JOURNAL OF
COMPUTATIONAL AND
APPLIED MATHEMATICS

ELSEVIER Journal of Computational and Applied Mathematics 64 (1995) 41–56

An introduction to wavelets with applications to Andrews' plots

Paul Embrechts[a,*], Agnes M. Herzberg[b], Heidi K. Kalbfleisch[b], William N. Traves[c], J. Robertson Whitla[d]

[a] *ETH Zentrum, CH-8092, Zürich, Switzerland*
[b] *Department of Mathematics and Statistics, Queen's University, Kingston, Canada*
[c] *University of Toronto, Toronto, Canada*
[d] *Arthur Andersen and Co., Toronto, Canada*

Received 13 July 1993; revised 27 May 1994

Abstract

A brief introduction to wavelets targeted to the statistician is given. Several types of wavelets are described and compared with the Fourier functions. It is shown how wavelets might extend the multivariate clustering technique of Andrews' plots.

Keywords: Wavelets; Fourier functions; Haar wavelet; Exploratory data analysis; Andrews' plots

1. Introduction

Wavelets are families of functions obtained by taking the dilations and translations of a particular function with sufficient decay in both the time and frequency domains. Several wavelet functions will be described and compared with the well-known Fourier basis functions. The similarities and differences between these two types of analysis will be illustrated. The wavelet transform, which maps functions in one variable to functions in two variables, will be introduced. An application of wavelets will be presented, namely the use of wavelets in Andrews' plots.

Haar [15] first constructed an alternative basis to the Fourier functions for $L^2[0, 1)$. Since the Haar functions have point discontinuities, Franklin [13] followed with an alternative basis for $L^2[0, 1)$ with continuous basis functions. It was not until the 1980s, however, that Haar's

* Corresponding author.

SSDI 0377-0427(95)00005-4

construction, and variations on Franklin's construction [18] were to be described as wavelet families.

2. Wavelets

Wavelets, in general, are constructed by taking the dilations and translations of a single function with sufficient decay in both the time and frequency domains. The definition adopted here for "sufficient" decay is that a function $\Psi(x)$ and its Fourier transform, denoted by $\hat{\Psi}(f)$, both decay faster than $|x|^{-1}$ and $|f|^{-1}$, respectively; see [6], i.e.,

$$\int_{-\infty}^{\infty} |x|^{-1} |\Psi(x)|^2 \,\mathrm{d}x < \infty, \tag{1}$$

and

$$\int_{-\infty}^{\infty} |f|^{-1} |\hat{\Psi}(f)|^2 \,\mathrm{d}f < \infty. \tag{2}$$

A set of wavelets shares many of the properties of the particular function, $\Psi(x)$, the basic wavelet, known also as the mother wavelet, such as regularity (differentiability), continuity and the magnitude of the dilations and translations. Typical examples of basic wavelets are:

(a) the Haar wavelet,

$$\Psi(x) = \begin{cases} 1, & 0 \leqslant x < \frac{1}{2}, \\ -1, & \frac{1}{2} \leqslant x < 1, \\ 0 & \text{elsewhere}; \end{cases}$$

(b) the Franklin wavelet,

$$\Psi(x) = \begin{cases} 0, & 0 \leqslant x < \frac{1}{2}, \\ 2x - 1, & \frac{1}{2} \leqslant x < 1; \end{cases}$$

(c) the stetson hat wavelet,

$$\Psi(x) = \begin{cases} -2x - 1, & -\frac{1}{2} \leqslant x < 0, \\ 6x - 1, & 0 \leqslant x < \frac{1}{2}, \\ -6x + 5, & \frac{1}{2} \leqslant x < 1, \\ 2x - 3, & 1 \leqslant x < \frac{3}{2}; \end{cases}$$

(d) the Mexican hat wavelet,

$$\Psi(x) = \frac{2}{\sqrt{3}} \pi^{-1/4} (1 - x^2) \mathrm{e}^{-x^2/2}.$$

These wavelets are illustrated in Fig. 1.

Let $\Psi(x)$ be a basic wavelet, and let a ($a \neq 0$) and b be real numbers; the family of wavelets corresponding to $\Psi(x)$ is

$$\Psi_{a,b}(x) = |a|^{-1/2}\Psi\{(x-b)/a\}, \tag{3}$$

where a gives the dilation and b gives the translation. The factor $|a|^{-1/2}$ normalizes the family of wavelets. A set of wavelets can be assembled by successively constructing each wavelet, $\Psi_{a,b}(x)$ for

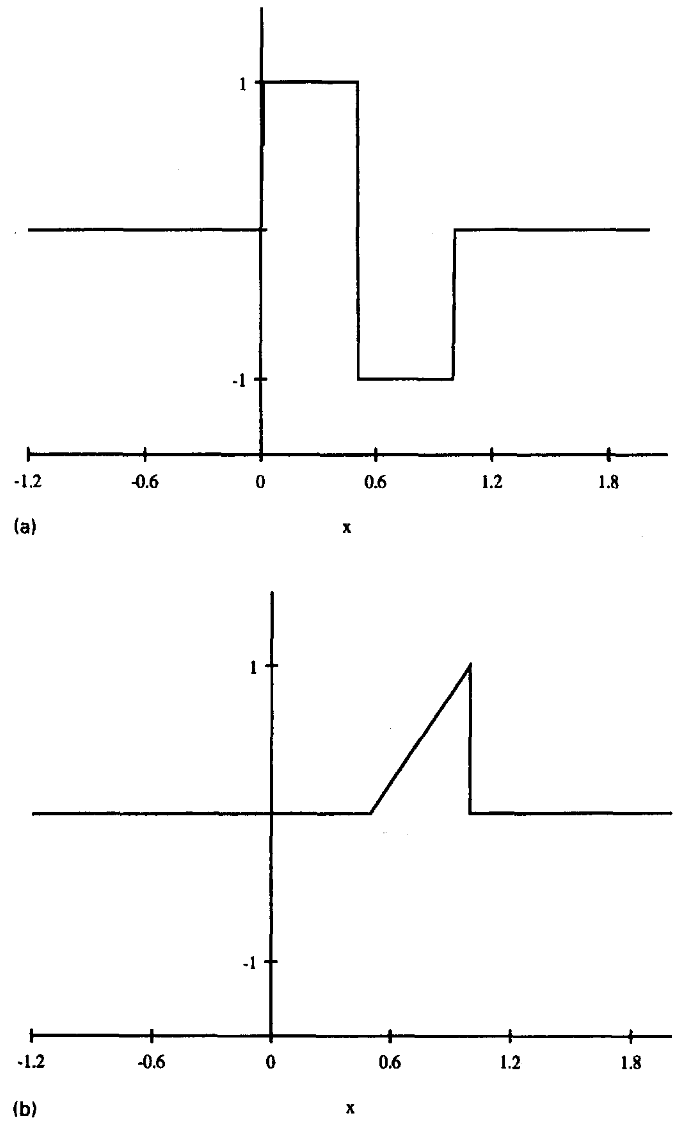

Fig. 1. Examples of wavelets: (a) the Haar wavelet; (b) the Franklin wavelet; (c) the stetson hat wavelet; (d) the Mexican hat wavelet.

