Visualization of High-Dimensional Data via Orthogonal Curves

César García-Osorio
(Departament of Civil Engineering
University of Burgos, Burgos, Spain
cgosorio@ubu.es)

Colin Fyfe
(School of Computing
University of Paisley, Paisley, UK
colin.fyfe@paisley.ac.uk)

Abstract: Computers are still much less useful than the ability of the human eye for pattern matching. This ability can be used quite straightforwardly to identify structure in a data set when it is two or three dimensional. With data sets with more than 3 dimensions some kind of transformation is always necessary. In this paper we review in depth and present and extension of one of these mechanisms: Andrews' curves. With the Andrews' curves we use a curve to represent each data point. A human can run his eye along a set of curves (representing the members of the data set) and identify particular regions of the curves which are optimal for identifying clusters in the data set. Of interest in this context, is our extension in which a moving three-dimensional image is created in which we can see clouds of data points moving as we move along the curves; in a very real sense, the data which dance together are members of the same cluster.

Key Words: grand tour methods, Andrews' curves, exploratory data analysis, visual clustering

Category: I.5.3 Clustering, I.5.5 Implementation, H.3.3 Information Search and Retrieval

1 Introduction

A strong desire of all data analysts is to have the ability to visualize data. Often this means taking a low dimensional projection of a data set and looking in turn at a variety of one, two or, at most, three dimensional projections of the data. An alternative is to transform the data in some way in order to make the data's properties visible via the transformation. Therefore the transformation has to maintain some inherent properties of the data if we are to be able to identify some inherent characteristics of the data after the transformation. In this paper we present a variation of Andrews' curves which performs that kind of transformation.

We start with a review in depth of Andrews' curves, then we focus on Wegman's curves, an extension of Andrews' curves that allows us to construct a two

dimensional grand tour. We give a new characterization of Wegman's curves, and by reformulating Wegman's contribution, we are able to extend that visualization method to three dimensions which makes the identification of structure very much easier. Then, we analyze the properties of this new transformation and the way we can use it for data analysis. We end with some extension suggestions and with the conclusions.

2   Andrews' curves

Andrews [Andrews, 1972] described his curves in 1972, early on in the computing era; it is an interesting observation that he thought it necessary to counsel "an output device with relatively high precision ... is required". Current standard PC software is quite sufficient for the purpose. The method is a way to visualize and hence to find structure in high dimensional data. Each data point $\mathbf{x} = \{x_1, x_2, ..., x_d\}$ defines a finite Fourier series

$$f_{\mathbf{x}}(t) = x_1/\sqrt{2} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + ... \quad (1)$$

and this function is then plotted for $-\pi < t < \pi$. Thus each data point may be viewed as a line between $-\pi$ and π. This formula can be thought of as the projection of the data point onto the vector

$$\left(\frac{1}{\sqrt{2}}, \sin(t), \cos(t), \sin(2t), \cos(2t), \dots \right) \quad (2)$$

If there is structure in the data, it may be visible in the Andrews' curves of the data. An example of Andrews' curves on the well known iris data set is shown in Figure 1. The data set is four dimensional and so only the first four terms of (1) are used. In varying the value of t in (1), we are moving along the curve; data points which are similar will behave similarly in that the locus of their movement will be similar. Thus in Figure 1, we see that $t = 3$ gives us a value for a linear projection of the data which differentiates one type of data from the other two but there is some difficulty in differentiating between these two (for any value between 2 and 3 we have a good separation of the curves corresponding to one of the classes, and hence a good projection to differentiate one class from the others, although there is still some overlapping between the other two).

These curves have been utilized in fields as different as biology [Murphy, 2003], neurology [Koziol and Hacke, 1991], sociology [Spencer, 2003] and semiconductor manufacturing [Rietman et al., 1998, Rietman and Layadi, 2000]. Some of their uses include the quality control of products [Kulkarmi and Paranjape, 1984], the detection of period and outliers in time series [Embrechts et al., 1986] or the visualization of learning in artificial neural networks [Gallagher, 2000]. Khattree and Naik [Khattree and Naik, 2002] have suggested their utilization in robust design and in correspondence analysis.
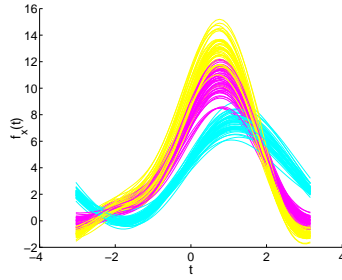
Figure 1: An Andrews' plot of the iris data set. It is clear that one type of iris is distinct from the other two but differentiating between the other two is less easy.

