solution of the phase problem. I. The centrosymmetric crystal*. ACA Monograph No. 3. Pittsburgh: Polycrystal Book Service.

Havighurst, R. J. (1925*a*). *The distribution of diffracting power in sodium chloride. Proc. Natl Acad. Sci. USA*, **11**, 502–507.

Havighurst, R. J. (1925*b*). *The distribution of diffracting power in certain crystals. Proc. Natl Acad. Sci. USA*, **11**, 507–512.

Heideman, M. T., Johnson, D. H. & Burrus, C. S. (1984). *Gauss and the history of the fast Fourier transform. IEEE Acoust. Speech Signal Process. Mag.* October 1984.

Hendrickson, W. A. & Konnert, J. H. (1980). *Incorporation of stereochemical information into crystallographic refinement*. In *Computing in crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.26. Bangalore: The Indian Academy of Science.

Herglotz, G. (1911). *Über Potenzreihen mit positiven, reellen Teil im Einheitskreis. Ber. Sächs. Ges. Wiss. Leipzig*, **63**, 501–511.

Hirschman, I. I. Jr & Hughes, D. E. (1977). *Extreme eigenvalues of Toeplitz operators. Lecture notes in mathematics*, Vol. 618. Berlin: Springer-Verlag.

Hodgson, M. L., Clews, C. J. B. & Cochran, W. (1949). *A punched card modification of the Beevers–Lipson method of Fourier synthesis. Acta Cryst.* **2**, 113–116.

Hoppe, W. & Gassmann, J. (1968). *Phase correction, a new method to solve partially known structures. Acta Cryst.* B**24**, 97–107.

Hoppe, W., Gassmann, J. & Zechmeister, K. (1970). *Some automatic procedures for the solution of crystal structures with direct methods and phase correction.* In *Crystallographic computing*, edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 26–36. Copenhagen: Munksgaard.

Hörmander, L. (1963). *Linear partial differential operators*. Berlin: Springer-Verlag.

Hörmander, L. (1973). *An introduction to complex analysis in several variables*, 2nd ed. Amsterdam: North-Holland.

Huber, R., Kulka, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974). *Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9Å resolution. J. Mol. Biol.* **89**, 73–101.

Hughes, E. W. (1941). *The crystal structure of melamine. J. Am. Chem. Soc.* **63**, 1737–1752.

Immirzi, A. (1973). *A general Fourier program for X-ray crystal-structure analysis which utilizes the Cooley–Tukey algorithm. J. Appl. Cryst.* **6**, 246–249.

Immirzi, A. (1976). *Fast Fourier transform in crystallography*. In *Crystallographic computing techniques*, edited by F. R. Ahmed, pp. 399–412. Copenhagen: Munksgaard.

Isaacs, N. W. (1982*a*). *The refinement of macromolecules*. In *Computational crystallography*, edited by D. Sayre, pp. 381–397. New York: Oxford University Press.

Isaacs, N. W. (1982*b*). *Refinement techniques: use of the FFT*. In *Computational crystallography*, edited by D. Sayre, pp. 398–408. New York: Oxford University Press.

Isaacs, N. W. (1984). *Refinement using the fast-Fourier transform least squares algorithm*. In *Methods and applications in crystallographic computing*, edited by S. R. Hall & T. Ashida, pp. 193–205. New York: Oxford University Press.

Isaacs, N. W. & Agarwal, R. C. (1978). *Experience with fast Fourier least squares in the refinement of the crystal structure of rhombohedral 2-zinc insulin at 1.5 Å resolution. Acta Cryst.* A**34**, 782–791.

Jack, A. (1973). *Direct determination of X-ray phases for tobacco mosaic virus protein using non-crystallographic symmetry. Acta Cryst.* A**29**, 545–554.

Jack, A. & Levitt, M. (1978). *Refinement of large structures by simultaneous minimization of energy and R factor. Acta Cryst.* A**34**, 931–935.

James, R. W. (1948*a*). *The optical principles of the diffraction of X-rays*. London: Bell.

James, R. W. (1948*b*). *False detail in three-dimensional Fourier representations of crystal structures. Acta Cryst.* **1**, 132–134.

Janssen, T. (1973). *Crystallographic groups*. Amsterdam: North-Holland.

Jaynes, E. T. (1957). *Information theory and statistical mechanics. Phys. Rev.* **106**, 620–630.

Jaynes, E. T. (1968). *Prior probabilities. IEEE Trans. SSC*, **4**, 227–241.

Jaynes, E. T. (1983). *Papers on probability, statistics and statistical physics*. Dordrecht: Kluwer Academic Publishers.

Johnson, H. W. & Burrus, C. S. (1983). *The design of optimal DFT algorithms using dynamic programming. IEEE Trans. Acoust. Speech Signal Process*. **31**, 378–387.

Jürgensen, H. (1970). *Calculation with the elements of a finite group given by generators and defining relations*. In *Computational problems in abstract algebra*, edited by J. Leech, pp. 47–57. Oxford: Pergamon Press.

Kac, M. (1954). *Toeplitz matrices, translation kernels, and a related problem in probability theory. Duke Math. J.* **21**, 501–509.

Kac, M., Murdock, W. L. & Szegö, G. (1953). *On the eigenvalues of certain Hermitian forms. J. Rat. Mech. Anal.* **2**, 767–800.