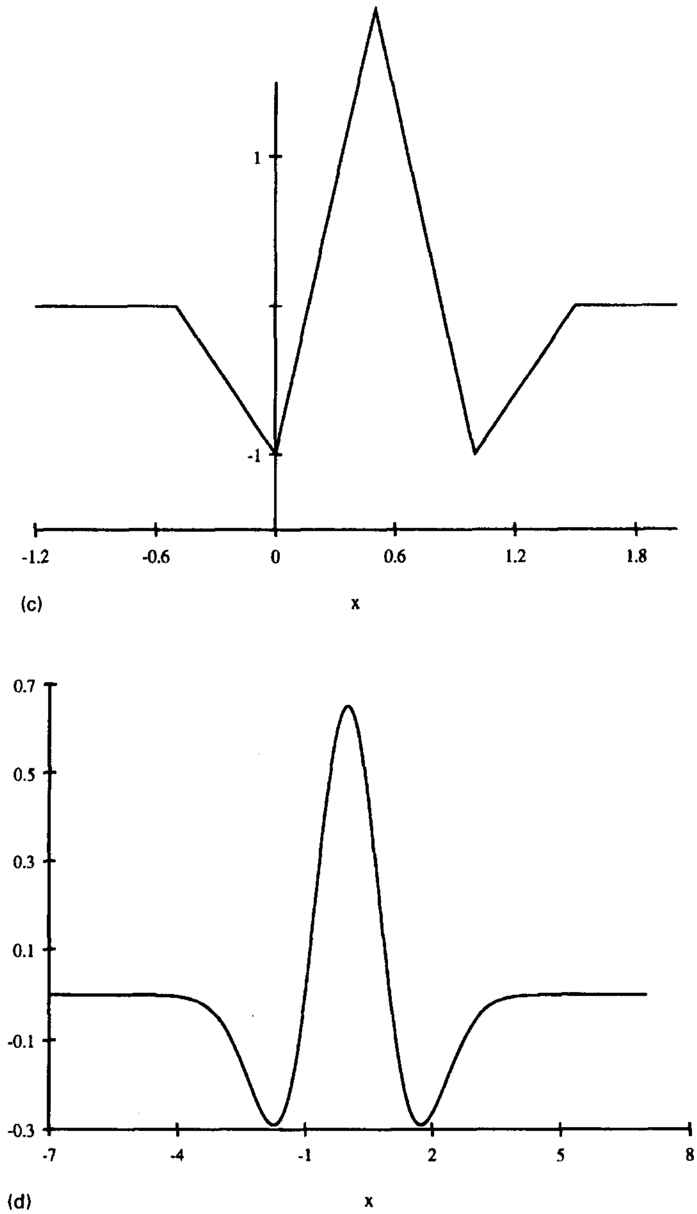

Fig. 1(c)–(d).

different values of a and/or b as in (3). Wavelet families which are constructed by allowing a and b to vary continuously as in (3) are called *continuous wavelet* families.

Discrete wavelets are constructed by constraining the values of a and b to a discrete lattice of points. A standard choice is to let $a = 2^{-j}$ and let $b = k\,(0 \leqslant k < 2^j)$, where j and k are integers. For a further discussion on the choice of lattice parameters a and b, see [8, p. 9] . The label *discrete* *wavelets* does not imply that the wavelets are discrete-valued, but that the dilates and translates themselves are discretized.

Consider, for example, the Haar wavelet family; see Fig. 2. Rewrite (3) as

$$\Psi_{j,k}(x) = 2^{j/2}\,\Psi(2^j x - k),$$

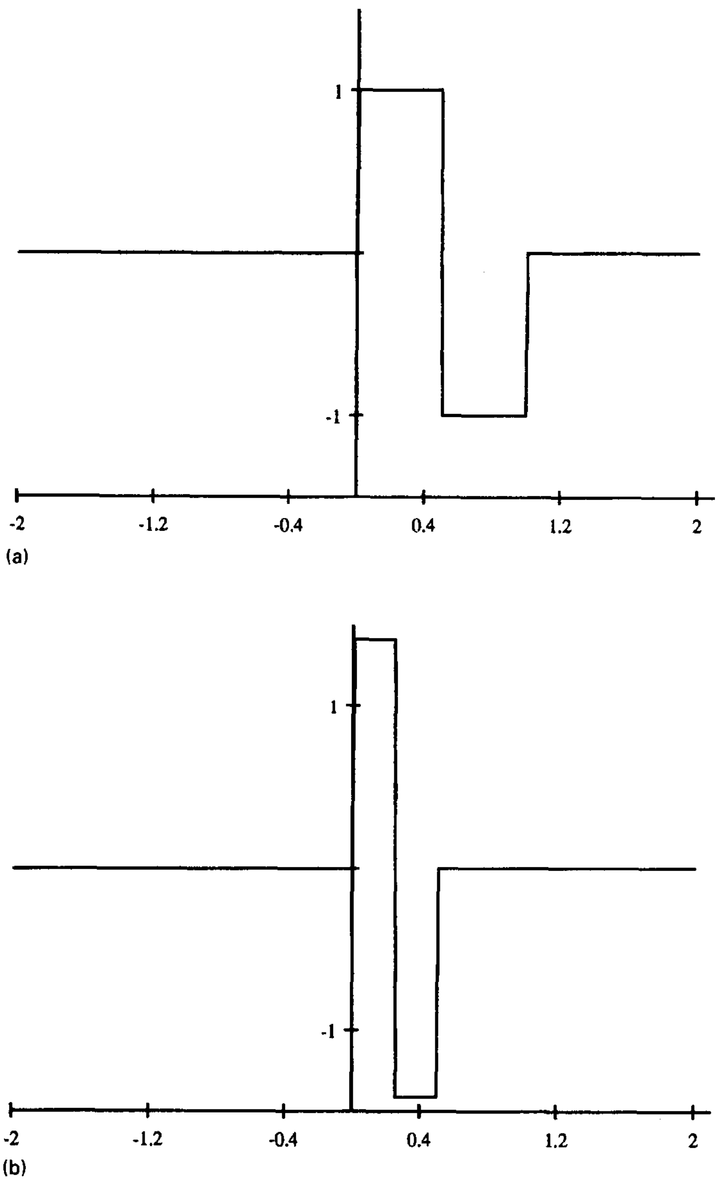

Fig. 2. Dilations and translations of the Haar wavelet: (a) the Haar wavelet: $\Psi_{1,0}(x)$; (b) compression by two, zero translation: $\Psi_{1,0}(x)$; (c) compression by two, translation by one: $\Psi_{1,1}(x)$; (d) compression by four, translation by one: $\Psi_{2,1}(x)$; (e) compression by four, translation by three: $\Psi_{2,3}(x)$.