2.1   Properties

These curves have several useful properties, some of which are:

1. Mean preservation. The function corresponding to the mean of a set of N multidimensional observations, is the pointwise mean of the functions corresponding to these observations:

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

2. Distance preservation. The distance between two functions defined as

$$||f_{\mathbf{x}}(t) - f_{\mathbf{y}}(t)||_{L_2} = \int_{-\pi}^{\pi} \left[f_{\mathbf{x}}(t) - f_{\mathbf{y}}(t) \right]^2 dt$$

is proportional to the Euclidean distance between the corresponding points since

$$||f_{\mathbf{x}}(t) - f_{\mathbf{y}}(t)||_{L_2} = \pi \sum_{i=1}^{d} (x_i - y_i)^2 = \pi ||\mathbf{x} - \mathbf{y}||^2$$

3. One-dimensional projections. For a particular value of $t = t_0$, the function value $f_{\mathbf{x}}(t_0)$ is proportional to the length of the projection of the vector $(x_1, x_2, \cdots, x_d)$ on the vector

$$\mathbf{f_1}(t_0) = \left(1/\sqrt{2}, \sin(t_0), \cos(t_0), \sin(2t_0), \cos(2t_0), \cdots \right)$$

This means that the curves are simultaneously showing all the projections onto that vector for the range $-\pi < t < \pi$.

4. Linear relationships. If a point **y** lies on a line joining **x** and **z**, then for all values of t, $f_\mathbf{y}(t)$ is between $f_\mathbf{x}(t)$ and $f_\mathbf{z}(t)$.

5. Also if the components of the data are uncorrelated with common variance σ^2, the Andrews' curves representations preserve that variance. This variance preservation property lets us perform a test of significance using the curves, although, as noted by Khattree and Naik [Khattree and Naik, 2002], this "*is less useful since most multivariate data are either correlated and/or have unequal variances across the variables*".

All these properties were noted by Andrews [Andrews, 1972]. The last one was generalized by Goodchild and Vijayan [Goodchild and Vijayan, 1974] to the case of unequal and not necessarily orthogonal variances. Tests of significance at particular values of t are still possible, but not so the overall tests mentioned by Andrews.

However, the Andrews' curves have also a drawback, in that they suffer from strong dependence on the order of the variables, i.e. if we change the order of variables the shape of the curves is completely different. That is why Embrechts and Herbeg [Embrechts and Herzberg, 1991] propose to try different arrangements of the variables to find the most suitable Andrews' curves. Also, as pointed by Andrews [Andrews, 1972], in the plots low frequencies are more readily seen than high frequencies. For this reason it is useful to associate the most important variables with low frequencies.

2.2 Variations

Some variations of the Andrews' curves have been proposed throughout the years. Andrews himself [Andrews, 1972] proposed the use of different integers to give the general formulation:

$$f_\mathbf{x}(t) = x_1 \sin(n_1 t) + x_2 \cos(n_1 t) + x_3 \sin(n_2 t) + x_4 \cos(n_2 t) + \ldots$$

The restriction to integers is because of the distance preserving property; without integers, this property is lost. Andrews compared the curve with values $n_1 = 2, n_2 = 4, n_3 = 8, \ldots$ with the original formulation and concluded that the former is more space filling but more difficult to interpret when it is used for visual inspection.

Embrechts and Herzberg investigate in [Embrechts and Herzberg, 1991] the effect of re-scaling and re-ordering the coefficients and the interpretation of the plots when one or more coordinates are made equal to zero. They propose the use of other kinds of orthogonal functions such as Legendre and Chebychev polynomials. They give many examples of these variations using the iris data set. In [Embrechts et al., 1995], Embrechts completes this study with a new

variation consisting of the use of wavelet functions. All these variations have been used later by Rietman and Layadi [Rietman and Layadi, 2000] as a help to monitor the manufacture of silicon wafers; they also point to a previous work [Rietman et al., 1998] which used another variation consisting of drawing the Andrews' curves using polar coordinates.

A bivariate version of Andrews' plots has been proposed by Kokiol and Hacke [Koziol and Hacke, 1991]:

Given two vectors of observations $x^T = (x_1, \cdots, x_p)$ and $y^T = (y_1, \cdots, y_p)$ where the $(x_i, y_i), i = 1, 2, \cdots, p$ are naturally paired, form the functions

$$f_x(t) = x_1/\sqrt{2} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \cdots$$
$$f_y(t) = y_1/\sqrt{2} + y_2 \sin(t) + y_3 \cos(t) + y_4 \sin(2t) + y_5 \cos(2t) + \cdots$$

and plot $(t, f_x(t), f_y(t))$ for a set of t-values in the range $-\pi \leq t \leq \pi$.