Karle, J. & Hauptman, H. (1950). *The phases and magnitudes of the structure factors. Acta Cryst.* **3**, 181–187.

Katznelson, Y. (1968). *An introduction to harmonic analysis*. New York: John Wiley.

Khinchin, A. I. (1949). *Mathematical foundations of statistical mechanics*. New York: Dover Publications.

Kitz, N. & Marchington, B. (1953). *A method of Fourier synthesis using a standard Hollerith senior rolling total tabulator. Acta Cryst.* **6**, 325–326.

Klug, A. (1958). *Joint probability distributions of structure factors and the phase problem. Acta Cryst.* **11**, 515–543.

Klug, A., Crick, F. H. C. & Wyckoff, H. W. (1958). *Diffraction by helical structures. Acta Cryst.* **11**, 199–213.

Kluyver, J. C. (1906). *A local probability problem. K. Ned. Akad. Wet. Proc.* **8**, 341–350.

Kolba, D. P. & Parks, T. W. (1977). *A prime factor FFT algorithm using high-speed convolution. IEEE Trans. Acoust. Speech Signal Process*. **25**, 281–294.

Konnert, J. H. (1976). *A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. Acta Cryst.* A**32**, 614–617.

Konnert, J. H. & Hendrickson, W. A. (1980). *A restrained-parameter thermal-factor refinement procedure. Acta Cryst.* A**36**, 344–350.

Korn, D. G. & Lambiotte, J. J. Jr (1979). *Computing the fast Fourier transform on a vector computer. Math. Comput.* **33**, 977–992.

Kuriyan, J., Brünger, A. T., Karplus, M. & Hendrickson, W. A. (1989). *X-ray refinement of protein structures by simulated annealing: test of the method on myohemerythrin. Acta Cryst.* A**45**, 396–409.

Lanczos, C. (1966). *Discourse on Fourier series*. Edinburgh: Oliver & Boyd.

Landau, H. J. & Pollack, H. O. (1961). *Prolate spheroidal wave functions, Fourier analysis and uncertainty (2). Bell Syst. Tech. J.* **40**, 65–84.

Landau, H. J. & Pollack, H. O. (1962). *Prolate spheroidal wave functions, Fourier analysis and uncertainty (3): the dimension of the space of essentially time- and band-limited signals. Bell Syst. Tech. J.* **41**, 1295–1336.

Lang, S. (1965). *Algebra*. Reading, MA: Addison-Wesley.

Larmor, J. (1934). *The Fourier discontinuities: a chapter in historical integral calculus. Philos. Mag.* **17**, 668–678.

1.3 (*cont.*)

Laue, M. von (1936). *Die aüßere Form der Kristalle in ihrem Einfluß auf die Interferenzerscheinungen an Raumgittern. Ann. Phys. (Leipzig)*, **26**, 55–68.

Lavoine, J. (1963). *Transformation de Fourier des pseudo-fonctions, avec tables de nouvelles transformées*. Paris: Editions du CNRS.

Ledermann, W. (1987). *Introduction to group characters*, 2nd ed. Cambridge University Press.

Leslie, A. G. W. (1987). *A reciprocal-space method for calculating a molecular envelope using the algorithm of B. C. Wang. Acta Cryst.* A**43**, 134–136.

Lighthill, M. J. (1958). *Introduction to Fourier analysis and generalized functions*. Cambridge University Press.

Linnik, I. Ju. (1975). *A multidimensional analogue of a limit theorem of G. Szegö. Math. USSR Izv*. **9**, 1323–1332.

Lipson, H. & Beevers, C. A. (1936). *An improved numerical method of two-dimensional Fourier synthesis for crystals. Proc. Phys. Soc. London*, **48**, 772–780.

Lipson, H. & Cochran, W. (1953). *The determination of crystal structures*. London: Bell.

Lipson, H. & Cochran, W. (1968). *The determination of crystal structures*. Revised and enlarged edition. London: G. Bell & Sons.

Lipson, H. & Taylor, C. A. (1951). *Optical methods in X-ray analysis. II. Fourier transforms and crystal-structure determination. Acta Cryst*. **4**, 458–462.

Lipson, H. & Taylor, C. A. (1958). *Fourier transforms and X-ray diffraction*. London: Bell.

Livesey, A. K. & Skilling, J. (1985). *Maximum entropy theory. Acta Cryst*. A**41**, 113–122.

Lonsdale, K. (1936). *Simplified structure factor and electron density formulae for the 230 space groups of mathematical crystallography*. London: Bell.

Lunin, V. Yu. (1985). *Use of the fast differentiation algorithm for phase refinement in protein crystallography. Acta Cryst*. A**41**, 551–556.

McClellan, J. H. & Rader, C. M. (1979). *Number theory in digital signal processing*. Englewood Cliffs: Prentice Hall.

MacGillavry, C. H. (1950). *On the derivation of Harker–Kasper inequalities. Acta Cryst*. **3**, 214–217.

MacLane, S. (1963). *Homology*. Berlin: Springer-Verlag.

Magnus, W., Karrass, A. & Solitar, D. (1976). *Combinatorial group theory: presentations of groups in terms of generators and relations*, 2nd revised ed. New York: Dover Publications.

