International Statistical Review (1991), **59**, 2, pp. 175–194. Printed in Great Britain

Variations of Andrews' Plots

Paul Embrechts[1] **and Agnes M. Herzberg**[2]

[1] *ETH Zürich, Switzerland.* [2] *Department of Mathematics and Statistics, Queen's University, Kingston, Canada K7L 3N6, and Department of Mathematics, Imperial College of Science and Technology, London SW7 2BZ, U.K.*

Summary

Andrews (1972) introduced a method of plotting high-dimensional data in two dimensions by using the values of the variables as coefficients in a Fourier function. In this paper, the influence of re-ordering the variables, re-scaling the variables and the choice of an alternative orthogonal basis, i.e. Legendre and Chebychev, are explored; some examples are given.

Key words: Andrews' plots; Chebychev polynomials; Exploratory data analysis; Graphical techniques; Iris data; Legendre polynomials.

1 Introduction

If statistical data are m-dimensional, then each set of m measurements can be presented as an m-dimensional point. Several authors have developed graphical techniques to plot high-dimensional data in two dimensions in order to visually cluster the data; see, for example, Andrews (1972); Chernoff (1973); Kleiner & Hartigan (1981). For a good review on possible techniques and further references, see Chambers et al. (1983) and Fisher (1987). More mathematical techniques have been given by Beale (1969); Banfield & Bassil (1977) and more recently projection pursuit methods by Friedman & Tukey (1974) and Huber (1985). Because of their clear mathematical properties, Andrews' plots offer an objective representation in two dimensions of multivariate data.

Variations of Andrews' plots will be discussed and illustrated via examples. These variations include projections of the original plots, re-scaling of the data, re-ordering of the data, other orthogonal bases for the plots rather than sines and cosines (Fourier coefficients), in particular Legendre polynomials and Chebychev polynomials. The Iris data figure prominently in the discussion because of the variety of ways they have already been analyzed in the literature; see, for example, Gnanadesikan (1977, pp. 207–). Based on this example and others, not all of which are included in this paper, a number of guidelines and strategies on how to use the Andrews' plots and their variations are given. The easy implementation should encourage the reader to build up further experience.

2 Andrews' Plots

Andrews (1972) proposed the following simple and useful method of plotting high-dimensional data in two dimensions. If the data are m-dimensional, each point $\mathbf{x} = (x_1, \ldots, x_m)$, where x_i $(i = 1, \ldots, m)$ are the measured variables, is represented by the function

$$f_{\mathbf{x}}(t) = x_1 2^{-\frac{1}{2}} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \ldots \quad (1)$$

plotted over the range $-\pi < t < \pi$. The functions given by (1) have several properties. If

$\mathbf{x}_i = (x_{1i}, \ldots, x_{mi})$ $(i = 1, \ldots, n)$ are n points in m-dimensional space, then if $\bar{\mathbf{x}}$ stands for the mean vector,

$$f_{\bar{\mathbf{x}}}(t) = \frac{1}{n} \sum_{i=1}^{n} f_{\mathbf{x}_i}(t),$$

$$\|f_{\mathbf{x}_j} - f_{\mathbf{x}_l}\|^2_{L_2} = \int_{-\pi}^{\pi} \{f_{\mathbf{x}_j}(t) - f_{\mathbf{x}_l}(t)\}^2 \, dt = \pi \sum_{k=1}^{m} (x_{kj} - x_{kl})^2.$$

If the X_{ki} $(k = 1, \ldots, m; i = 1, \ldots, n)$ are uncorrelated random variables with common variance σ^2, then

$$\begin{aligned} \operatorname{var}\{f_{\mathbf{X}}(t)\} &= \sigma^2(2^{-1} + \sin^2 t + \cos^2 t + \sin^2 2t + \cos^2 2t + \ldots) \\ &= \begin{cases} 2^{-1}\sigma^2 m & (m \text{ odd}), \\ 2^{-1}\sigma^2\left\{m - 1 + 2\sin^2\left(\dfrac{mt}{2}\right)\right\} & (m \text{ even}). \end{cases} \end{aligned} \tag{2}$$

Thus

$$2^{-1}\sigma^2(m-1) \leq \operatorname{var}\{f_{\mathbf{X}}(t)\} \leq 2^{-1}\sigma^2(m+1) \quad (-\pi < t < \pi). \tag{3}$$

Therefore, overall tests and tests of significance may be performed; see Andrews (1972).

Thus Andrews' plots preserve means, distances and variances, and by virtue of the definition yield infinitely many one-dimensional projections on the vectors $(2^{-\frac{1}{2}}, \sin t, \cos t, \ldots)$ $(-\pi < t < \pi)$. Since the Andrews' plot representation is distance preserving up to a constant, plots exhibiting functions that are close together imply that the corresponding data points are close together. For a good overall discussion of the use of Andrews' plots, see Gnanadesikan (1977), Everitt (1987), Jolliffe (1986, pp. 89–91) and Jolliffe, Jones & Morgan (1980).

3 Variations

3.1 Projections

Several possibilities present themselves for the case of projections. Here only the presence or absence of a co-ordinate of $\mathbf{x}$ will be illustrated. More generally, various orthogonal projections on co-ordinate planes and indeed on arbitrary subspaces could be considered. The resulting equations involve combinations of $x_j = 0$ $(j = 1, \ldots, m)$ for some j; thus the Andrews' plot of the projected data is obtained from (1) by putting the relevant x_j's equal to zero. If, for example, the separation between the plots is hardly affected by putting certain $x_j = 0$ in (1), then this would imply that that co-ordinate has 'low discriminating power'. If, on the other hand, well-separated curves of the complete m-dimensional data collapse together if one projects on, for example, $x_j = 0$, then x_j carries a lot of information for the data. This technique could be considered, from an exploratory data analysis point of view, to be like all-subset regression.

3.2 Re-scaling and Re-ordering

The function $f_{\mathbf{x}}(t)$ is the sum of many functions. Thus, consistently large co-ordinates in $\mathbf{x}$ may hide visually the effect of other co-ordinates in $f_{\mathbf{x}}(t)$. Thus, it is sometimes advisable to re-scale the data. This is done as follows.

Let

$$\mathbf{x}_i = (x_{1i}, \ldots, x_{mi}) \quad (i = 1, \ldots, n)$$

be a set of n points in m dimensions; let

$$\bar{x}_{j.}=\frac{1}{n}\sum_{i=1}^{n}x_{ji},\quad s_j^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_{ji}-\bar{x}_{j.})^2,$$

where $\bar{x}_{j.}$ and s_j^2 are the sample mean and variance of the jth variable, respectively $(j=1,\ldots,m)$. Then, let the re-scaled points be

$$\mathbf{y}_i=(y_{1i},\ldots,y_{mi})\quad(i=1,\ldots,n),$$

where

$$y_{ji}=\frac{x_{ji}-\bar{x}_{j.}}{s_j}\quad(j=1,\ldots,m).$$

The squared Euclidean distance between the points $\mathbf{x}_j$ and $\mathbf{x}_l$ is

$$\|\mathbf{x}_j-\mathbf{x}_l\|^2=\sum_{k=1}^{m}(x_{kj}-x_{kl})^2$$

and that between the corresponding transformed points is

$$\|\mathbf{y}_j-\mathbf{y}_l\|^2=\sum_{k=1}^{m}\frac{(x_{kj}-x_{kl})^2}{s_k^2}.$$

Let $\mathbf{y}$ be the re-scaled point corresponding to $\mathbf{x}$. Then, distances on the y-scale will be preserved for the corresponding Andrews' plots. Moreover, for uncorrelated normally distributed data,

$$\operatorname{var}\{f_{\mathbf{Y}}(t)\}=\begin{cases}\dfrac{(n-1)^2m}{2n(n-3)} & (m\text{ odd}),\\[2ex] \dfrac{(n-1)^2}{2n(n-3)}\left\{m-1+2\sin^2\left(\dfrac{mt}{2}\right)\right\} & (m\text{ even}).\end{cases}\tag{4}$$

A proof of (4) is given in the Appendix.

In §§ 4.2 and 4.3, two examples of data before re-scaling and after will be given. It is shown that the re-scaled results can make quite a difference.