A similar idea to the previous one, but this time to obtain a three dimensional Andrews' plot, has been proposed in [Wegman and Shen, 1993]. As our extension is inspired in this variation, we postpone its detailed discussion to the next section.

More recently, Khattree and Naik [Khattree and Naik, 2002] have suggested the function:

$$g_{\mathbf{x}}(t) = \frac{1}{\sqrt{2}} \Big\{ x_1 + x_2 \big(\sin(t) + \cos(t)\big) + x_3 \big(\sin(t) - \cos(t)\big)$$
$$+ x_4 \big(\sin(2t) + \cos(2t)\big) + x_5 \big(\sin(2t) - \cos(2t)\big) + \cdots \Big\}, \quad -\pi \leq t \leq \pi$$

$$(3)$$

So, every y_i is exposed to a sine function as well as a cosine function. As they note, one of the advantages of this formulation is that the trigonometric terms in (3) do not simultaneously vanish at any given t. They also establish an interesting relation between the Andrews' curves and the eigenvectors of a symmetric positive definite circular covariance matrix.

3   A new perspective on Wegman's algorithm

Wegman and Shen [Wegman and Shen, 1993] discuss the benefits of using a slightly different projection, namely that onto

$$\mathbf{w}_1 = \sqrt{\frac{2}{d}} \big(\sin(\lambda_1 t), \cos(\lambda_1 t), ..., \sin(\lambda_{\frac{d}{2}} t), \cos(\lambda_{\frac{d}{2}} t)\big)$$
$$\mathbf{w}_2 = \sqrt{\frac{2}{d}} \big(\cos(\lambda_1 t), -\sin(\lambda_1 t), ..., \cos(\lambda_{\frac{d}{2}} t), -\sin(\lambda_{\frac{d}{2}} t)\big)$$

with the λ_j linearly independent over the rationals[1] (that is, there are not rationals r_i for which $\sum_{i=1}^{\frac{d}{2}} r_i \lambda_i = 0$) and the implicit requirement of an even number of terms (if the dimension of the data is odd we made it even by adding one additional 0).

They were concerned with the connection between Andrews' curves and the grand tour noted by Crawford and Fall [Crawford and Fall, 1990] (the grand tour is a multivariate visualization method that consists of looking at the data from all points of view by presenting a continuous sequence, an animation, of low dimensional projections; see [Asimov, 1985], [Buja and Asimov, 1986] and [Wegman and Solka, 2002] for a deeper treatment). They show that Andrews' curves are not a real one-dimensional grand tour. The problem is that Andrews' curves do not exhaust all possible orientations of a one-dimensional vector. Their generalization of Andrews' curves is more space filling, although it has lost the distance preservation property, and can be used to obtain a bi-dimensional *pseudo* grand tour.

Clearly $(\mathbf{w}_1, \mathbf{w}_2)$ form a set of 2 orthonormal basis vectors. If we define

$$y_1 = \mathbf{w}_1^T \mathbf{x} \propto x_1 \sin(\lambda_1 t) + x_2 \cos(\lambda_1 t) + ... + x_d \cos(\lambda_{\frac{d}{2}} t)$$
$$y_2 = \mathbf{w}_2^T \mathbf{x} \propto x_1 \cos(\lambda_1 t) - x_2 \sin(\lambda_1 t) + ... - x_d \sin(\lambda_{\frac{d}{2}} t)$$

then we have a two dimensional display on which to project $\mathbf{x}$ so that we can look for structure by eye. Visually from this projection, we can identify clusters of points which are nearby and whose trajectories as we change t (i.e. as we move along the Andrews' curves) keep close together. When we use these curves in this way we obtain a two dimensional "grand tour" of the data. Figure 2 shows the obtained curves and a snapshot of the grand tour.

Other points may approach a particular cluster for a brief period of time but will not remain within the cluster throughout the tour. This may be seen as

$$\frac{\partial y_1}{\partial t} \propto x_1 \lambda_1 \cos(\lambda_1 t) - x_2 \lambda_1 \sin(\lambda_1 t) + ... - x_d \lambda_{\frac{d}{2}} \sin(\lambda_{\frac{d}{2}} t)$$
$$\frac{\partial y_2}{\partial t} \propto -x_1 \lambda_1 \sin(\lambda_1 t) - x_2 \lambda_1 \cos(\lambda_1 t) - ... - x_d \lambda_{\frac{d}{2}} \cos(\lambda_{\frac{d}{2}} t)$$

with similar patterns holding at higher orders of derivatives. The points within the cluster have a characteristic dance associated with the joint behaviour of the rates of change which are determined by the derivatives which are sinusoids. Now since we can identify clusters from the position and motion of individual points, this suggests a second projection might be useful and so we now investigate a