Magnus, W., Oberhettinger, F. & Soni, R. P. (1966). *Formulas and theorems for the special functions of mathematical physics*. Berlin: Springer-Verlag.

Main, P. & Rossmann, M. G. (1966). *Relationships among structure factors due to identical molecules in different crystallographic environments. Acta Cryst*. **21**, 67–72.

Main, P. & Woolfson, M. M. (1963). *Direct determination of phases by the use of linear equations between structure factors. Acta Cryst*. **16**, 1046–1051.

Mayer, S. W. & Trueblood, K. N. (1953). *Three-dimensional Fourier summations on a high-speed digital computer. Acta Cryst*. **6**, 427.

Mersereau, R. & Speake, T. C. (1981). *A unified treatment of Cooley–Tukey algorithms for the evaluation of the multidimensional DFT. IEEE Trans. Acoust. Speech Signal Process*. **29**, 1011–1018.

Mersereau, R. M. (1979). *The processing of hexagonally sampled two-dimensional signals. Proc. IEEE*, **67**, 930–949.

Montroll, E. W., Potts, R. B. & Ward, J. C. (1963). *Correlations and spontaneous magnetization of the two-dimensional Ising model. J. Math. Phys.* **4**, 308–322.

Moore, D. H. (1971). *Heaviside operational calculus. An elementary foundation*. New York: American Elsevier.

Morris, R. L. (1978). *A comparative study of time efficient FFT and WFTA programs for general purpose computers. IEEE Trans. Acoust. Speech Signal Process*. **26**, 141–150.

Narayan, R. & Nityananda, R. (1982). *The maximum determinant method and the maximum entropy method. Acta Cryst*. A**38**, 122–128.

Natterer, F. (1986). *The mathematics of computerized tomography*. New York: John Wiley.

Navaza, J. (1985). *On the maximum-entropy estimate of the electron density function. Acta Cryst*. A**41**, 232–244.

Nawab, H. & McClellan, J. H. (1979). *Bounds on the minimum number of data transfers in WFTA and FFT programs. IEEE Trans. Acoust. Speech Signal Process*. **27**, 393–398.

Niggli, A. (1961). *Small-scale computers in X-ray crystallography*. In *Computing methods and the phase problem*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 12–20. Oxford: Pergamon Press.

Nussbaumer, H. J. (1981). *Fast Fourier transform and convolution algorithms*. Berlin: Springer-Verlag.

Nussbaumer, H. J. & Quandalle, P. (1979). *Fast computation of discrete Fourier transforms using polynomial transforms. IEEE Trans. Acoust. Speech Signal Process*. **27**, 169–181.

Onsager, L. (1944). *Crystal statistics. I. Two-dimensional model with an order–disorder transition. Phys. Rev.* **65**, 117–149.

Paley, R. E. A. C. & Wiener, N. (1934). *Fourier transforms in the complex domain*. Providence, RI: American Mathematical Society.

Patterson, A. L. (1934). *A Fourier series method for the determination of the components of interatomic distances in crystals. Phys. Rev.* **46**, 372–376.

Patterson, A. L. (1935*a*). *A direct method for the determination of the components of interatomic distances in crystals. Z. Kristallogr*. **90**, 517–542.

Patterson, A. L. (1935*b*). *Tabulated data for the seventeen plane groups. Z. Kristallogr.* **90**, 543–554.

Patterson, A. L. (1959). In *International tables for X-ray crystallography*, Vol. II, pp. 9–10. Erratum, January 1962. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

Pearson, K. (1905). *The problem of the random walk. Nature (London)*, **72**, 294, 342.

Pease, M. C. (1968). *An adaptation of the fast Fourier transform for parallel processing. J. Assoc. Comput. Mach.* **15**, 252–264.

Pepinsky, R. (1947). *An electronic computer for X-ray crystal structure analyses. J. Appl. Phys.* **18**, 601–604.

Pepinsky, R. (1952). *The use of positive kernels in Fourier syntheses of crystal structures*. In *Computing methods and the phase problem in X-ray crystal analysis*, edited by R. Pepinsky, pp. 319–338. State College: Penn. State College.

Pepinsky, R., van den Hende, J. & Vand, V. (1961). *X-RAC and digital computing methods*. In *Computing methods and the phase problem*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 154–160. Oxford: Pergamon Press.

Pepinsky, R. & Sayre, D. (1948). *Quantitative electron-density contour delineation in the electronic Fourier synthesizer for crystal structure analysis. Nature (London)*, **162**, 22–23.

Petrov, V. V. (1975). *Sums of independent random variables*. Berlin: Springer-Verlag.

Pollack, H. O. & Slepian, D. (1961). *Prolate spheroidal wave functions, Fourier analysis and uncertainty (1). Bell Syst. Tech. J.* **40**, 43–64.

Qurashi, M. M. (1949). *Optimal conditions for convergence of steepest descents as applied to structure determination. Acta Cryst*. **2**, 404–409.

Qurashi, M. M. (1953). *An analysis of the efficiency of convergence of different methods of structure determination. I. The methods of least squares and steepest descents: centrosymmetric case. Acta Cryst*. **6**, 577–588.

Qurashi, M. M. & Vand, V. (1953). *Weighting of the least-squares and steepest-descents methods in the initial stages of the crystal-structure determination. Acta Cryst*. **6**, 341–349.