It will also follow that the Andrews' plots suffer from strong dependence on the order of the variables in (1); see § 4.4. As a rule of thumb, one would suggest grouping strongly correlated variables together and putting strongly discriminating variables at the extreme frequencies of sines and cosines. These points will be discussed further in the examples in § 4.4.

3.3 *Other Orthogonal Functions*

The properties of the Andrews' plots mentioned in § 2 hold for every basis of orthogonal functions, not just the Fourier functions. Therefore, if

$$\Psi=\{\psi_k\mid\psi_k:\mathbb{D}\subset\mathbb{R}\to\mathbb{R}\quad(k=1,2,3,\ldots)\},$$

where $\mathbb{D}$ is some suitable domain, denotes a set of L^2-orthogonal functions, then for every $\mathbf{x}\in\mathbb{R}^m$,

$$f_{\mathbf{x}}^{\psi}(t)=\sum_{k=1}^{m}x_k\psi_k(t)$$

may give a useful two-dimensional representation $(t,f_{\mathbf{x}}^{\psi}(t))_{t\in\mathbb{D}}$. The main problem related to these more general orthogonal representations stems from the fact that the 'variance function' $\sum \psi_k^2(t)$, where the sum is over $k = 1, \ldots, m$, can be hard to calculate and moreover lacks the 'near independence of t' property exhibited in (2) and (3) for the Fourier function; see also Nevai (1979, p. 85). For data with uncorrelated variables, admittedly a rather uninteresting case, this confounds the interpretation of the plot, as measurement of significant width changes over the plot; see for instance Goodchild & Vijayan (1974) for a related discussion.

For illustrative reasons, the basic definitions and properties for the Legendre and Chebychev polynomials are given.

Chebychev Polynomials. The Chebychev polynomial of degree n, $T_n(x)$, is defined by

$$T_n(x) = \cos n\theta, \quad x = \cos\theta \quad (0 \leq \theta \leq \pi).$$

It is well known that

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_n(x) &= 2xT_{n-1}(x) - T_{n-2}(x) \quad (n \geq 2), \end{aligned}$$

where $-1 \leq x \leq 1$. Thus the coefficients of the Chebychev polynomials may be calculated; see Table 1. For further details on Chebychev polynomials, see Spanier & Oldham (1987, pp. 193–).

For $\mathbf{x} = (x_1, \ldots, x_m)$, the Andrews' plot based on Chebychev polynomials is

$$f_{\mathbf{x}}(t) = (1 - t^2)^{-1/4}\left\{2^{-\frac{1}{2}}x_1 T_0(t) + \sum_{k=1}^{m-1} x_{k+1}T_k(t)\right\} \quad (-1 < t < 1). \tag{5}$$

If again $\mathbf{X} = (X_1, \ldots, X_m)$, with the X_i's uncorrelated random variables with common variance σ^2, then it can be shown that

$$\begin{aligned} \operatorname{var}\{f_{\mathbf{X}}(t)\} &= 2^{-1}\sigma^2(1 - t^2)^{-\frac{1}{2}}\left\{1 + 2\sum_{k=1}^{m-1} T_k^2(t)\right\} \\ &\leq 2^{-1}\sigma^2(1 - t^2)^{-\frac{1}{2}}(2m - 1). \end{aligned}$$

From a practical point of view, it makes sense to ignore the boundary effects ($t = \pm 1$) and

Table 1

Coefficients of the Chebychev polynomials

Degree of polynomial	Power of x 0	1	2	3	4	5	6	7	8	9	10
0	1										
1	0	1									
2	−1	0	2								
3	0	−3	0	4							
4	1	0	−8	0	8						
5	0	5	0	−20	0	16					
6	−1	0	18	0	−48	0	32				
7	0	−7	0	56	0	−112	0	64			
8	1	0	−32	0	160	0	−256	0	128		
9	0	9	0	−120	0	432	0	−570	0	256	
10	−1	0	50	0	−400	0	1120	0	−1280	0	512

Table 2
Coefficients of the Legendre polynomials

Degree of polynomial	Power of x 0	1	2	3	4	5
0	1					
1	0	1				
2	−0·5	0	1·5			
3	0	−1·5	0	2·5		
4	0·375	0	−3·75	0	4·375	
5	0	1·875	0	−8·75	0	7·875

to restrict to the interval $[-0{\cdot}8, 0{\cdot}8]$ say. In this interval, it can be shown that

$$\operatorname{var}\{f_{\mathbf{x}}(t)\} \leqslant \frac{5\sigma^2}{6}(-0{\cdot}377 + 1{\cdot}11m) \tag{6}$$

holds. From this and further estimates on the variance, approximations to the usual test procedures can be deduced. We shall not go into further details here.

Legendre Polynomials. The Legendre polynomial of degree n, $P_n(x)$, is defined by

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x, \\ P_n(x) &= \frac{2n-1}{n} x P_{n-1}(x) - \frac{n-1}{n} P_{n-2}(x) \quad (n \geqslant 2), \end{aligned} \tag{7}$$

where $-1 \leqslant x \leqslant 1$. Thus the coefficients of the Legendre polynomial may be calculated; see Table 2. For further details on Legendre polynomials, see Spanier & Oldham (1987, pp. 183–).

For $\mathbf{x} = (x, \ldots, x_m)$, the Andrews' plot based on Legendre polynomials is

$$f_{\mathbf{x}}(t) = 2^{-\frac{1}{2}}\{x_1 + 3^{\frac{1}{2}}x_2P_1(t) + \ldots + (2m-1)^{\frac{1}{2}}x_mP_{m-1}(t)\} \quad (-1 \leqslant t \leqslant 1). \tag{8}$$

All properties, except the variance estimates, of the original Andrews' plots are retained. For uncorrelated data with constant variance σ^2, one can show that

$$\begin{aligned} \operatorname{var}\{f_{\mathbf{x}}(t)\} &= 2^{-1}\sigma^2\left\{\sum_{i=1}^{m}(2i-1)P_{i-1}^2(t)\right\} \\ &\leqslant 2^{-1}\sigma^2(1{\cdot}28 - 0{\cdot}208m + 0{\cdot}175m^2) \end{aligned}$$

if t is restricted to a suitable subinterval of $[-1, 1]$. Again we shall not go into further details here.

4 Examples

4.1 Some Preliminary Remarks

The best way to appreciate the usefulness of a graphical technique is to show its versatility by way of examples. This section concentrates on: re-ordering of the variables, re-scaling of the variables, co-ordinate projections and the use of other orthogonal bases for some specific sets of data.

As already explained in § 1, we shall concentrate on the Iris data in order to be able to compare and contrast more easily the various alternative ways for exploiting the proposed techniques.

4.2 Iris Data; A First Look

The well-known set of Iris data contains measurements of the sepal length, x_1, and width, x_2, and the petal length, x_3, and width, x_4, of 50 irises from each of three species *Iris setosa, Iris versicolor* and *Iris virginica*. The set of data has been used considerably in the discussion of various clustering techniques, the first being that of Fisher (1936). For a full listing of the data, see Andrews & Herzberg (1985, p. 8).

Figure 1 shows the Andrews' plots based on trigonometric functions. The plots separate clearly the *Iris setosa* from the other two classes. Various other characteristics show up. For instance, the curves are very tightly banded near $t_0 = -2{\cdot}5$. This implies that, in a direction orthogonal to the vector

$$\left(\frac{1}{\sqrt{2}}, \sin t_0, \cos t_0, \sin 2t_0\right),$$

the data-cloud in $\mathbb{R}^4$ looks rather 'flat' so that a dimension reduction from four to three is likely. A canonical variate analysis would confirm this. The same characteristic is shown in Fig. 5. As would be expected, it is hard to separate *Iris virginica* from *Iris versicolor*