---

[1] Khattree and Naik [Khattree and Naik, 2002] point to Gnanadesikan [Gnanadesikan, 1977] who attribute a special case of this formulation to Tukey. Tukey used as lambdas $1, \sqrt{2}, \sqrt{3}, \sqrt{5}, \dots$.
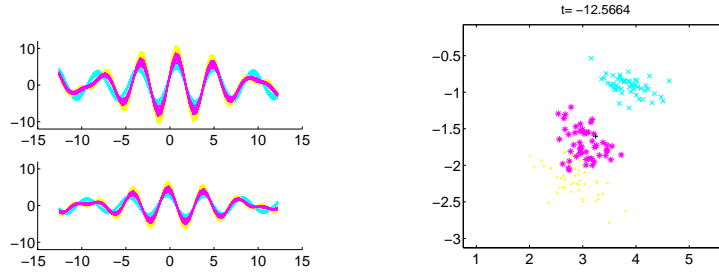
Figure 2: *Left*: the two Wegman's curves for the iris data set $(-4\pi \leq t \leq 4\pi)$. *Right*: a snapshot of the grand tour.

different set of 2 orthogonal basis vectors. We define

$$y_1 = \mathbf{w}_1^T \mathbf{x} \propto x_1 \sin(\lambda_1 t) + x_2 \cos(\lambda_1 t) + ... + x_d \cos(\lambda_{\frac{d}{2}} t)$$

$$y_2 = \frac{\partial \mathbf{w}_1^T \mathbf{x}}{\partial t} \propto x_1 \lambda_1 \cos(\lambda_1 t) - x_2 \lambda_1 \sin(\lambda_1 t) + ... - x_d \lambda_{\frac{d}{2}} \sin(\lambda_{\frac{d}{2}} t)$$

where λ_i are now integers as they were with the Andrews original curves. We will call this basis the *derivative basis* in the following. Since

$$\left| \frac{\partial \mathbf{w}_1^T}{\partial t} \right|^2 = \lambda_1^2 + \lambda_2^2 + ... + \lambda_{\frac{d}{2}}^2 \tag{4}$$

we may readily re-normalize these vectors to get a set of orthonormal vectors the second of which differs slightly from Wegman's basis in that each term is related by

$$(\mathbf{w}_2)_i(\text{derivative}) = \sqrt{\frac{d}{2}} \frac{\lambda_i}{\sqrt{\sum_{i=1}^{\frac{d}{2}} \lambda_i^2}} (\mathbf{w}_2)_i(\text{ Wegman }) \tag{5}$$

where we have used $(\mathbf{w}_2)_i$ to denote the ith element of the vector $\mathbf{w}_2$. Therefore Wegman's basis is a special case of the derivative basis discussed herein. In practice, subjectively we have seen little difference between the projections of the data onto the two bases.

The derivative basis also gives us an insight into the characteristic dance of points in a cluster: such points appear to move about in a group as though joined together by a set of springs. Let $\bar{\mathbf{x}} = \{\bar{x}_1, \bar{x}_2, ..., \bar{x}_d\}$ be the mean vector of a cluster of data points which have some underlying relation e.g. they all belong to a specific class. Let $\bar{\mathbf{x}}$ be mapped to $\bar{y}$ using Andrews' Curves so that

$$\bar{y} = \bar{x}_1 \sin(\lambda_1 t) + \bar{x}_2 \cos(\lambda_1 t) + ...\bar{x}_d \cos(\lambda_{\frac{d}{2}} t) \tag{6}$$

Now consider a specific data point, $\mathbf{x}^1$, a member of this cluster, which is mapped to y^1. Let

$$\mathbf{x}^1 = \bar{\mathbf{x}} + \epsilon^1 = \{\bar{x}_1, \bar{x}_2, ..., \bar{x}_d\} + \{\epsilon_1^1, \epsilon_2^1, ..., \epsilon_d^1\} \tag{7}$$

Then

$$y^1 = \bar{y} + \epsilon_1^1 \sin(\lambda_1 t) + \epsilon_2^1 \cos(\lambda_1 t) + ... \epsilon_d^1 \cos(\lambda_{\frac{d}{2}} t) \tag{8}$$

Now the distance of y^1 from $\bar{y}$ is dominated by those ϵ_i^1 terms which have corresponding trigonometric terms tending to 1 i.e. where $|\cos(\lambda_i t)| \to 1$ i.e. $\lambda_i t \to 0, \pi, 2\pi$ etc or $|\sin(\lambda_i t)| \to 1$ i.e. $\lambda_i t \to \pi/2, 3\pi/2$ etc. But these terms are exactly the terms where $\frac{\partial w}{\partial t} \to 0$ i.e. there is liable to be a low rate of change of their position. Thus we tend to see groups moving in a relatively fixed position for reasonably long spells as we change t; it is not too fanciful to describe the resulting movement as a gentle dance.