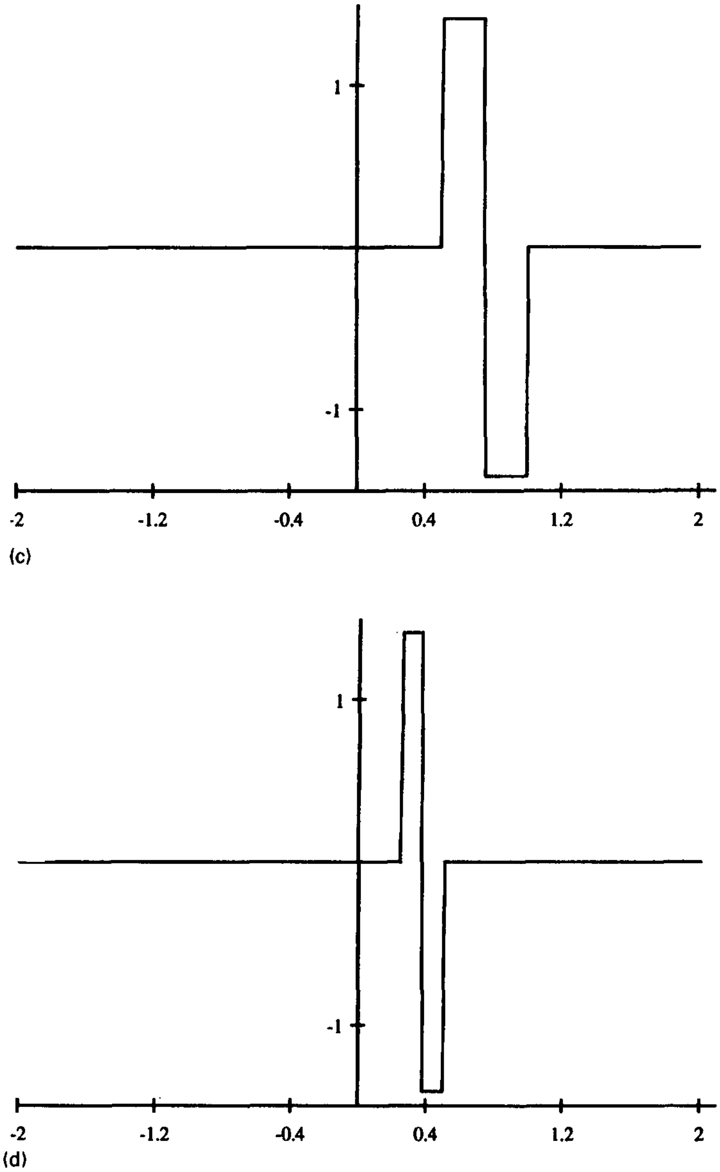

Fig. 2(c)–(d).

where in terms of the previous notation $a = 2^{-j}$ and $b = 2^{-j}k$ on the right-hand side of (3), k and j being integers. The function $\Psi_{1,1}(x)$ is a translate of $\Psi_{1,0}(x)$ and vice versa. The same is true for j held constant at a different value. The function $\Psi_{2,1}(x)$ in Fig. 2(d) is a translate of $\Psi_{2,3}(x)$ in Fig. 2(e). Changing the parameter k changes the location of the wavelet in x at a fixed resolution; thus, changing k represents translation in time. Changing j dilates the basic wavelet and hence changes the *frequency resolution*. For discrete wavelets based upon dilates of two, changing the parameter j by one corresponds to a change in frequency by a factor of two. Wavelets with the same

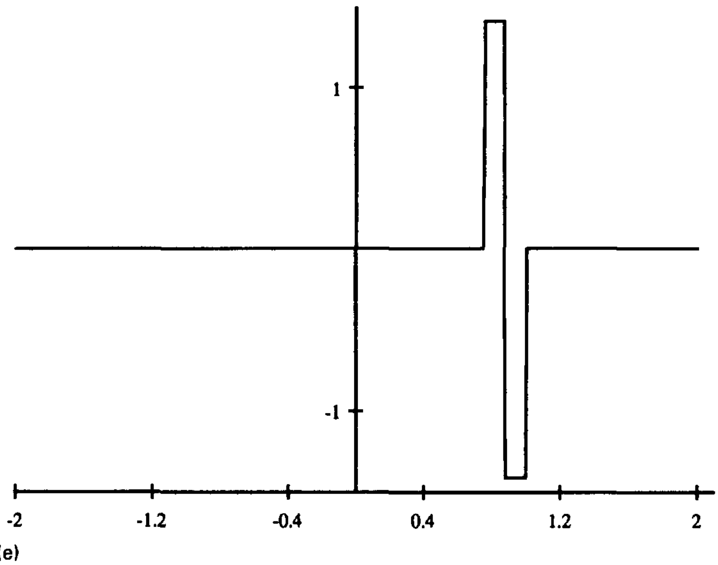

Fig. 2(e).

parameter j are at the same frequency resolution described as resolution 2^j. For further information about the construction of wavelets, see [10].

For any particular wavelet basis, one may choose to represent a function as a linear combination of a projection onto basis vectors. Typically, the Fourier representation over an interval of a function depends heavily on cancellation; see, for example, [1] or [4]. By contrast, the superposition of wavelets in a representation is more of a stacking up of building blocks. To illustrate this difference, one can compare the basic building blocks of each representation; see Fig. 3. As the figure illustrates, the Fourier functions rely on high frequencies and low frequencies cancelling out to obtain convergence. The Haar unnormalized wavelets, in this example, are clearly localized.

Wavelets have some clear advantageous characteristics over the Fourier series. One important idea to note is that properties of wavelet functions are translated directly into their wavelet coefficients unlike the situation with the Fourier series; see [19].