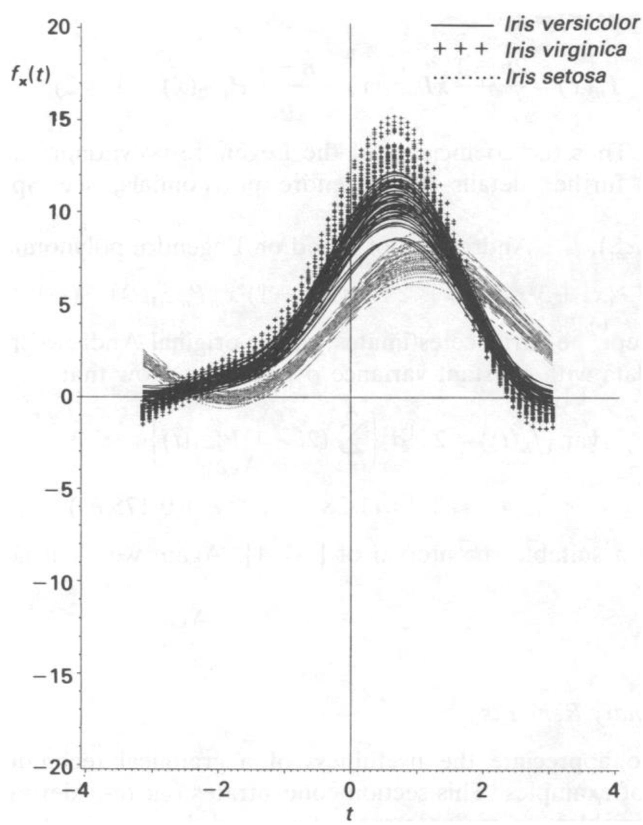

Figure 1. *Andrews' plots for the iris data based on trigonometric functions.*

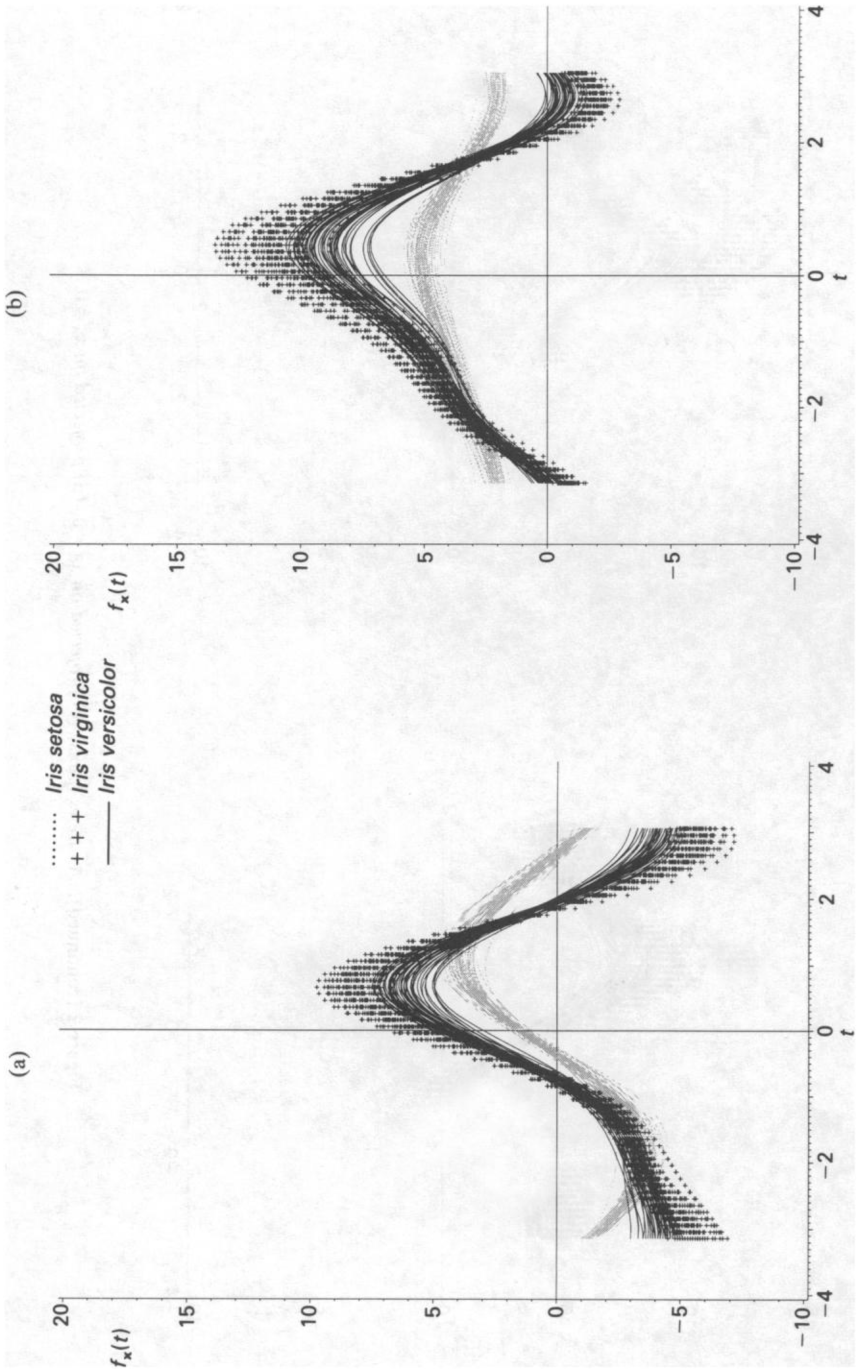

Figure 2. *Andrews' plots for the iris data with projections,* (a) *projected on* $x_1 = 0$, (b) *projected on* $x_2 = 0$.

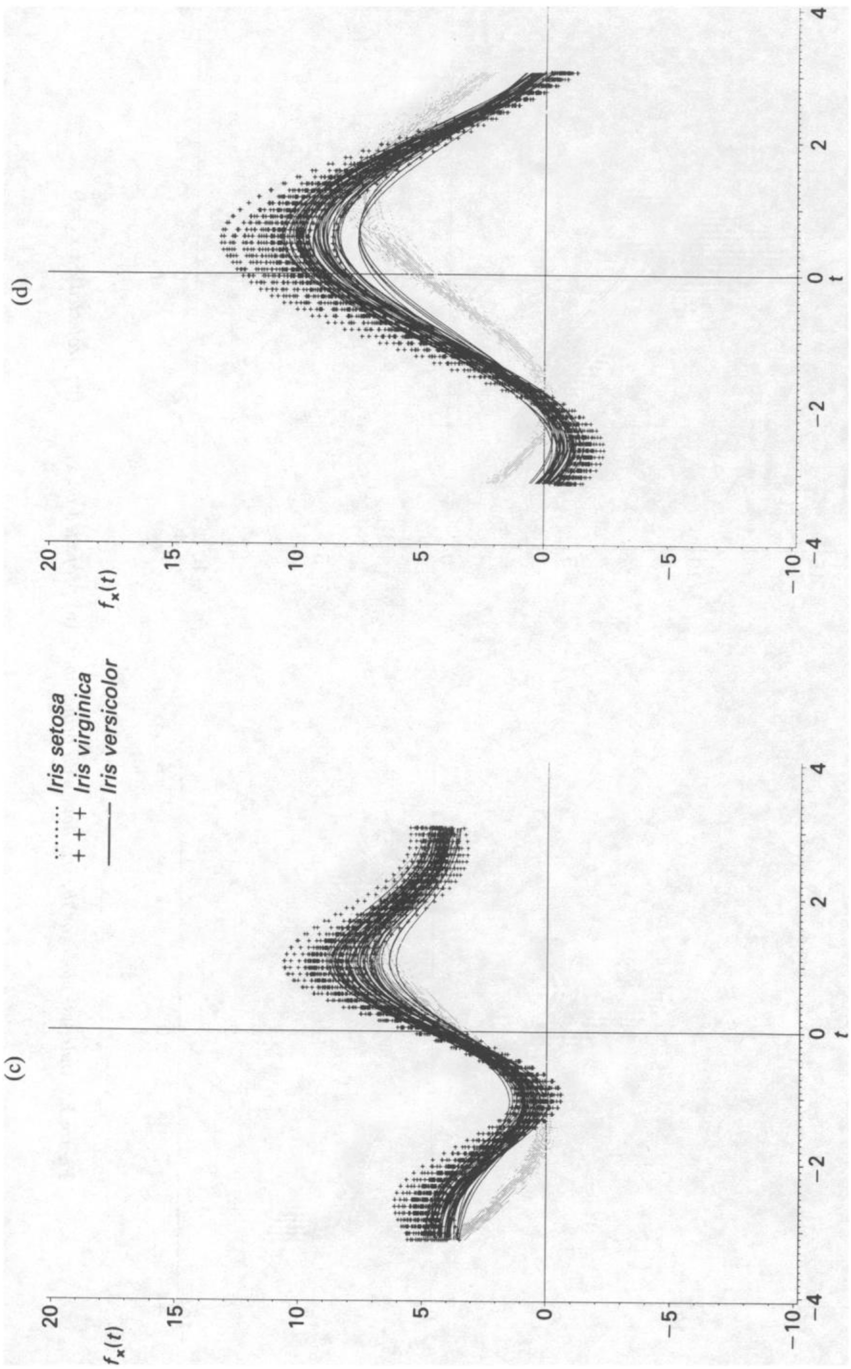

Figure 2 (continued). *Andrews' plots,* (c) *projected on* $x_3 = 0$, (d) *projected on* $x_4 = 0$.

although at some intervals for t, separation is clear; take for instance the interval $[-2{\cdot}5, -1{\cdot}5]$, in Fig. 1, where both groups exhibit rather different curvature in the plots. An interactively constructed plot would reveal this pattern much more clearly. Also note some potential outlying curves in the cloud of the *Iris versicolor*. This pattern shows up more clearly in other plots; see for instance Figs. 4(c) and 7(a).