4   Extending the derivative curves

Now the Andrews' Curves were derived during the infancy of computational power. We now have very much more powerful machines and, in particular, visual representation on screen is much more sophisticated than it was in those times. Specifically, we now have hardware accelerators which will allow real time plotting of three dimensional projections of literally millions of points in real time. Given that the human visual system tends to operate in a three dimensional universe, and has no difficulty in determining the third dimension even when only two are really available (as on a computer monitor), it seems natural to extend the derivative curves to three dimensional representations. In fact, our subjective findings are that this facilitates the extraction of structure from high dimensional data sets by human observers.

Therefore this perception of the projections of data points moving in the plane can be extended to data points moving in space so that now $y_i = f(t, s)$ so that

$$y_1 = \mathbf{w}_1^T \mathbf{x} \propto x_1 \cos(\lambda_1 t) \cos(\mu_1 s) + x_2 \cos(\lambda_1 t) \sin(\mu_1 s) + x_3 \sin(\lambda_1 t) + ... \tag{9}$$

$$y_2 = \mathbf{w}_2^T \mathbf{x} \propto x_1 \sin(\lambda_1 t) \cos(\mu_1 s) + x_2 \sin(\lambda_1 t) \sin(\mu_1 s) - x_3 \cos(\lambda_1 t)... \tag{10}$$

$$y_3 = \mathbf{w}_3^T \mathbf{x} \propto x_1 \sin(\mu_1 s) - x_2 \cos(\mu_1 s) + x_3 * 0 + ... \tag{11}$$

where we have the implicit requirement that the number of terms in each expansion is a multiple of 3 rather than 2 as previously (in the equation for y_3 we have left $x_3 * 0$ for parallelism with equations for y_1 and y_2). Note that the second curve is the derivative of the first one with respect to t, and the third is constructed to be orthogonal to the other two and is proportional to the derivative with respect to s. We have omitted the λ_i, μ_i factors for ease of exposition and because in practice there seems to be very little loss of comprehension when
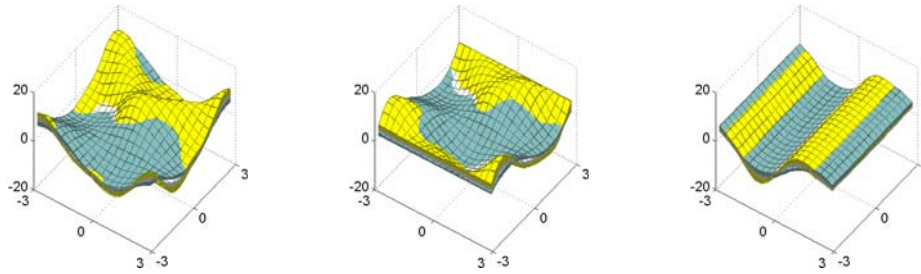
Figure 3: The three groups of surfaces for the Iris data set.

we view the movement of the projections without these factors compared to the corresponding movement when we include these factors.

Note that these equations really give three different groups of surfaces in 3D space but this gives a diagrammatic representation which is very difficult to understand (see Figure 3). Thus we prefer to change t and s independently and view the movement of the groups of points through 3D space. We call the curves obtained when the value of t is fixed, *S-slices*, each corresponding to a particular slice of the surface with a specific t value. Similarly, we call *T-slices* the curves obtained when we fix the value of s. Figure 4 illustrates this for one group of surfaces. An alternative is to let $t = s$ and view the equations as a curve moving in 3D space (Figure 5). One interpretation of this is that the first component represents the (unit) tangent vector to the curve, the second the (unit) normal vector and the third the binormal vector so that these three vectors represent a natural local basis for that space.