The span of a basis of arbitrary functions constructed using (3) may not be complete, and is unlikely to have orthogonal basis vectors. It is no accident, for example, that the Haar wavelets form a complete orthonormal basis for $L^2[0, 1)$. In contrast, the Mexican hat wavelets do not form a complete orthonormal basis for $L^2(\mathbb{R})$ over dilates of two, but can form a useful nonorthogonal span of $L^2(\mathbb{R})$ by dilating by powers of $2^{-1/4}$; see [7, 8]. Further development of these theories is best examined formally as in [6, 8].

Special classes of wavelets possess a transform analogous to the Fourier transform enabling dual time–frequency domain analysis. In many areas of signal analysis, the Fourier transform is used to provide a measure of the frequency content of a signal over its entire domain [4]. In contrast, the wavelet transform is used when a measure of the frequency content in bands localized in time is desired [16]. Fourier transforms and wavelet transforms are, therefore, complementary; Fourier transforms give the time average of frequencies in the signal and wavelet transforms give the frequency strengths in a signal over arbitrary intervals of time.

The continuous wavelet transform is defined in an analogous manner to the Fourier transform, i.e., the wavelet transform [7] of the function $f(x)$ is

$$\tilde{f}(a, b) = \int_{-\infty}^{\infty} |a|^{-1/2} f(x) \overline{\Psi\{(x-b)/a\}}\, \mathrm{d}x; \tag{4}$$

the inverse continuous wavelet transform [16] is

$$f(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{f}(a, b)\, \Psi\{(x-b)/a\}\, \mathrm{d}a\, \mathrm{d}b. \tag{5}$$

The discrete wavelet transform is a transform of a continuous function $f(x)$, but is so named because it is the transform relative to discrete wavelets. For orthogonal wavelets, the wavelet transform projects a function onto distinct frequency channels localized in time to yield a series representation of the function. For more on the discrete wavelet transform, see [7], and the inversion of nonorthogonal wavelets, see [9].

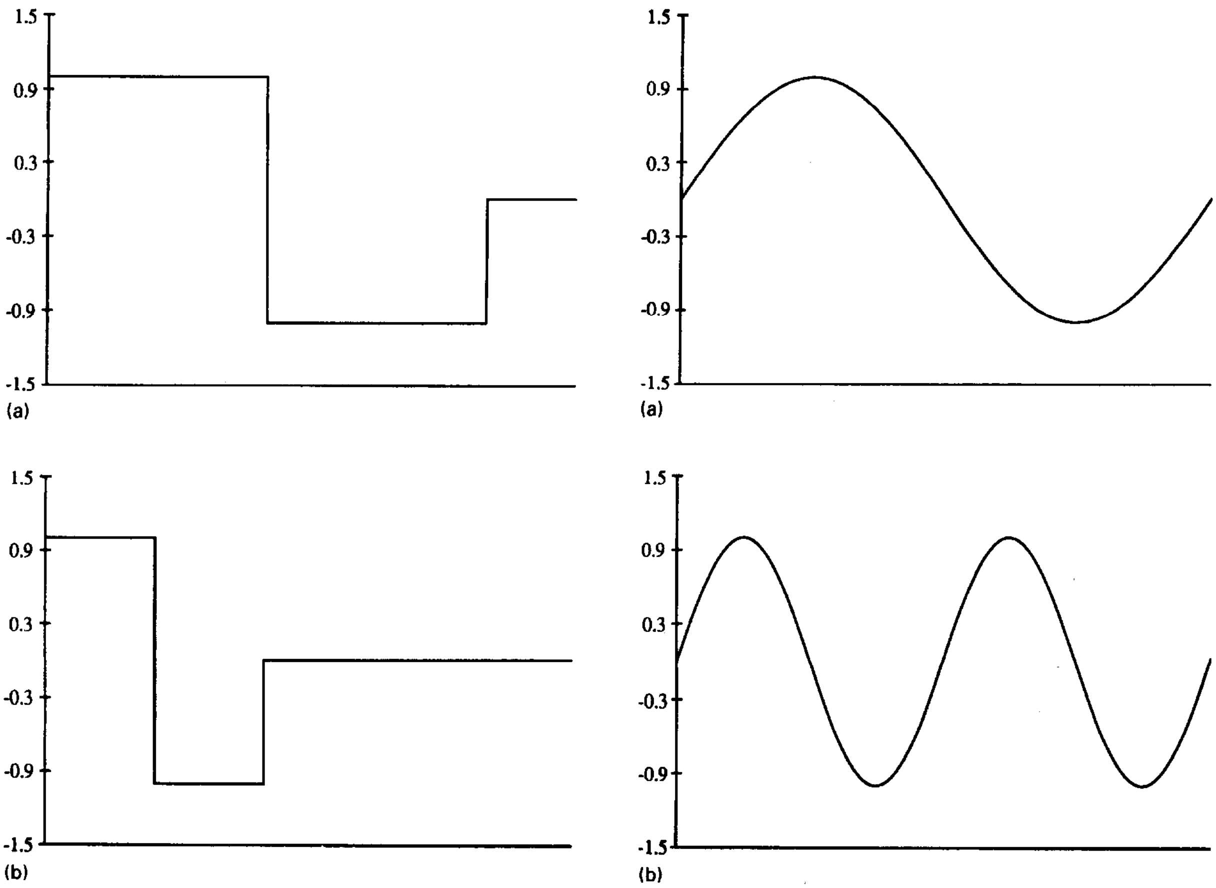

Fig. 3. A comparison of Haar wavelets and Fourier functions: (a) $\Psi_{0,0}(x)$ and $\sin(x)$; (b) $\Psi_{1,0}(x)$ and $\sin(2x)$; (c) $\Psi_{1,1}(x)$ and $\cos(2x)$; (d) $\Psi_{2,0}(x)$ and $\sin(3x)$; (e) $\Psi_{2,1}(x)$ and $\cos(3x)$.