Figures 2 and 3 illustrate the projection method given in § 3.1, using trigonometric functions. Figure 2 shows what happens when the data are projected onto the subspaces, $x_k = 0$ ($k = 1, 2, 3, 4$). In the subspace, $x_3 = 0$, Fig. 2(c), all three species are rather close together over almost all of the whole range of t. It follows that x_3, which corresponds to the petal length, is an important discriminating characteristic of the irises. On the other hand, x_1 (sepal length) and x_2 (sepal width) carry little discriminating 'power'; as can be seen from Figs. 2(a) and 2(b) as these plots are very similar to Fig. 1.

Figure 3 shows a further projection, namely, $x_3 = x_4 = 0$; this stresses further that the measurements made on sepals yield almost no discriminatory information. Another way to consider this would be to combine all two-dimensional scatter plots of the data in a so-called *draftsman's display* (Chambers et al., 1983, p. 136).

Figure 4 indicates the influence of re-scaling. In Fig. 4(a) the Andrews' plots for the re-scaled Iris data are given. The influence of re-scaling can clearly be seen. The clearest impression of the presence of three subgroups is obtained when plotting these graphs interactively. The presence of possible outliers can be seen from Fig. 4.

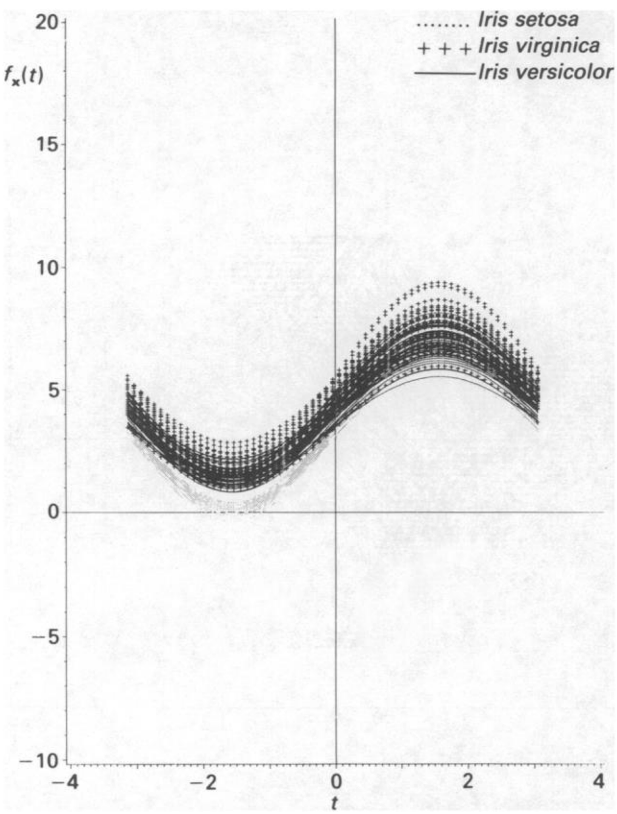

Figure 3. *Andrews' plots for the iris data projected on $x_3 = x_4 = 0$.*

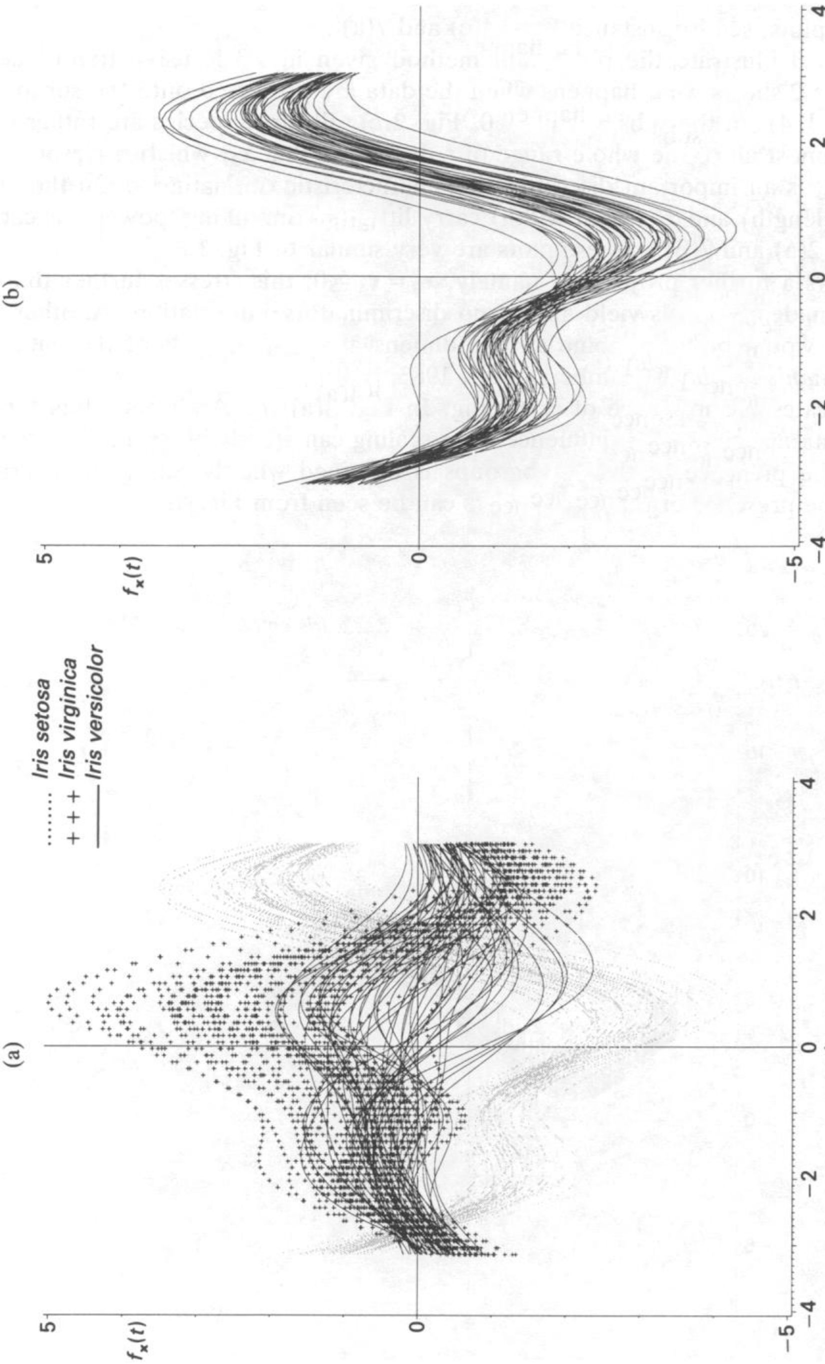

Figure 4. *Andrews' plots for the iris data,* (a) *for the re-scaled iris data,* (b) *for the re-scaled iris setosa data.*