Rader, C. M. (1968). *Discrete Fourier transforms when the number of data samples is prime. Proc. IEEE*, **56**, 1107–1108.

Rayleigh (J. W. Strutt), Lord (1880). *On the resultant of a large number of vibrations of the same pitch and arbitrary phase. Philos. Mag.* **10**, 73–78.

Rayleigh (J. W. Strutt), Lord (1899). *On James Bernoulli's theorem in probabilities. Philos. Mag.* **47**, 246–251.

1.3 (*cont.*)

Rayleigh (J. W. Strutt), Lord (1905). *The problem of the random walk. Nature (London)*, **72**, 318.

Rayleigh (J. W. Strutt), Lord (1918). *On the light emitted from a random distribution of luminous sources. Philos. Mag.* **36**, 429–449.

Rayleigh (J. W. Strutt), Lord (1919). *On the problem of random flights in one, two or three dimensions. Philos. Mag.* **37**, 321–347.

Reif, F. (1965). *Fundamentals of statistical and thermal physics*, Appendix A.6. New York: McGraw-Hill.

Reijen, L. L. van (1942). *Diffraction effects in Fourier syntheses and their elimination in X-ray structure investigations. Physica*, **9**, 461–480.

Rice, S. O. (1944, 1945). *Mathematical analysis of random noise. Bell Syst. Tech. J.* **23**, 283–332 (parts I and II); **24**, 46–156 (parts III and IV). [Reprinted in *Selected papers on noise and stochastic processes* (1954), edited by N. Wax, pp. 133–294. New York: Dover Publications.]

Riesz, M. (1938). *L'intégrale de Riemann–Liouville et le problème de Cauchy pour l'équation des ondes. Bull. Soc. Math. Fr.* **66**, 153–170.

Riesz, M. (1949). *L'intégrale de Riemann–Liouville et le problème de Cauchy. Acta Math.* **81**, 1–223.

Rivard, G. E. (1977). *Direct fast Fourier transform of bivariate functions. IEEE Trans. Acoust. Speech Signal Process.* **25**, 250–252.

Robertson, J. M. (1932). *A simple harmonic continuous calculating machine. Philos. Mag.* **13**, 413–419.

Robertson, J. M. (1935). *An X-ray study of the structure of phthalocyanines. Part I. The metal-free, nickel, copper, and platinum compounds. J. Chem. Soc.* pp. 615–621.

Robertson, J. M. (1936*a*). *Numerical and mechanical methods in double Fourier synthesis. Philos. Mag.* **21**, 176–187.

Robertson, J. M. (1936*b*). *Calculation of structure factors and summation of Fourier series in crystal analysis: non-centrosymmetrical projections. Nature (London)*, **138**, 683–684.

Robertson, J. M. (1936*c*). *An X-ray study of the phthalocyanines. Part II. Quantitative structure determination of the metal-free compound. J. Chem. Soc.* pp. 1195–1209.

Robertson, J. M. (1937). *X-ray analysis and application of Fourier series methods to molecular structures. Phys. Soc. Rep. Prog. Phys.* **4**, 332–367.

Robertson, J. M. (1954). *A fast digital computer for Fourier operations. Acta Cryst.* **7**, 817–822.

Robertson, J. M. (1955). *Some properties of Fourier strips with applications to the digital computer. Acta Cryst.* **8**, 286–288.

Robertson, J. M. (1961). *A digital mechanical computer for Fourier operations*. In *Computing methods and the phase problem*, edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 21–24. Oxford: Pergamon Press.

Rollett, J. S. (1965). *Structure factor routines*. In *Computing methods in crystallography*, edited by J. S. Rollett, pp. 38–46. Oxford: Pergamon Press.

Rollett, J. S. (1970). *Least-squares procedures in crystal structure analysis*. In *Crystallographic computing*, edited by F. R. Ahmed, pp. 167–181. Copenhagen: Munksgaard.

Rollett, J. S. & Davies, D. R. (1955). *The calculation of structure factors for centrosymmetric monoclinic systems with anisotropic atomic vibration. Acta Cryst.* **8**, 125–128.

Rossmann, M. G. & Blow, D. M. (1962). *The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst.* **15**, 24–31.

Rossmann, M. G. & Blow, D. M. (1963). *Determination of phases by the conditions of non-crystallographic symmetry. Acta Cryst.* **16**, 39–45.

Sayre, D. (1951). *The calculation of structure factors by Fourier summation. Acta Cryst.* **4**, 362–367.

Sayre, D. (1952*a*). *The squaring method: a new method for phase determination. Acta Cryst.* **5**, 60–65.

Sayre, D. (1952*b*). *Some implications of a theorem due to Shannon. Acta Cryst.* **5**, 843.

Sayre, D. (1952*c*). *The Fourier transform in X-ray crystal analysis*. In *Computing methods and the phase problem in X-ray crystal analysis*, edited by R. Pepinsky, pp. 361–390. State College: Penn. State University.

Sayre, D. (1972). *On least-squares refinement of the phases of crystallographic structure factors. Acta Cryst.* A**28**, 210–212.

Sayre, D. (1974). *Least-squares phase refinement. II. High-resolution phasing of a small protein. Acta Cryst.* A**30**, 180–184.

Sayre, D. (1980). *Phase extension and refinement using convolutional and related equation systems*. In *Theory and practice of direct methods in crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 271–286. New York and London: Plenum.