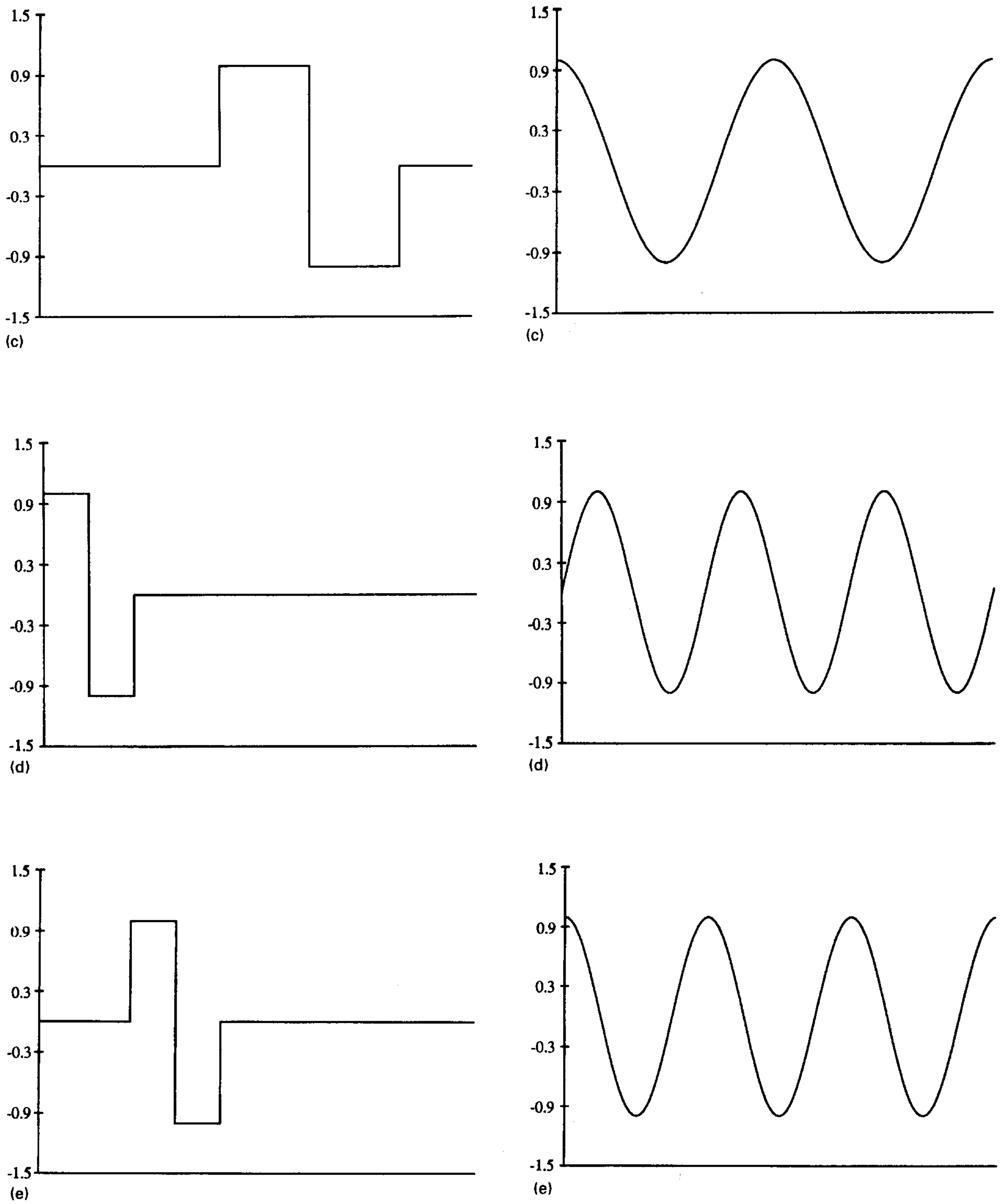

Fig. 3(c)–(e).

For sampled or discrete signals, a third form of the wavelet transform exists. The discrete wavelet pyramid transform is the analog of the discrete Fourier transform. For band-limited signals that are sampled at greater than the Nyquist rate, and for orthogonal wavelet families, the function's sample points may be reconstructed, and subsequently interpolated to regenerate the original signal. Because of the logarithmic structure of dyadic wavelets, there are numerically efficient algorithms for decomposing a sampled signal via the discrete wavelet transform; see [23].

Wavelet constructions can become considerably more complex than those presented thus far. Mallat [17] introduces a technique, known as multiresolution analysis, which allows the construction of complete bases of orthogonal wavelets. By the use of dilates by powers of constants other than two, nonorthogonal bases of wavelets may be constructed [9, 18]. Daubechies [8, Ch. 3] gives an excellent explanation of "frames", the abstraction through which one may obtain a complete set of nonorthogonal vectors to span a space.

Orthogonal, continuous, discontinuous, symmetric and antisymmetric wavelets can be constructed with known regularity and other properties. The properties of the basic wavelet and the size of the dilations and translations completely specify the nature of any subsequent time–frequency analysis. This has led many analysts to specify properties a priori and then construct a basic wavelet subject to certain corresponding constraints. If they exist, the resulting basic wavelets may be defined strictly by a recursion having no simple closed form. These wavelets are often unusual in form; see, for example, [7]. Further, wavelets may be orthogonal with compact support and arbitrarily high regularity [7]. Strang [22, p. 618] presents a number of properties for orthogonal wavelets with compact support and known regularity. They may be nonorthogonal but form complete bases for $L^2(\mathbb{R})$ [9]. They may also have discontinuities [15]. It is in fact the Haar wavelets which have some of the most interesting properties for statistical applications.

A related family of functions is the Walsh functions; see [21].

3. Andrews' plots

Andrews [2] proposed a method for plotting high-dimensional data in two dimensions. If the data are m-dimensional, each point $\boldsymbol{x} = (x_1, \ldots, x_m)$, where x_i $(i = 1, \ldots, m)$ are the measured variables, is represented by the function

$$f_{\boldsymbol{x}}(t) = x_1 2^{-1/2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \cdots \tag{6}$$

plotted over the range $-\pi < t < \pi$. Andrews gave several properties of these functions, namely (6) preserves means, distances and variances, and gives one-dimensional projections. The orthogonality of the Fourier functions in (6) implies that plots of the functions that are close together in the L^2-metric imply that the corresponding points of the data are close together in the Euclidean metric. When (6) is plotted for each point of the data, $\boldsymbol{x}$, clustering of points will be seen by a banding together of the plots of the functions. Further, the plots enable one to see if points are close together in some projection of the m-measured variables [11]. Although many graphical techniques for multivariate data analysis exist, one advantage of Andrews' plotting technique is that its properties are based on mathematical theory and, therefore, the visual interpretation is not as subjective as some other methods [5]. As an example consider the well-known iris data; see

[12, 3]. The data consist of measurements of the sepal length, x_1, and width, x_2, and the petal length, x_3, and width, x_4, of 50 irises of three species, *Iris setosa, Iris versicolor* and *Iris virginica.* Fig. 4 shows Andrews plots for the iris data with Fourier functions as in (6).