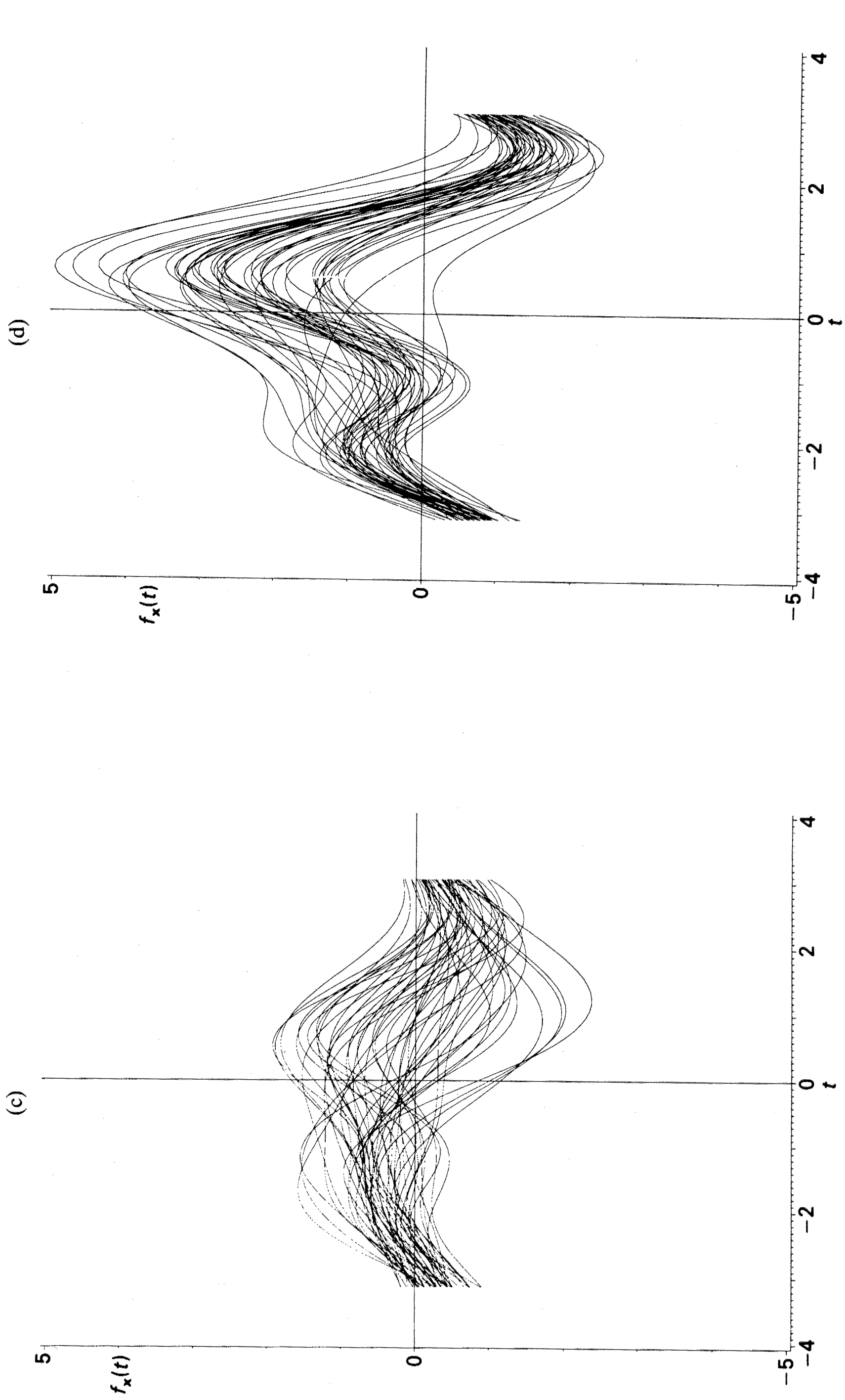

Figure 4 (continued). *Andrews' plots,* (c) *for the re-scaled iris versicolor data,* (d) *for the re-scaled iris virginica data.*

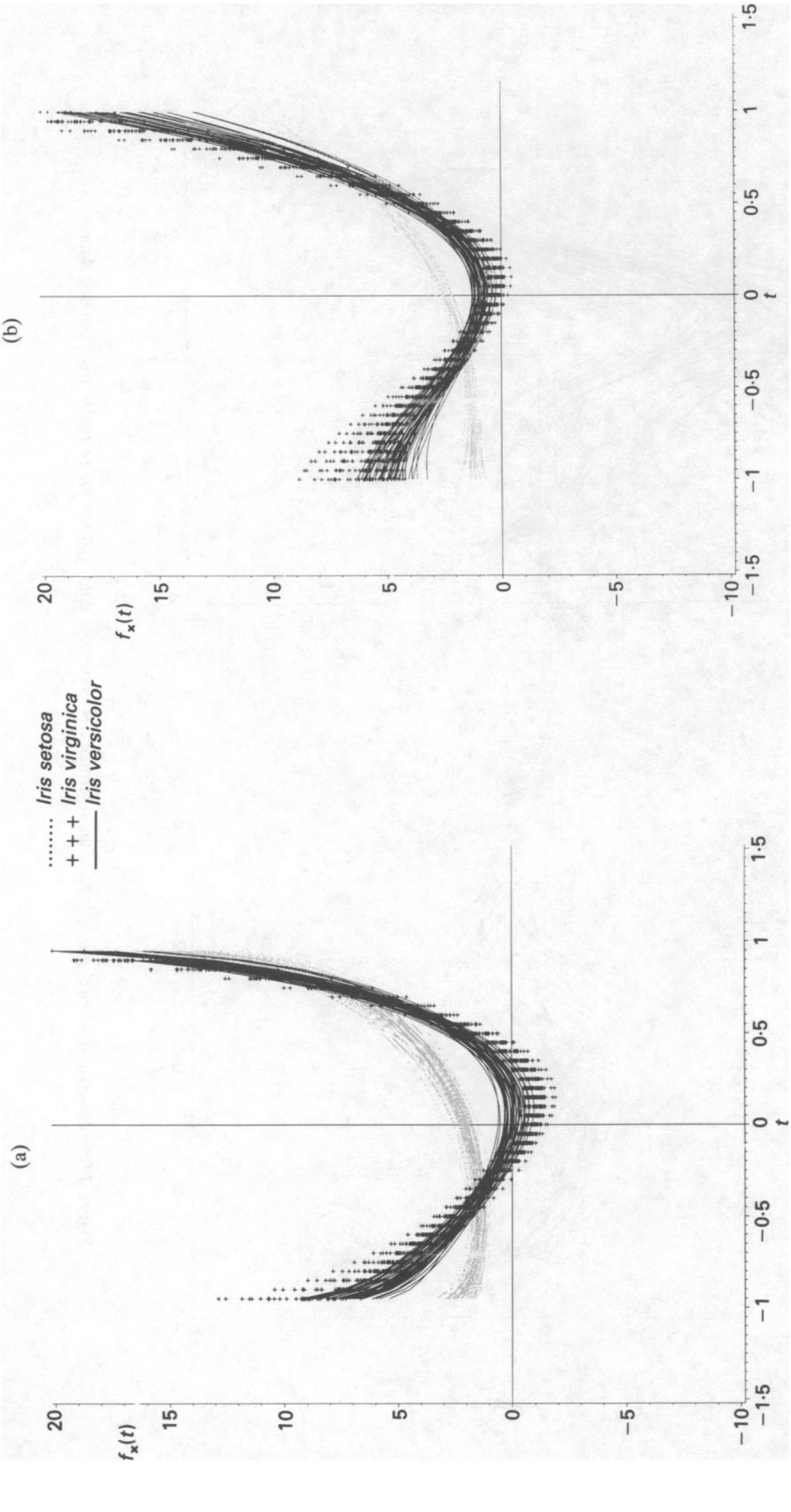

Figure 5. *Andrews' plots for the iris data,* (a) *based on Chebychev polynomials,* (b) *based on Legendre polynomials.*

Figure 5 shows the Andrews' plots for Chebychev polynomials and Legendre polynomials. Both the graphs again show that *Iris setosa* is well separated from the other two types of iris.

4.3 *Another Example: The Automobile Data*

Another striking example indicating the need for re-scaling is given in Fig. 6. These plots depict 66 observations from a 12-dimensional set of data on automobile models. For a detailed description of this well-known set of data, see for instance Chambers et al. (1983, pp. 352–355). Whereas the Andrews' plots for the original data (Fig. 6(a)) show little structure, it is obvious from Fig. 6(b) that a plot based on the re-scaled data clearly clusters the cars. In general terms, Fig. 6(b) separates the larger models from the smaller ones, together with a cluster of intermediate cars, larger and smaller for the purpose of illustration being based on the turning circle of the automobile in Fig. 6(c). Some potential outliers are also present. For these data, it is again important to judge the applicability of the Andrews' plotting technique *via* an interactive, sequential plotting device so that separate curves can be high-lighted more easily. A full analysis would of course include re-ordering, projections and highlighting potential group separation *via* other variables other than the turning circle; these graphs are not shown.