4.1   Some properties of the new transformation

As we the original Andrews's curves, our surfaces have several useful properties:

1. The surface representation preserves means . If $\bar{\mathbf{x}}$ is the mean of a set of n multivariate observations $\mathbf{x}_i$, the surface corresponding to $\bar{\mathbf{x}}$ is the pointwise mean of the surfaces corresponding to the n observations:

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

2. Distance preservation. The distance between two surfaces defined as

$$||f_{\mathbf{x}}(t, s) - f_{\mathbf{y}}(t, s)||_{L_2} = \iint_S [f_{\mathbf{x}}(t, s) - f_{\mathbf{y}}(t, s)]^2 dt\, ds$$
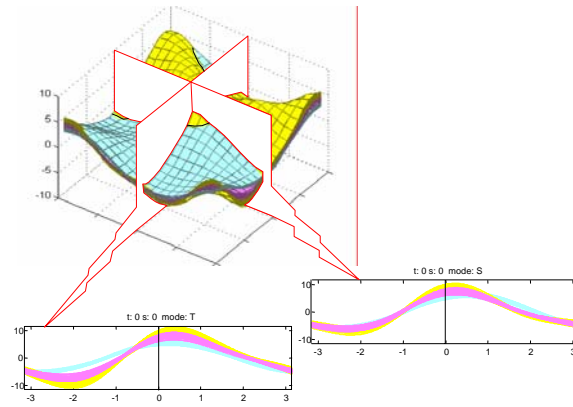
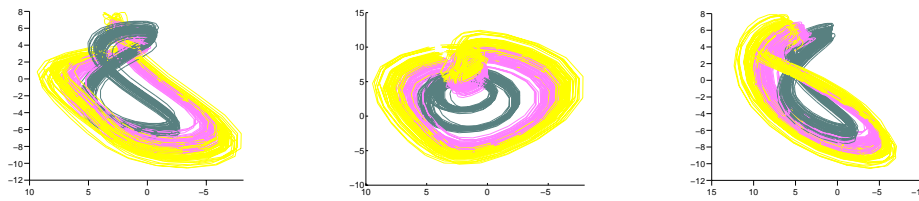**Figure 4:** One of the surfaces and its T and S slices.

Figure 5: Three different perspectives (front, top and side views) of the curves obtained from the Iris data set when $t = s$.

where $S = [-\pi, \pi] \times [-\pi, \pi]$, is proportional to the Euclidean distance between the corresponding points since

$$||f_{\mathbf{x}}(t,s) - f_{\mathbf{y}}(t,s)||_{L_2} = \pi^2 \sum_{i=1}^{d} (x_i - y_i)^2 = \pi^2 ||\mathbf{x} - \mathbf{y}||^2$$

However, this property only holds for the first two surfaces. Besides, it is necessary to introduce a normalization term, $1/\sqrt{2}$, in all the coordinates which are a multiple of three.

3. The representation yields a one-dimensional projection when we simultaneously fix the values of s and t. If we only fix s (t) we obtain the T-slices (S-slices) in which we can simultaneously see the different projections for the range $-\pi < t < \pi$ ($-\pi < s < \pi$).

4. Linear relationships . If a point $\mathbf{y}$ lies on a line joining $\mathbf{x}$ and $\mathbf{z}$, then for all values of t and s, $f_{\mathbf{y}}(t,s)$ is between $f_{\mathbf{x}}(t,s)$ and $f_{\mathbf{z}}(t,s)$.

5   Data exploration with curves

Now we have two representations of the data: we have three curves which are a simple extension of Andrews' curves which we saw earlier; but we also have a representation of each data point as a point in three dimensional space — it is difficult to convey the visual clustering effect of this representation as we move along either curve in a static diagram but the impression of clusters in the dynamically changing environment is very strong.

This suggests the following interactive process for identifying clusters in data: calculate the curve for each data point (in terms of parameters s and t) and search for clusters of points which are performing similarly over at least a local part of the curve and which are also distinct from other points' curves over the same part of the curve. We will initialize our curves with $s = 0$ and $t = 0$ and then simply progress along the curves looking for small sections in which such a group of points can be identified. Of course such a group will not remain distinct from the other curves throughout all its length but, if it is to qualify as a group, it must remain coherent, forming a small bundle of curves through all values of s and t. When we identify such a group, we will remove it from the data set and then repeat the process with the remaining data points.

Since the method is interactive, we are actually happy to work with two types of displays of the data: the first is that described above ; the second is the view in 3 dimensions of the data moving in space. It is difficult to do justice to this second view on a static page but the impression of clusters of objects moving together is very strong in this display. In the web page `http://pisuerga.inf.ubu.es/cgosorio/Andrews/` there is a video demo of an application that implement these surfaces.

One of the criticisms that Andrews himself makes of his curves is that the curves are only useful when the number of points is not too large [Andrews, 1972]. Actually the Andrews' curves and the variant that we propose in this paper can be used without major problems with a greater number of points if the curves are combined with a brushing mechanism [Becker and Cleveland, 1987] that allows us to highlight those points/curves to see if they constitute a cluster (of course when Andrews wrote his paper computing facilities were much less powerful than those we have now). In our implementation, we can instantly highlight a group by changing the colour of the curves of its members and the corresponding points in 3D space using point and click operations (this is known as brushing and linking, [Buja et al., 1991], and allows us to relate information in one plot to the information in another). As noted above, when we are convinced that such a group is moving coherently through all values of s and t, we remove them from the display and continue our search for new clusters.