Consider again the Haar wavelet pictured in Fig. 2. Analogously to (6), one can construct an Andrews plot with the Haar functions rather than the Fourier functions as

$$f_x(t) = x_1 + x_2 \Psi_{0,0}(t) + x_3 \Psi_{1,0}(t) + x_4 \Psi_{1,1}(t) + \cdots \tag{7}$$

with $0 \leqslant t \leqslant 1$, the coefficient of x_1 having been chosen to satisfy orthogonality constraints. The plots can also be formed with other families of wavelets, but the Haar wavelet has many desirable properties because of its piecewise constant nature.

It is important to note here that the functional representation given in (7) does not necessarily yield the near-independence-of-t property of the variance of $f_x(t)$ for uncorrelated data; see [11, Eq. (2)]. Intervals of the t-axis may be contained in the support of different numbers of wavelet functions. In this case the values of $f_x(t)$ on different intervals are influenced by different numbers of wavelet functions, and hence different numbers of variables. For example, in a plot arising from five-dimensional data, the interval $[0, \frac{1}{2})$ is influenced by three wavelet functions, and therefore by three variables, and the interval $[\frac{1}{2}, 1)$ is influenced by two wavelet functions. The idea of the stacking up of building blocks, Fig. 2, illustrates this comment. For a discussion of how this independence-of-t property can be used for constructing confidence intervals and statistical tests, see [2]. Besides a desirability of this property on these statistical grounds, it is not clear in general how the discriminating power of an Andrews-like plotting technique may depend on it.

There are a number of possible ways to obtain the independence-of-t property. One method is to use a selection of translated and dilated wavelets such that for all t the number of wavelets which are non-zero at t is constant. In this case, the variance of the plot does not depend on t; however, there may be only a few linear combinations of the variables plotted. Many permutations of the variables are then necessary to see sufficient numbers of linear combinations.

Another approach involves creating a new function γ_r which is the normalization of the sum of all the wavelets of resolution r, i.e.

$$\gamma_r(t) = 2^{-r/2}\{\Psi_{r,0}(t) + \Psi_{r,1}(t) + \cdots + \Psi_{r,2r-1}(t)\}.$$

To each point, $\boldsymbol{x}$, associate the function

$$f_x(t) = x_1 + x_2\gamma_1(t) + x_3\gamma_2(t) + x_4\gamma_3(t) + \cdots \quad (0 \leqslant t \leqslant 1). \tag{8}$$

This method is more like Andrews original method because the functions have support over the whole range plotted. This method has two advantages in that it both preserves the independence of t and gives a large number of linear combinations with coefficients 1 and -1. Unfortunately, a linear combination coefficient cannot be 0 and the variance of $f_x(t)$ is considerably higher in this method than in that of the previous case. Here, the variance is $m\sigma^2$ versus a variance of $q\sigma^2$ for the method first suggested, where $q \leqslant 1 + s$ and s is largest integer p such that $2^p \leqslant m$.

Fig. 5 shows the Andrews plots with Haar wavelets, as in (7); Fig. 6 shows the Andrews plot with the functional representation given in (8). In Figs. 5 and 6, the Haar functions have been *tapered* so that the eye can more easily follow each line. Since all Haar wavelet functions of resolutions 1 and

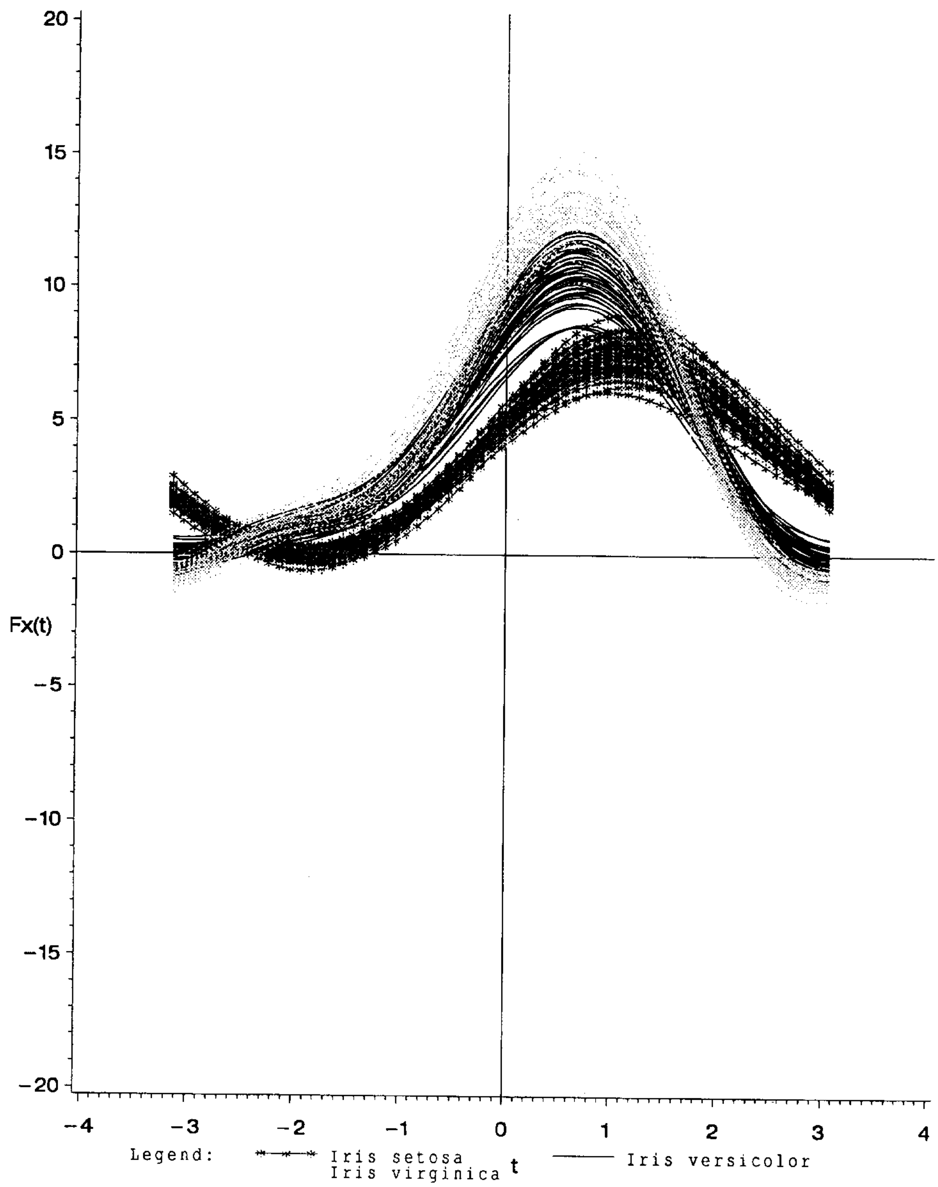

Fig. 4. Andrews' plots of the iris data using Fourier functions.