4.4 *Re-ordering*

Finally, in Fig. 7 the influence of re-ordering of the variables is illustrated. If Ψ is a general orthogonal family and the data $\mathbf{x} = (x_1, \ldots, x_m)$ is m-dimensional, then

$$f_{\mathbf{x}}^{\Psi}(t) = \sum_{k=1}^{m} x_k \psi_k(t)$$

corresponds to the projection from $\mathbf{x}$ onto $\{\psi_1(t), \ldots, \psi_m(t)\}$. Changing the order amounts to different projections so that a combination of various orders will always yield a richer overall-picture of the data. If one compares, for instance, Figs. 7(a) and 7(d), it is clearly seen how in the latter case putting the important factor x_3 at the lowest frequency essentially amounts to a shift in the curves; see also Fig. 7(b). Whereas in Fig. 7(a), x_3 influences the plots *via* the highest frequency and consequently gives a better visual separation over this range of projections. For a related discussion of this point and some further references, see Krzanowski (1988, p. 43).

5 Discussion

The Andrews' plot forms an important tool for exploratory data analysis. The mathematical definition of the Andrews' plot makes the interpretation of the results and the proposal of variants much easier than the interpretation for most existing techniques.

Their software implementation in an interactive environment is straightforward and the overlay of graphs, although this process becomes rather unwieldly for large sets of data, is a major advantage. As a teaching device in courses on exploratory data analysis or multivariate statistics, Andrews' plots form an ideal technique to start a general discussion on the topic of graphical representation of multi-dimensional observations.

From our experience, we would like to stress the following points.

(i) The ideal software and hardware environment has to allow for interactive, high-resolution, colour plotting.

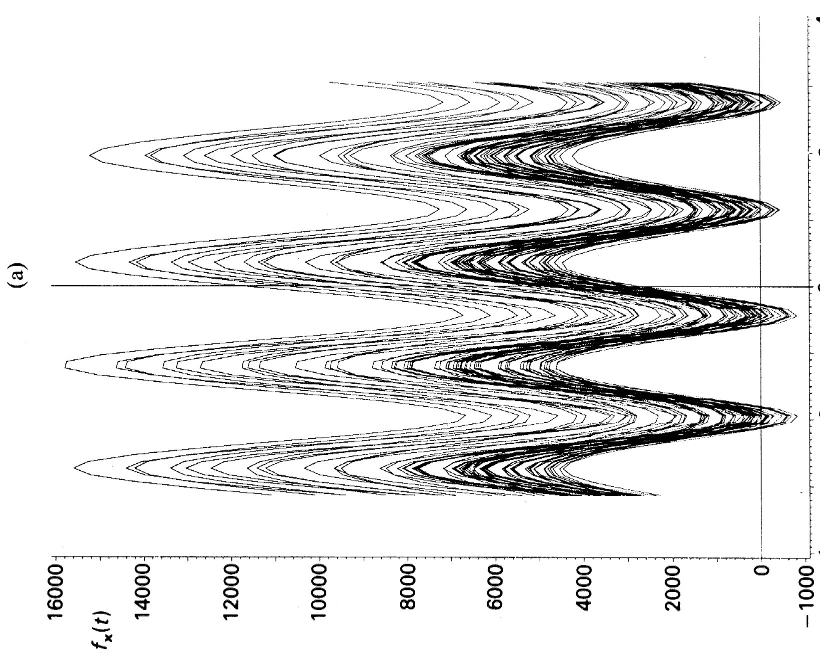

Figure 6. *Andrews' plots for the automobile data,* (a) *without scaling.*

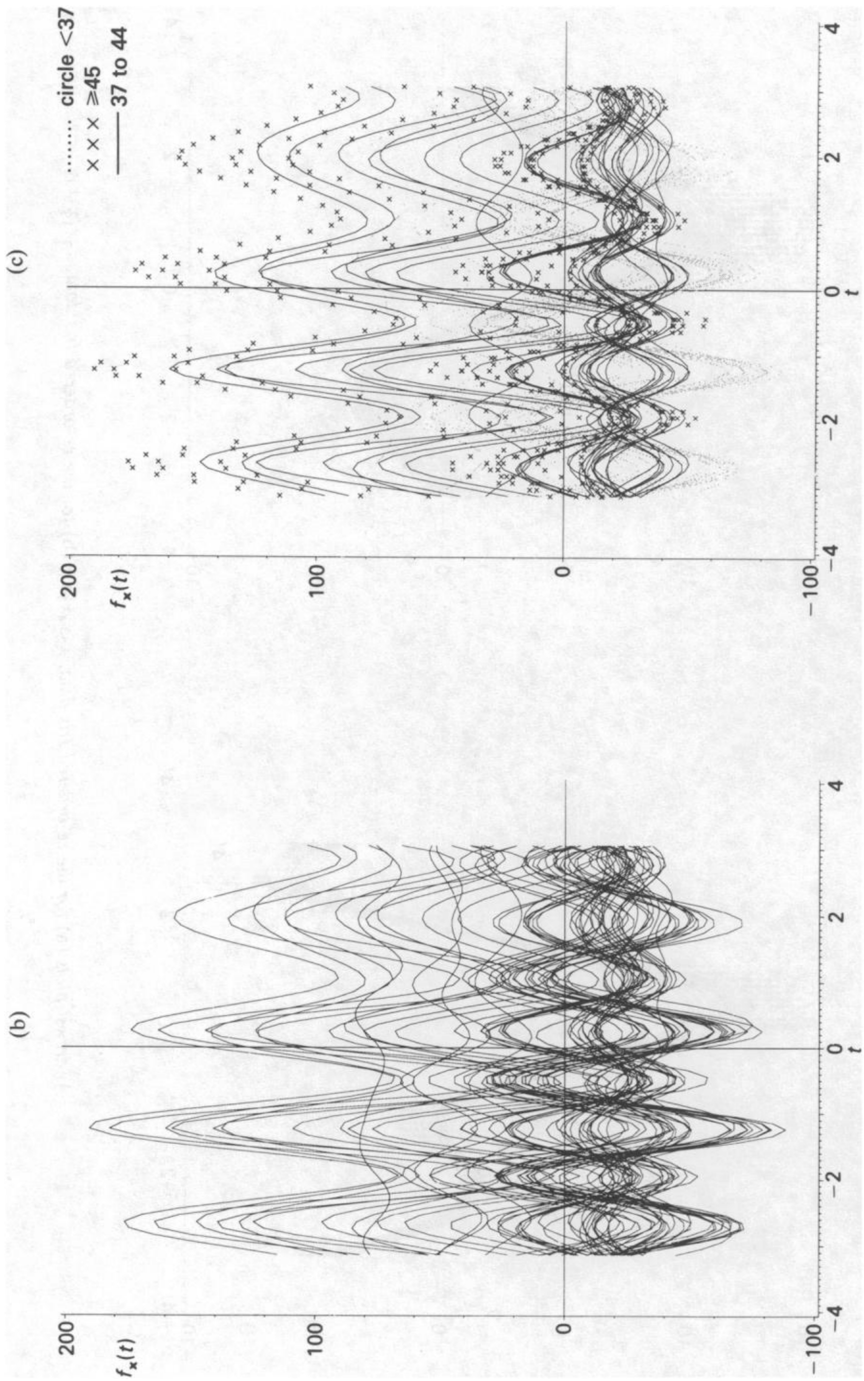

Figure 6 (continued). *Andrews' plots* (b) *for the re-scaled automobile data,* (c) *for the re-scaled automobile data grouped for turning circles.*