The tool that we have developed lets the user visualize interactively different slices of our surface. A vertical line indicates the point where the current point

representation is taken from. It is possible to change between the two types of slices, and also it is possible to move interactively changing the s value of a T-slice, or the t value of a S-slice (that is, changing the point where we cut the surfaces to obtain the slices); this gives a kind of grand tour that is multidimensional, since in the slices we have a simultaneous view of all the projections in the range $[-\pi, \pi]$. Embrechts and Herzberg [Embrechts and Herzberg, 1991] suggest trying different orders for the coefficients. With our tool we can also change the order of the variables used to obtain the curves or to choose not to use some of them.

6   Extending our extension of Andrews' curves

Our initial motivation to extend the Wegman and Solka curves was to jump from a 2D representation to a 3D one. We show below the obvious generalization to a four dimensional representation: each of the basis vectors shown is of length one and each is orthogonal to the other three i.e. we have an orthonormal basis. We can use the first three values as coordinates and the fourth one to get the colour of the points from a colour map, close points in the high-dimensional space will give close points in the display with similar colours. The colour idea could be used as well with our original extension to combine in the same display the information from another different group of surfaces (obtained from a different perspective of the data, for example, the principal component projection, or from a different arrangement of the variables), one group of surfaces can give the 3D coordinates and the other the RGB components of the colour of the points[2] (this use of our surfaces is in some way related to the image grand tour [Wegman et al., 1998, Symanzik et al., 2002], but instead of using multi-spectral images we use an arbitrary data set).

$$
\begin{pmatrix}
\cos(\lambda_1 t)\cos(\mu_1 s)\cos(\nu_1 r), & \cos(\lambda_1 t)\cos(\mu_1 s)\sin(\nu_1 r), & \cos(\lambda_1 t)\sin(\mu_1 s), & \sin(\lambda_1 t), & ... \\
\sin(\lambda_1 t)\cos(\mu_1 s)\cos(\nu_1 r), & \sin(\lambda_1 t)\cos(\mu_1 s)\sin(\nu_1 r), & \sin(\lambda_1 t)\sin(\mu_1 s), & -\cos(\lambda_1 t), & ... \\
\sin(\mu_1 s)\cos(\nu_1 r), & \sin(\mu_1 s)\sin(\nu_1 r), & -\cos(\mu_1 s), & 0, & ... \\
\sin(\nu_1 r), & -\cos(\nu_1 r), & 0, & 0, & ...
\end{pmatrix}
$$

We have arranged these four vectors in a manner which should make it clear that the proposal can be extended to any required dimensionality. This statement, though, must come with the caveat that the limit to the assistance which such a sequence of projections can make is liable to be determined by the difficulty which the human mind has on keeping track of several characteristics at one time. Even for five dimensions, one must keep track of position, speed, acceleration, colour and size simultaneously. Also we note that the final basis

---

[2] We can think even in dividing the variables in two groups, and using one group to obtain a set of surfaces, the other group to obtain a second set, and combining both in the same display.

vector in any such basis is dependent on only two out of every five fields in the sample vector and so care must be taken in determining the order of the basis vectors.

7   Conclusions

We have presented a new extension to an existing, indeed, rather old technique for exploratory data analysis. The advance of computational power has made possible extensions which Andrews could only dream about thirty years ago.

   We have pictured how the derivative projections can be used in two quite different but complementary ways:

1. The first allows us to walk along the curves using either of two parameters and find groups of curves which remain as a group for all possible values of the parameters.

2. The second allows us to use the human facility of identifying structure in moving three dimensional displays, something for which our evolution in a three dimensional visual environment has created excellent pattern matchers.

   The combined use of the new display with techniques such as brushing and linking allows us to identify visually the clusters and outliers present in high-dimensional data.

References

[Andrews, 1972] Andrews, D. F. (1972). Plots of High Dimensional Data. *Biometrics*, 28:125–136.

[Asimov, 1985] Asimov, D. (1985). The Grand Tour: a Tool for Viewing Multidimensional Data. *SIAM Journal on Scientific and Statistical Computing*, 6(1):128–143.

[Becker and Cleveland, 1987] Becker, R. A. and Cleveland, W. S. (1987). Brushing Scatterplots. *Technometrics*, 29:127–142.

[Buja and Asimov, 1986] Buja, A. and Asimov, D. (1986). Grand Tour Methods: An Outline. In Allen, D., editor, *Computer Science and Statistics: Proceedings of the Seventeenth Symposium on the Interface*, pages 63–67, Amsterdam: North Holland. Elsevier Sicence Publisher B.V.