Schroeder, M. R. (1986). *Number theory in science and communication*, 2nd ed. Berlin: Springer-Verlag.

Schwartz, L. (1965). *Mathematics for the physical sciences*. Paris: Hermann, and Reading: Addison-Wesley.

Schwartz, L. (1966). *Théorie des distributions*. Paris: Hermann.

Schwarzenberger, R. L. E. (1980). *N-dimensional crystallography. Research notes in mathematics*, Vol. 41. London: Pitman.

Scott, W. R. (1964). *Group theory*. Englewood Cliffs: Prentice-Hall. [Reprinted by Dover, New York, 1987.]

Shaffer, P. A. Jr, Schomaker, V. & Pauling, L. (1946*a*). *The use of punched cards in molecular structure determinations. I. Crystal structure calculations. J. Chem. Phys.* **14**, 648–658.

Shaffer, P. A. Jr, Schomaker, V. & Pauling, L. (1946*b*). *The use of punched cards in molecular structure determinations. II. Electron diffraction calculations. J. Chem. Phys.* **14**, 659–664.

Shannon, C. E. (1949). *Communication in the presence of noise. Proc. Inst. Radio Eng. NY*, **37**, 10–21.

Shenefelt, M. (1988). *Group invariant finite Fourier transforms*. PhD thesis, Graduate Centre of the City University of New York.

Shmueli, U. & Weiss, G. H. (1985). *Exact joint probability distribution for centrosymmetric structure factors. Derivation and application to the Σ_1 relationship in the space group $P\bar{1}$. Acta Cryst.* A**41**, 401–408.

Shmueli, U. & Weiss, G. H. (1986). *Exact joint distribution of $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ and $E_{\mathbf{h}+\mathbf{k}}$, and the probability for the positive sign of the triple product in the space group $P\bar{1}$. Acta Cryst.* A**42**, 240–246.

Shmueli, U. & Weiss, G. H. (1987). *Exact random-walk models in crystallographic statistics. III. Distributions of $|E|$ for space groups of low symmetry. Acta Cryst.* A**43**, 93–98.

Shmueli, U. & Weiss, G. H. (1988). *Exact random-walk models in crystallographic statistics. IV. P.d.f.'s of $|E|$ allowing for atoms in special positions. Acta Cryst.* A**44**, 413–417.

Shmueli, U., Weiss, G. H. & Kiefer, J. E. (1985). *Exact random-walk models in crystallographic statistics. II. The bicentric distribution for the space group $P\bar{1}$. Acta Cryst.* A**41**, 55–59.

Shmueli, U., Weiss, G. H., Kiefer, J. E. & Wilson, A. J. C. (1984). *Exact random-walk models in crystallographic statistics. I. Space groups $P\bar{1}$ and P1. Acta Cryst.* A**40**, 651–660.

Shoemaker, D. P., Donohue, J., Schomaker, V. & Corey, R. B. (1950). *The crystal structure of L-threonine. J. Am. Chem. Soc.* **72**, 2328–2349.

Shoemaker, D. P. & Sly, W. G. (1961). *Computer programming strategy for crystallographic Fourier synthesis: program MIFR1. Acta Cryst.* **14**, 552.

Shohat, J. A. & Tamarkin, J. D. (1943). *The problem of moments. Mathematical surveys*, No. 1. New York: American Mathematical Society.

Silverman, H. F. (1977). *An introduction to programming the Winograd Fourier transform algorithm (WFTA). IEEE Trans. Acoust. Speech Signal Process.* **25**, 152–165.

Singleton, R. C. (1969). *An algorithm for computing the mixed radix fast Fourier transform. IEEE Trans. AU*, **17**, 93–103.

Sneddon, I. N. (1951). *Fourier transforms*. New York: McGraw-Hill.

Sneddon, I. N. (1972). *The use of integral transforms*. New York: McGraw-Hill.

Sparks, R. A., Prosen, R. J., Kruse, F. H. & Trueblood, K. N. (1956). *Crystallographic calculations on the high-speed digital computer SWAC. Acta Cryst.* **9**, 350–358.

Sprecher, D. A. (1970). *Elements of real analysis*. New York: Academic Press. [Reprinted by Dover Publications, New York, 1987.]

Stout, G. H. & Jensen, L. H. (1968). *X-ray structure determination. A practical guide*. New York: Macmillan.

Suryan, G. (1957). *An analogue computer for double Fourier series*

1.3 (*cont.*)

summation for X-ray crystal-structure analysis. Acta Cryst. **10**, 82–84.

Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. (1977). *A structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters. Acta Cryst.* A**33**, 800–804.

Swartzrauber, P. N. (1984). *FFT algorithms for vector computers. Parallel Comput.* **1**, 45–63.

Szegö, G. (1915). *Ein Grenzwertsatz uber die Toeplitzschen Determinanten einer reellen positiven Funktion. Math. Ann.* **76**, 490–503.

Szegö, G. (1920). *Beitrage zur Theorie der Toeplitzchen Formen (Erste Mitteilung). Math. Z.* **6**, 167–202.

Szegö, G. (1952). *On certain Hermitian forms associated with the Fourier series of a positive function. Comm. Sém. Math. Univ. Lund* (Suppl. dedicated to Marcel Riesz), pp. 228–238.