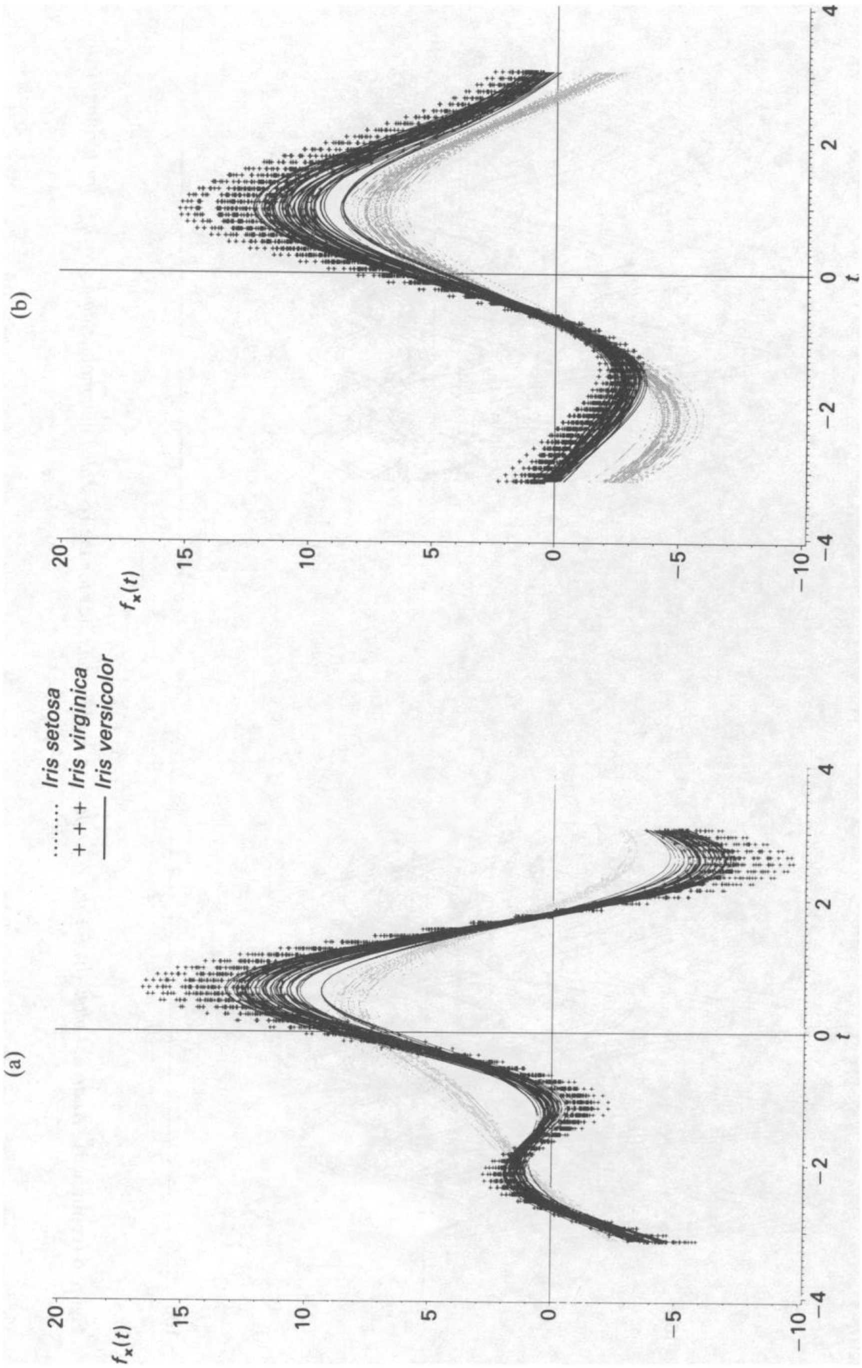

Figure 7. *Andrews' plots* (a) *for the re-ordered Iris data,* $x_2x_4x_1x_3$; (b) *for the re-ordered Iris data,* $x_3x_1x_2x_4$.

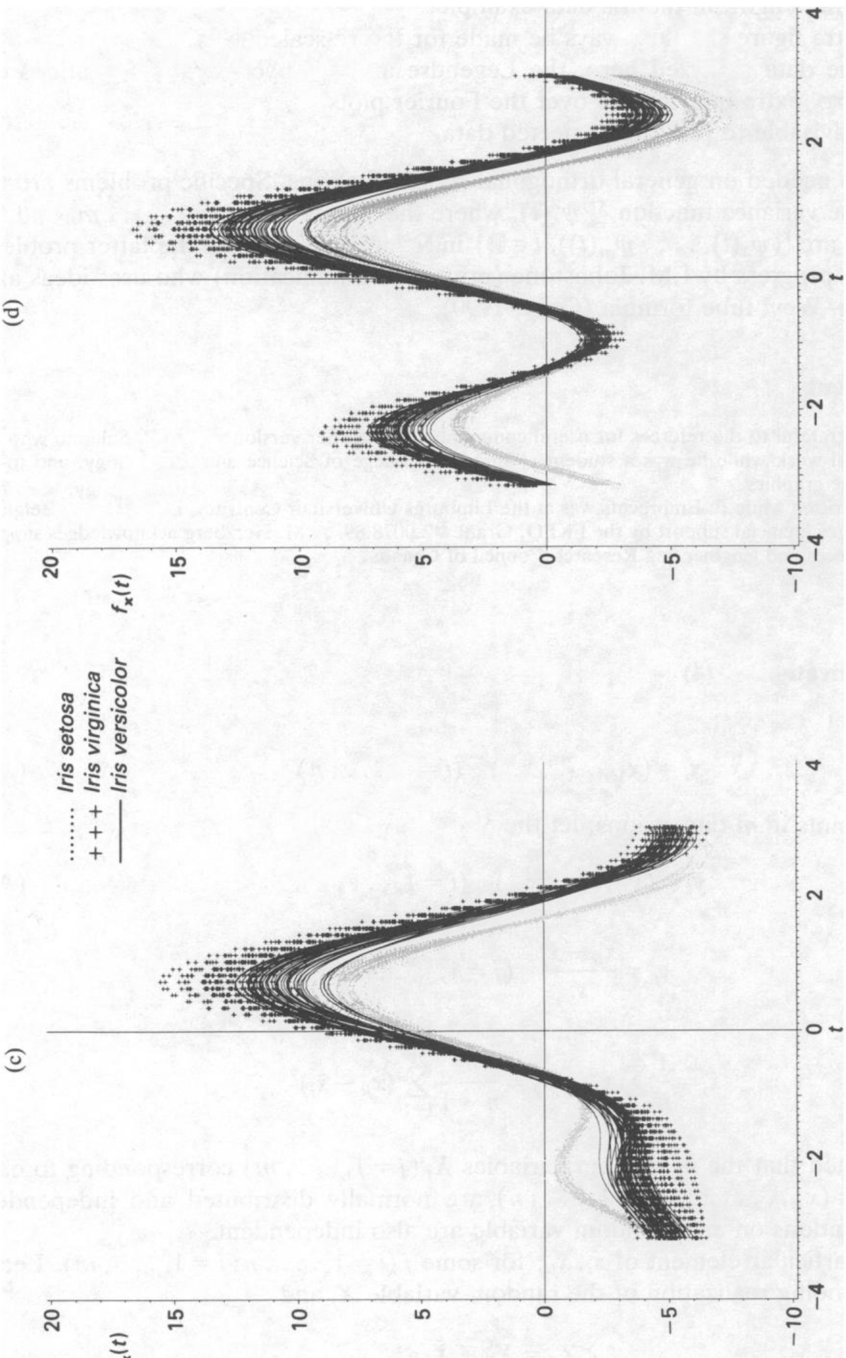

Figure 7 (continued). *Andrews' plots* (c) *for the re-ordered Iris data*, $x_4x_3x_1x_2$; (d) *for the re-ordered Iris data*, $x_3x_4x_2x_1$.

(ii) After a first inspection, if possible the variables should be re-ordered in order to put strongly correlated variables together. Variables which appear to discriminate groups should be put at the extreme frequencies in the Fourier case like x_3, the petal length, in the iris data example.
(iii) An extra figure should always be made for the re-scaled data.
(iv) For the data analyzed here, the Legendre and Chebychev representations did not carry extra information over the Fourier plots.
(v) It is advisable to plot the projected data.

More work is needed on general orthogonal representations. Specific problems are the estimation of the variance function $\sum \psi_k^2(t)$, where the sum is over $k = 1, \ldots, m$, and the coverage of the arc $\{(\psi_1(t), \ldots, \psi_m(t)), t \in \mathbb{D}\}$ in $\mathbb{R}^m$. With regard to the latter problem, some work is in progress by I.M. Johnstone (private communication) who uses ideas akin to the Hotelling–Weyl tube formula (Gray, 1990).