2 are used in Fig. 5, the variance of the plots is independent of t. In all three figures the plots for *Iris setosa* may be distinguished from those of *Iris versicolor* and *Iris virginica*. This clustering seems most evident in Figs. 4 and 6.

Wavelets can be used to construct an alternative functional representation to (6) for use in Andrews' plots. The Haar wavelet's piecewise constant nature may offer the researcher an

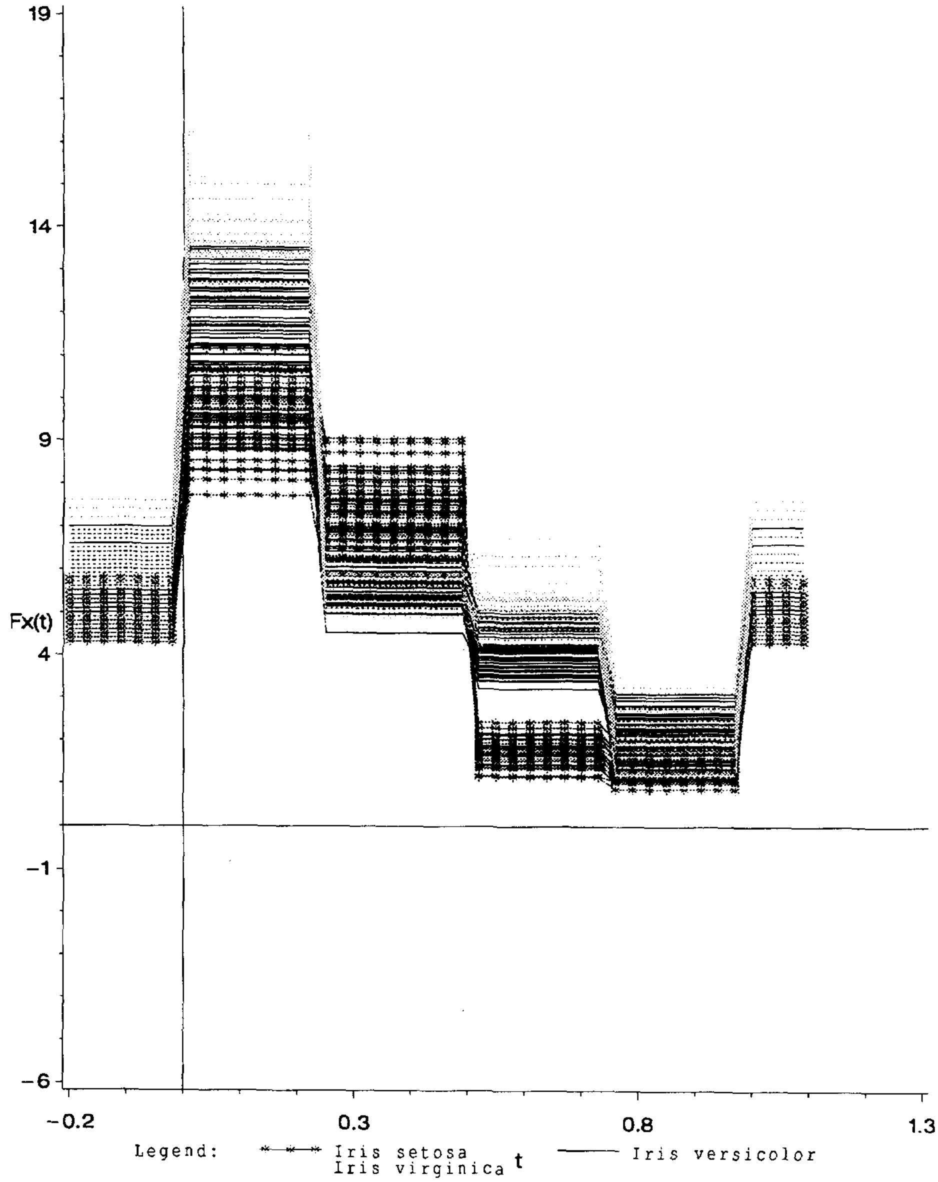

Fig. 5. Andrews' plots of the iris data using the Haar wavelet.

interesting additional way to discriminate graphically the clusters in multivariate data. It is well known that the eye can judge areas and distances better with straight lines than with curves. This is especially important when a large number of variables are included in the Andrews plots. Of the Haar wavelet representations, that proposed in (6) may be best.