Takano, T. (1977*a*). *Structure of myoglobin refined at 2.0 Å resolution. I. Crystallographic refinement of metmyoglobin from sperm whale. J. Mol. Biol.* **110**, 537–568.

Takano, T. (1977*b*). *Structure of myoglobin refined at 2.0 Å resolution. II. Structure of deoxymyoglobin from sperm whale. J. Mol. Biol.* **110**, 569–584.

Temperton, C. (1983*a*). *Self-sorting mixed-radix fast Fourier transforms. J. Comput. Phys.* **52**, 1–23.

Temperton, C. (1983*b*). *A note on the prime factor FFT algorithm. J. Comput. Phys.* **52**, 198–204.

Temperton, C. (1983*c*). *Fast mixed-radix real Fourier transforms. J. Comput. Phys.* **52**, 340–350.

Temperton, C. (1985). *Implementation of a self-sorting in-place prime factor FFT algorithm. J. Comput. Phys.* **58**, 283–299.

Ten Eyck, L. F. (1973). *Crystallographic fast Fourier transforms. Acta Cryst.* A**29**, 183–191.

Ten Eyck, L. F. (1977). *Efficient structure factor calculation for large molecules by fast Fourier transform. Acta Cryst.* A**33**, 486–492.

Ten Eyck, L. F. (1985). *Fast Fourier transform calculation of electron density maps*. In *Diffraction methods for biological macromolecules* (*Methods in enzymology*, Vol. 115), edited by H. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 324–337. New York: Academic Press.

Titchmarsh, E. C. (1922). *Hankel transforms. Proc. Camb. Philos. Soc.* **21**, 463–473.

Titchmarsh, E. C. (1948). *Introduction to the theory of Fourier integrals*. Oxford: Clarendon Press.

Toeplitz, O. (1907). *Zur Theorie der quadratischen Formen von unendlichvielen Variablen. Nachr. der Kgl. Ges. Wiss. Göttingen, Math. Phys. Kl.* pp. 489–506.

Toeplitz, O. (1910). *Zur Transformation der Scharen bilinearer Formen von unendlichvielen Veränderlichen. Nachr. der Kgl. Ges. Wiss. Göttingen, Math. Phys. Kl.* pp. 110–115.

Toeplitz, O. (1911*a*). *Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. I. Teil: Theorie der L-formen. Math. Ann.* **70**, 351–376.

Toeplitz, O. (1911*b*). *Über die Fouriersche Entwicklung positiver Funktionen. Rend. Circ. Mat. Palermo*, **32**, 191–192.

Tolimieri, R. (1985). *The algebra of the finite Fourier transform and coding theory. Trans. Am. Math. Soc.* **287**, 253–273.

Tolimieri, R., An, M. & Lu, C. (1989). *Algorithms for discrete Fourier transform and convolution*. New York: Springer-Verlag.

Tolstov, G. P. (1962). *Fourier series*. Englewood Cliffs, NJ: Prentice-Hall.

Trèves, F. (1967). *Topological vector spaces, distributions, and kernels*. New York and London: Academic Press.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Cryst.* A**43**, 489–501.

Trueblood, K. N. (1956). *Symmetry transformations of general anisotropic temperature factors. Acta Cryst.* **9**, 359–361.

Truter, M. R. (1954). *Refinement of a non-centrosymmetrical structure: sodium nitrite. Acta Cryst.* **7**, 73–77.

Uhrich, M. L. (1969). *Fast Fourier transforms without sorting. IEEE Trans. AU*, **17**, 170–172.

Van der Pol, B. & Bremmer, H. (1955). *Operational calculus*, 2nd ed. Cambridge University Press.

Vand, V. (1948). *Method of steepest descents: improved formula for X-ray analysis. Nature (London)*, **161**, 600–601.

Vand, V. (1951). *A simplified method of steepest descents. Acta Cryst.* **4**, 285–286.

Vand, V., Eiland, P. F. & Pepinsky, R. (1957). *Analytical representation of atomic scattering factors. Acta Cryst.* **10**, 303–306.

Wang, B. C. (1985). *Resolution of phase ambiguity in macromolecular crystallography*. In *Diffraction methods for biological macromolecules* (*Methods in enzymology*, Vol. 115), edited by H. Wyckoff, C. H. W. Hirs & S. N. Timasheff, pp. 90–112. New York: Academic Press.

Warren, B. & Bragg, W. L. (1929). *The structure of diopside, $CaMg(SiO_3)_2$. Z. Kristallogr.* **69**, 168–193.

Warren, B. E. (1969). *X-ray diffraction*. Reading: Addison-Wesley.

Warren, B. E. & Gingrich, N. S. (1934). *Fourier integral analysis of X-ray powder patterns. Phys. Rev.* **46**, 368–372.

Waser, J. (1955*a*). *Symmetry relations between structure factors. Acta Cryst.* **8**, 595.

Waser, J. (1955*b*). *The anisotropic temperature factor in triclinic coordinates. Acta Cryst.* **8**, 731.

Waser, J. & Schomaker, V. (1953). *The Fourier inversion of diffraction data. Rev. Mod. Phys.* **25**, 671–690.

Watson, G. L. (1970). *Integral quadratic forms*. Cambridge University Press.

Watson, G. N. (1944). *A treatise on the theory of Bessel functions*, 2nd ed. Cambridge University Press.