Acknowledgements

The authors are grateful to the referees for useful comments to an earlier version and to P. Salamin who did some of the original work while he was a student at Imperial College of Science and Technology, and to M. Fleming who did the graphics.

This paper was written while P. Embrechts was at the Limburgs Universitair Centrum, Diepenbeek, Belgium. He also acknowledges financial support by the FKFO, Grant #2.0078.89. A.M. Herzberg acknowledges support of the Natural Sciences and Engineering Research Council of Canada.

Appendix Derivation of (4)

As in § 3.2, let

$$\mathbf{x}_i = (x_{1i}, \ldots, x_{mi}) \quad (i = 1, \ldots, n) \tag{A1}$$

be a set of n points in m dimensions; let the re-scaled points be

$$\mathbf{y}_i = (y_{1i}, \ldots, y_{mi}) \quad (i = 1, \ldots, n), \tag{A2}$$

where

$$y_{ji} = \frac{x_{ji} - \bar{x}_j}{s_j} \quad (j = 1, \ldots, m).$$

Further,

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ji}, \quad s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ji} - \bar{x}_j)^2.$$

It will be assumed that the m random variables X_j $(j = 1, \ldots, m)$ corresponding to each element of $\mathbf{x}_i = (x_{1i}, \ldots, x_{mi})$ $(i = 1, \ldots, n)$ are normally distributed and independent and that observations on each random variable are also independent.

Consider a particular element of $\mathbf{x}_i$, x_{ji}, for some j $(i = 1, \ldots, n; j = 1, \ldots, m)$. Let x_{ji} be the corresponding realization of the random variable X_j and

$$Z_{ji} = X_{ji} - \bar{X}_j,$$

where

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^{n} X_{ji}.$$

Then it follows that

$$E(Z_{ji}) = 0, \quad \operatorname{var}(Z_{ji}) = \frac{n-1}{n}\sigma^2.$$

Since the X_{ji}'s $(j = 1, \ldots, m)$ are normally distributed, the Z_{ji}'s $(j = 1, \ldots, m)$ are also normally distributed with mean zero and variance $(n-1)n^{-1}\sigma^2$. Hence, $(n/(n-1))^{\frac{1}{2}}\sigma^{-1}Z_{ji}$ is distributed as a standard normal distribution with mean zero and variance 1. It follows that

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} Z_{ji}^2$$

is distributed as a χ^2-distribution with $n-1$ degrees of freedom. Now

$$\sum_{i=1}^{n} Z_{ji}^2 = (n-1)S_j^2 \quad (j = 1, \ldots, m),$$

where

$$S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_{ji} - \bar{X}_j)^2.$$

Thus, $(n-1)\sigma^{-2}S_j^2$ is distributed as a χ^2-distribution with $n-1$ degrees of freedom.

Let

$$T_{ji} = \frac{\{n/(n-1)\}^{\frac{1}{2}}\sigma^{-1}Z_{ji}}{\sigma^{-1}S_j};$$

then T_{ji} is distributed as a t-distribution with $n-1$ degrees of freedom. Further, if the random variables Y_{ji} $(j = 1, \ldots, m; i = 1, \ldots, n)$ correspond to the re-scaled data given by (A2), then

$$T_{ji} = \left(\frac{n}{n-1}\right)^{\frac{1}{2}} Y_{ji},$$

where

$$Y_{ji} = \frac{Z_{ji}}{S_j}.$$

Further,

$$Y_{ji} = \left(\frac{n-1}{n}\right)^{\frac{1}{2}} T_{ji}$$

are independent $(j = 1, \ldots, m)$. Also Y_{ji} is distributed as a t-distribution with $n-1$ degrees of freedom with mean zero and variance

$$\frac{n-1}{n-3} \times \frac{n-1}{n} = \frac{(n-1)^2}{n(n-3)}.$$

Consider $\mathbf{Y} = (Y_1, \ldots, Y_m)$, Y_j $(j = 1, \ldots, m)$ corresponding to the re-scaled random variables. Then

$$f_{\mathbf{Y}}(t) = Y_1 2^{-\frac{1}{2}} + Y_2 \sin t + Y_3 \cos t + Y_4 \sin 2t + Y_5 \cos t2t + \ldots,$$

$$\operatorname{var}\{f_{\mathbf{Y}}(t)\} = \operatorname{var}(Y_1)\{\tfrac{1}{2} + \sin^2 t + \cos^2 t + \sin^2 2t + \cos^2 2t + \sin^2 3t + \ldots\}$$

$$= \begin{cases} \dfrac{2^{-1}(n-1)^2}{n(n-3)} m & (m \text{ odd}), \\ \dfrac{2^{-1}(n-1)^2}{n(n-3)} \left\{ m - 1 + 2\sin^2\left(\dfrac{mt}{2}\right) \right\} & (m \text{ even}). \end{cases}$$

Thus

$$\frac{(n-1)^2(m-1)}{n(n-3)2} \leqslant \operatorname{var}\{f_{\mathbf{Y}}(t)\} \leqslant \frac{(n-1)^2}{n(n-3)} \tfrac{1}{2}(m+1) \quad (-\pi < t < \pi).$$

References

Andrews, D.F. (1972). Plots of high-dimensional data. *Biometrics* **28,** 125–136.

Andrews, D.F. & Herzberg, A.M. (1985). *Data.* A Collection of Problems from Many Fields for the Student and Research Worker. New York: Springer-Verlag.

Banfield, C.A. & Bassil, L.C. (1977). A transfer algorithm for non-hierarchical classification. *Appl. Statist.* **26,** 206–210.

Beale, E.M.L. (1969). Euclidean cluster analysis. *Bull. Int. Statist. Inst.* **43,** II, 92–94.

Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. (1983). *Graphical Methods for Data Analysis.* Belmont, California: Wadsworth.

Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *J. Am. Statist. Assoc.* **68,** 361–368.

Everitt, B.S. (1987). Graphical displays of complex data-scientific tools or simply art for art's sake. *Proceedings of the 46th Session, ISI,* Tokyo, *III,* 353–367.

Fisher, N.I. (1987). Graphical methods in statistics: current and prospective views. *Proceedings of the 46th Session, ISI,* Tokyo, *III,* 387–400.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics* **7,** (II), 179–188.

Friedman, J.H. & Tukey, J.W. (1974). A projection pursuit algorithm for exploratory data analysis. IEEE Trans. Comput. **C-23,** 881–890.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* New York, Wiley.

Goodchild, N.A. & Vijayan, K. (1974). Significance tests in plots of multidimensional data in two dimensions. *Biometrics* **30,** 209–210.

Gray, A. (1990). *Tubes.* Redwood City, California: Addison-Wesley.

Huber, P.J. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13,** 435–525.

Jolliffe, I.T. (1986). *Principal Component Analysis.* New York: Springer-Verlag.

Jolliffe, I.T. Jones, B. & Morgan, B.J.T. (1980). Cluster analysis of the elderly at home: a case study. In *Data Analysis and Informatics,* Ed. E. Diday, L. Lebart, J.P. Pages and R. Tomassone, pp. 745–757. Amsterdam: North Holland.

Kleiner, B. & Hartigan, J. A. (1981). Representing points in many dimensions by trees and castles (with discussion). *J. Am. Statist. Assoc.* **76,** 260–276.

Krzanowski, W.J. (1988). *Principles of Multivariate Analysis: A User's Perspective.* Oxford: Clarenden Press.

Nevai, P.G. (1979). *Orthogonal Polynomials.* Memoirs of the American Mathematical Society Number 213. Providence, Rhode Island: American Mathematical Society.

Spanier, J. & Oldham, K.B. (1987). *An Atlas of Functions.* Berlin: Springer-Verlag.

Résumé

Andrews (1972) introduit une méthode de représentation graphique en 2 dimensions de données multidimensionelles en utilisant les valeurs des variables en tant que coefficients d'une fonction de Fourier. Ce travail explore l'influence du changement d'ordre et d'échelle des variables, ainsi que le choix d'une nouvelle base orthogonale (c-à-d Legendre ou Chebychev); quelques exemples sont données.

[*Received January* 1989, *accepted September* 1990]