[Buja et al., 1991] Buja, A., McDonald, J. A., Michalak, J., and Stuetzle, W. (1991). Interactive Data Visualization Using Focusing and Linking. In *IEEE Visualization '91*, pages 22–25, San Diego, California, USA. IEEE Computer Society Press.

[Crawford and Fall, 1990] Crawford, S. L. and Fall, T. C. (1990). Projection Pursuit Techniques for Visualizing High-dimensional Data Sets. In Nielson, G. M., Shrivers, B., and Rosenblum, L. J., editors, *Visualization in Scientific Computing*, pages 94–108. IEEE Computer Society Press, Los Alamitos, CA.

[Embrechts and Herzberg, 1991] Embrechts, P. and Herzberg, A. M. (1991). Variations of Andrews' Plots. *International Statistical Review*, 59(2):175–194.

[Embrechts et al., 1995] Embrechts, P., Herzberg, A. M., Kalbfleisch, H. K., Traves, W. N., and Whitla, J. R. (1995). An Introduction to Wavelets with Applications to Andrews' Plots. *Journal of Computational and Applied Mathematics*, 64:41–56.

[Embrechts et al., 1986] Embrechts, P., Herzberg, A. M., and Ng, A. C. (1986). An Investigation of Andrews' Plots to Detect Period and Outliers in Time Series Data. *Communications in Statistics – Simulation and Computation*, 15(4):1027–1051.

[Gallagher, 2000] Gallagher, M. (2000). *Multi-layer Perceptron Error Surfaces: Visualization, Structure and Modelling.* PhD thesis, Dept. Computer Science and Electrical Engineering, University of Queensland.

[Gnanadesikan, 1977] Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* John Wiley & son, New York.

[Goodchild and Vijayan, 1974] Goodchild, N. A. and Vijayan, K. (1974). Significance Tests in Plots of Multi-dimensional Data in two Dimensions. *Biometrics*, 30:209–210.

[Khattree and Naik, 2002] Khattree, R. and Naik, D. N. (2002). Andrews Plots for Multivariate Data: some new Suggestions and Applications. *Journal of Statistical Planning and Inference*, 100:411–425.

[Koziol and Hacke, 1991] Koziol, J. A. and Hacke, W. (1991). A Bivariate Version of Andrews Plots. *IEEE Transactions on Biomedical Engineering*, 38(12):1271–1274.

[Kulkarmi and Paranjape, 1984] Kulkarmi, S. R. and Paranjape, S. R. (1984). Use of Andrews' Function Plot Technique to Construct Control Curves for Multivariate Process. *Communications in Statistics – Theory Methods*, 13(20):2511–2533.

[Murphy, 2003] Murphy, J. F. (2003). *Methods for Collection and Processing of Gene Expression Data.* PhD thesis, California Institute of Technology, Pasadena, California 91125.

[Rietman and Layadi, 2000] Rietman, E. A. and Layadi, N. (2000). A Study on $\mathbb{R}^m \to \mathbb{R}^1$ Maps: Application to a 0.16-μm Via Etch Process Endpoint. *IEEE Transactions on Semiconductor Manufacturing*, 13(4):457–468.

[Rietman et al., 1998] Rietman, E. A., Lee, J. T. C., and Layadi, N. (1998). Dynamic Images of Plasma Processes: Use of Fourier Blobs for Endpoint Detection during Plasma Etching of Patterned Wafers. *Journal of Vacuum Science and Technology*, 16(3):1449–1453.

[Spencer, 2003] Spencer, N. H. (2003). Investigating Data with Andrews Plots. *Social Science Computer Review*, 21(2):244–249.

[Symanzik et al., 2002] Symanzik, J., Wegman, E. J., Braverman, A. J., and Luo, Q. (2002). New Applications of the Image Grand Tour. *Computing Science and Statistics*, 34:500–512.

[Wegman et al., 1998] Wegman, E. J., Poston, W. L., and Solka, J. L. (1998). Image grand tour. Technical Report TR 150, The Center for Computational Statistics, `ftp://www.galaxy.gmu.edu/pub/papers/Image_Tour.pdf`.

[Wegman and Shen, 1993] Wegman, E. J. and Shen, J. (1993). Three-Dimensional Andrews Plots and the Grand Tour. *Computing Science and Statistics*, 25:284–288.

[Wegman and Solka, 2002] Wegman, E. J. and Solka, J. L. (2002). On Some Mathematics for Visualising High Dimensional Data. *Indian Journal of Statistics*, 64(Series A, 2):429–452.