Wells, M. (1965). *Computational aspects of space-group symmetry. Acta Cryst.* **19**, 173–179.

Weyl, H. (1931). *The theory of groups and quantum mechanics*. New York: Dutton. [Reprinted by Dover Publications, New York, 1950.]

Whittaker, E. J. W. (1948). *The evaluation of Fourier transforms by a Fourier synthesis method. Acta Cryst.* **1**, 165–167.

Whittaker, E. T. (1915). *On the functions which are represented by the expansions of the interpolation-theory. Proc. R. Soc. (Edinburgh)*, **35**, 181–194.

Whittaker, E. T. (1928). *Oliver Heaviside. Bull. Calcutta Math. Soc.* **20**, 199–220. [Reprinted in Moore (1971).]

Whittaker, E. T. & Robinson, G. (1944). *The calculus of observations*. London: Blackie.

Whittaker, E. T. & Watson, G. N. (1927). *A course of modern analysis*, 4th ed. Cambridge University Press.

Widom, H. (1960). *A theorem on translation kernels in n dimensions. Trans. Am. Math. Soc.* **94**, 170–180.

Widom, H. (1965). *Toeplitz matrices*. In *Studies in real and complex analysis*, edited by I. I. Hirschmann Jr, pp. 179–209. *MAA studies in mathematics*, Vol. 3. Englewood Cliffs: Prentice-Hall.

Widom, H. (1975). *Asymptotic inversion of convolution operators. Publ. Math. IHES*, **44**, 191–240.

Wiener, N. (1933). *The Fourier integral and certain of its applications*. Cambridge University Press. [Reprinted by Dover Publications, New York, 1959.]

Wilkins, S. W., Varghese, J. N. & Lehmann, M. S. (1983). *Statistical geometry. I. A self-consistent approach to the crystallographic inversion problem based on information theory. Acta Cryst.* A**39**, 47–60.

Winograd, S. (1976). *On computing the discrete Fourier transform. Proc. Natl Acad. Sci. USA*, **73**, 1005–1006.

Winograd, S. (1977). *Some bilinear forms whose multiplicative complexity depends on the field of constants. Math. Syst. Theor.* **10**, 169–180.

Winograd, S. (1978). *On computing the discrete Fourier transform. Math. Comput.* **32**, 175–199.

Winograd, S. (1980). *Arithmetic complexity of computations*. CBMS-NST Regional Conf. Series Appl. Math, Publ. No. 33. Philadelphia: SIAM Publications.

1.3 (*cont.*)

Wolf, J. A. (1967). *Spaces of constant curvature.* New York: McGraw-Hill.

Wondratschek, H. (1995). *Introduction to space-group symmetry.* In *International tables for crystallography*, Vol. A. *Space-group symmetry*, edited by Th. Hahn, pp. 712–735. Dordrecht: Kluwer Academic Publishers.

Yosida, K. (1965). *Functional analysis.* Berlin: Springer-Verlag.

Zachariasen, W. H. (1945). *Theory of X-ray diffraction in crystals.* New York: John Wiley.

Zassenhaus, H. (1948). *Über eine Algorithmus zur Bestimmung der Raumgruppen. Commun. Helv. Math.* **21**, 117–141.

Zemanian, A. H. (1965). *Distribution theory and transform analysis.* New York: McGraw-Hill.

Zemanian, A. H. (1968). *Generalised integral transformations.* New York: Interscience.

Zygmund, A. (1959). *Trigonometric series*, Vols. 1 and 2. Cambridge University Press.

Zygmund, A. (1976). *Notes on the history of Fourier series*. In *Studies in harmonic analysis*, edited by J. M. Ash, pp. 1–19. *MAA studies in mathematics*, Vol. 13. The Mathematical Association of America.

1.4

Altermatt, U. D. & Brown, I. D. (1987). *A real-space computer-based symmetry algebra. Acta Cryst.* A**43**, 125–130.

Authier, A. (1981). *The reciprocal lattice*. Edited by the IUCr Commission on Crystallographic Teaching. Cardiff: University College Cardiff Press.

Bertaut, E. F. (1964). *On the symmetry of phases in the reciprocal lattice: a simple method. Acta Cryst.* **17**, 778–779.

Bienenstock, A. & Ewald, P. P. (1962). *Symmetry of Fourier space. Acta Cryst.* **15**, 1253–1261.

Bourne, P. E., Berman, H. M., McMahon, B., Watenpaugh, K. D., Westbrook, J. D. & Fitzgerald, P. M. D. (1997). *mmCIF: macromolecular crystallographic information file. Methods Enzymol.* **277**, 571–590.

Bragg, L. (1966). *The crystalline state*. Vol. I. *A general survey*. London: Bell.

Buerger, M. J. (1942). *X-ray crystallography*. New York: John Wiley.

Buerger, M. J. (1949). *Crystallographic symmetry in reciprocal space. Proc. Natl Acad. Sci. USA*, **35**, 198–201.

Buerger, M. J. (1960). *Crystal structure analysis*. New York: John Wiley.

Computers in the New Laboratory – a Nature Survey (1981). *Nature (London)*, **290**, 193–200.

Dowty, E. (2000). *ATOMS: a program for the display of atomic structures.* Commercial software available from http://www.shapesoftware.com.