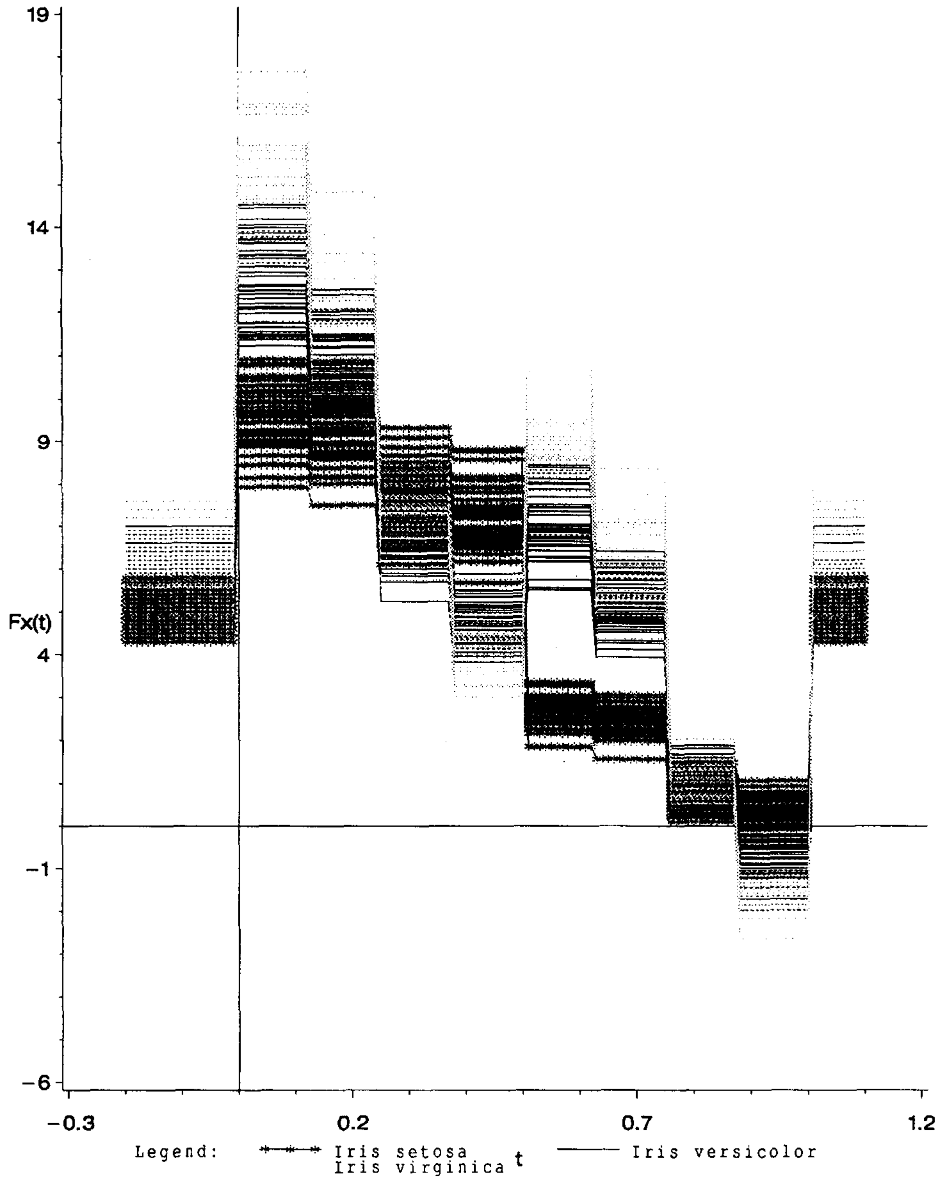

Fig. 6. Andrews' plots of the iris data with uniform variance.

4. Discussion

Historically, Haar [15] was probably the first to construct a wavelet set. The form of the modern dyadic wavelet, however, was first used by Strömberg [24] in the development of a modified Franklin system. The term wavelet itself may stem from its use by Grossman and Morlet [14] and others since the modern study of wavelets arises from work with splines in seismic analysis. There are also known applications in quantum field theory, pure mathematics and in data compression and the analysis of signals; see the introduction in [19]. More recent applications may be found in [20].

The properties of the basic wavelet and the size of the dilations and translations completely specify the nature of any subsequent time–frequency analysis. Because wavelets provide localization simultaneously in both time and frequency, wavelets and the wavelet transform are useful tools in signal analysis [6]. Wavelets also have graphical properties which make them useful for the pictorial representation of information, such as in Andrews' plots or in data compression and reconstruction algorithms. Here it has been shown that wavelets are a useful addition to the arsenal of pictorial representations of data. Further study is needed on how the analytic properties of more general wavelets may imply discriminating power in the Andrews plots based on them. For a discussion of some of the mathematical questions involved, see [11].

Acknowledgements

The authors would like to thank Professor D.F. Andrews for his suggestion regarding the pictorial representation of wavelets in Andrews' plots and Mr. M. Fleming who did some of the graphics. The authors are also grateful for the remarks of a referee. This paper was supported by grants from Employment Canada, the Natural Sciences and Engineering Research Council of Canada and the FKFO, Grant No. 2.0078.89.

References

[1] T.W. Anderson, *The Statistical Analysis of Time Series* (Wiley, New York, 1971).
[2] D.F. Andrews, Plots in high-dimensional data, *Biometrics* **28** (1972) 125–136.
[3] D.F. Andrews and A.M. Herzberg, *Data: A Collection of Problems in Many Fields for the Student and Research Worker* (Springer, Berlin, 1985).
[4] E.D. Brigham, *The Fast Fourier Transform* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[5] H. Chernoff, The use of faces to represent points in k-dimensional space graphically, *J. Amer. Statist. Assoc.* **68** (1973) 361–368.
[6] C.K. Chui, Ed., *Wavelets: A Tutorial in Theory and Applications* (Academic Press, San Diego, 1992).
[7] I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **41** (1988) 909–996.
[8] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).
[9] I. Daubechies, A. Grossman and Y. Meyer, Painless nonorthogonal expansions, *J. Math. Phys.* **27** (1986) 1271–1283.
[10] R.A. DeVore and B. J. Lucie, Wavelets, in: *Acta Numerica* 1992 (Cambridge Univ. Press, Cambridge, 1991) 1–56.
[11] P. Embrechts and A.M. Herzberg, Variations of Andrews' plots, *Internat. Statist. Rev.* **59** (1991) 175–194.
[12] R.A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. Eugenics* **7** (II) (1936) 179–188.
[13] P. Franklin, *A Set of Continuous Orthogonal Functions*, Mathematische Annalen, No. 100 (Springer, Berlin, 1928).
[14] A. Grossman and J. Morlet, Decomposition of Hardy functions with square integrable wavelets of constant shape, *SIAM J. Math. Ann.* **15** (1984) 723–736.
[15] A. Haar, Orthogonale Funktionensystem, *Math. Ann.* **69** (1910) 337–371.
[16] C.E. Heil and D.F. Walnut, Continuous and discrete wavelet transforms, *SIAM Rev.* **31** (1989) 628-666.
[17] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Machine Intelligence* **11** (1989) 674–693.

[18] Y. Meyer, Ed., *Wavelets and Applications* (Springer, Paris, 1989).
[19] Y. Meyer, *Ondelettes et Operateurs* (Hermann, Paris, 1990).
[20] M.B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer and L. Raphael, Eds., *Wavelets and their Applications* (Jones and Bartlett, Boston, MA, 1992).
[21] D.S. Stoffer, Walsh-Fourier analysis and its statistical applications, *J. Amer. Statist. Assoc.* **86** (1991) 461–485.
[22] G. Strang, Wavelets and dilation equations: a brief introduction, *SIAM Rev.* **31** (1989) 614–662.
[23] G. Strang, Wavelet transforms versus Fourier transforms, *Bull. (New Series) Amer. Math. Soc.* **31** (1993) 288–305.
[24] J.O. Strömberg, A modified Franklin system and higher-order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces, in: W. Beckner, A.P. Calderón, R. Fefferman and P.W. Jones, Eds., *Conference on Harmonic Analysis in Honor of Antoni Zygmund Volume II* (Wadsworth, Belmont, CA, 1982) 475–494.