Ewald, P. P. (1921). *Das 'reziproke Gitter' in der Strukturtheorie. Z. Kristallogr. Teil A*, **56**, 129–156.

Grosse-Kunstleve, R. W. (1995). *SGINFO: a comprehensive collection of ANSI C routines for the handling of space-group symmetry*. Freely available from http://www.kristall.ethz.ch/LFK/software/sginfo/.

Grosse-Kunstleve, R. W. (1999). *Algorithms for deriving crystallographic space-group information. Acta Cryst.* A**55**, 383–395.

Hall, S. R. (1981*a*). *Space-group notation with an explicit origin. Acta Cryst.* A**37**, 517–525.

Hall, S. R. (1981*b*). *Space-group notation with an explicit origin: erratum. Acta Cryst.* A**37**, 921.

Hall, S. R. (1997). *LoopX: a script used to loop Xtal.* Copyright University of Western Australia. Freely available from http://www.crystal.uwa.edu.au/~syd/symmetry/.

Hall, S. R., du Boulay, D. J. & Olthof-Hazekamp, R. (2000). *Xtal: a system of crystallographic programs.* Copyright University of Western Australia. Freely available from http://xtal.crystal.uwa.edu.au.

Hearn, A. C. (1973). *REDUCE 2.0 user's manual*. University of Utah, Salt Lake City, Utah, USA.

Hovmöller, S. (1992). *CRISP: crystallographic image processing on a personal computer. Ultramicroscopy*, **41**, 121–135.

International Tables for Crystallography (1983). Vol. A. *Space-group symmetry*, edited by Th. Hahn. Dordrecht: Reidel. (Present distributor Kluwer Academic Publishers, Dordrecht.)

International Tables for Crystallography (1987). Vol. A. *Space-group symmetry*, edited by Th. Hahn, 2nd ed. Dordrecht: Reidel. (Present distributor Kluwer Academic Publishers, Dordrecht.)

International Tables for Crystallography (1992). Vol. A. *Space-group symmetry*, edited by Th. Hahn, 3rd ed. Dordrecht: Kluwer Academic Publishers.

International Tables for Crystallography (1993). Vol. B. *Reciprocal space*, edited by U. Shmueli, 1st ed. Dordrecht: Kluwer Academic Publishers.

International Tables for X-ray Crystallography (1952). Vol. I. *Symmetry groups*, edited by N. F. M. Henry & K. Lonsdale. Birmingham: Kynoch Press.

International Tables for X-ray Crystallography (1965). Vol. I. *Symmetry groups*, edited by N. F. M. Henry & K. Lonsdale, 2nd ed. Birmingham: Kynoch Press.

Internationale Tabellen zur Bestimmung von Kristallstrukturen (1935). 1. Band. Edited by C. Hermann. Berlin: Borntraeger.

Koch, E. & Fischer, W. (1983). In *International tables for crystallography*, Vol. A, ch. 11. *Symmetry operations*. Dordrecht: Reidel. (Present distributor Kluwer Academic Publishers, Dordrecht.)

Larine, M., Klimkovich, S., Farrants, G., Hovmöller, S. & Xiaodong, Z. (1995). *Space Group Explorer: a computer program for obtaining extensive information about any of the 230 space groups.* Freely available from http://www.calidris-em.com/.

Lipson, H. & Cochran, W. (1966). *The crystalline state*. Vol. II. *The determination of crystal structures*. London: Bell.

Lipson, H. & Taylor, C. A. (1958). *Fourier transforms and X-ray diffraction*. London: Bell.

Seitz, F. (1935). *A matrix-algebraic development of crystallographic groups. III. Z. Kristallogr.* **90**, 289–313.

Shmueli, U. (1984). *Space-group algorithms. I. The space group and its symmetry elements. Acta Cryst.* A**40**, 559–567.

Waser, J. (1955). *Symmetry relations between structure factors. Acta Cryst.* **8**, 595.

Wells, M. (1965). *Computational aspects of space-group symmetry. Acta Cryst.* **19**, 173–179.

Zachariasen, W. H. (1945). *Theory of X-ray diffraction in crystals*. New York: John Wiley.

1.5

Altmann, S. L. (1977). *Induced representations in crystals and molecules.* London: Academic Press.

Bouckaert, L. P., Smoluchowski, R. & Wigner, E. P. (1936). *Theory of Brillouin zones and symmetry properties of wave functions in crystals. Phys. Rev.* **50**, 58–67.

Boyle, L. L. (1986). *The classification of space group representations.* In *Proceedings of the 14th international colloquium on group-theoretical methods in physics*, pp. 405–408. Singapore: World Scientific.

Boyle, L. L. & Kennedy, J. M. (1988). *Raumgruppen Charaktereintafeln. Z. Kristallogr.* **182**, 39–42.

Bradley, C. J. & Cracknell, A. P. (1972). *The mathematical theory of symmetry in solids: representation theory for point groups and space groups.* Oxford: Clarendon Press.

Cracknell, A. P., Davies, B. L., Miller, S. C. & Love, W. F. (1979). *Kronecker product tables*, Vol. 1. *General introduction and tables of irreducible representations of space groups.* New York: IFI/Plenum.

Davies, B. L. & Cracknell, A. P. (1976). *Some comments on and addenda to the tables of irreducible representations of the*