Tutorial: Transforms, orthogonality, eigenvectors, and eigenvalues

Ian F. Jones[1]

Abstract

Seismic signal processing often involves the use of various data transforms. The concept of a transform is reviewed along with the notion of orthogonality, which is one of the main properties that make many transforms useful in segregating different parts of the input signal from each other. In addition, digital sampling of underlying continuous functions induces artefacts in the transforms, which are related to the sampling process itself. The nature and origin of these artefacts are described.

Introduction

In one form or another, data transforms underpin a large part of contemporary signal processing, whether in geophysics, communications, or video and image processing. I begin this overview by describing why transforms are of interest, and how we can conceive of a transform. I use three well known, but distinct, classes of transform, namely the Fourier, Radon, and eigenvector transforms. Although these methods are quite different, between them their properties serve to exemplify the salient details involved in most methods.

The underlying motivation for using a transform can be manifold, for example:

- To try to find a representation of some input data object where desired parts of the input signal can be segregated from undesired parts, e.g., primary and multiple separation in seismic processing.
- To find a more compact way of representing and/or storing information, as with data compression.
- To identify common parts of a suite of data objects to facilitate characterization and recognition of features and patterns in these and other objects.

To begin the description of how we can think of a transform, we must first have some basic hypothesis about how we can describe an object. The object in question might be, for example, a single seismic data trace, a 2D data gather, a photograph of a person's face, or a 3D medical image of a human torso. In the case of the Fourier transform, the hypothesis asserts that a single 1D series of samples of data can be represented by a summation of sine waves, each with an amplitude weight and relative phase delay. Or, in facial recognition image processing, we might assert that a set of photographs of faces could be represented by some average representative face, plus a sum of additional 2D correction images representing various deviations from this average face.

The starting object is the input data, and the things that it can be represented by are the transform objects. The set of transform objects we use to describe the input is called the basis of the transform because it is the basis on which we are making the description of the input object. The transform objects are variously referred to as the basis functions, basis vectors, or in some cases principal components, in different publications.

The numerical dimensions of the data define its domain. For example, a 3D volume of recorded seismic data has dimensions of two-way travel time in seconds, inline direction (metres) and crossline direction (metres), and so is in the (t,x,y) domain. A 1D seismic trace is said to be in the time domain, and Fourier transforming that trace would produce data in the frequency domain.

With these notions in mind, we have to assess what contribution the input data make to each of the transform domain objects or, conversely, what weight is applied to each transform object when summing them together to recreate a version of the original input. To do this, we need to assess how similar each input object is to each of the transform objects. Stated differently, what are the characteristics of the transform? For some specific classes of transform, these transform weights constitute the elements of the eigenvectors. (The German word for a characteristic in the context of self is 'eigen'). To assess how much each transform object contributes to the input, or what proportion of the input contributes to the transform objects, we need the notion of similarity.

Similarity

To introduce the concept of digital similarity, consider the following example. Take two digital photographs of the same

[1] *GX Technology EAME Ltd, 1st Floor Integra House, Vicarage Road, Egham, Surrey TW20 9JZ, UK.*
E-mail: ian.jones@iongeo.com

face where the colour scale for each pixel in the photograph is represented by a number in the range, say, ±100, and overall the numbers representing the picture have zero mean. If these images were identical and we overlaid them, then pixel by pixel, all overlapping pixels will have the same numerical values. If we multiply these overlapping numbers and sum their products, i.e., perform a correlation at zero-lag, we will have a single number which can characterize the similarity between the two pictures. This single number is called the variance when the distribution of numbers has zero mean. If the pictures differ slightly, then the sum of these products will, in general, be smaller than for the identical images. Furthermore, if the images were totally different, in general the sum of products would tend to zero.

Orthogonality, eigenvectors, and eigenvalues

If we took a set of vaguely similar photographs, and performed these zero-lag correlations between all possible pairs of images, we would have a table, or matrix, of the variances between the images, referred to as the covariance matrix. If we then formed an average representative image, perhaps by a least-squares fitting process, and subtracted the weighted average image from all of the original images, we would have a new set of images. None of the new images would resemble the average because we would just have removed that component from each image. These resulting images are said to be orthogonal to the average image, as the correlation of any of the new images with the average image would be zero. This kind of analysis underpins the advances in some facial recognition software, e.g., the eigenface decomposition techniques described by Muller et al. (2004) and Papadakis et al. (2007).

If we continued this procedure of removing the common parts of the images, we would end up with a new set of images, each of which was orthogonal to the others. And each of the original images could be reconstructed as a weighted sum of the new orthogonal images. The new orthogonal images constitute the principal component images of the set of original input images, and the weighting functions constitute the eigenvectors of the system. If we computed the sum of squares of the numerical values constituting each orthogonal image, this would be the amount of energy in each of the principal component images, and this single representative number is called its eigenvalue, the square root of which is referred to as the singular value.

The same kind of decomposition can be performed on collections of single traces, or 2D ensembles, as with the photographs, or on higher dimensional objects. For example, geophysicists will be very familiar with the Fourier transform, which decomposes a single trace into a representative set of sinusoids. Each sinusoid is orthogonal to the others: correlating them with each other produces a zero. The amplitude spectrum constitutes the eigenvector weighting function, and the power spectrum (the power in each sinusoidal component) gives the eigenvalue for each frequency component (e.g., Easton, 2010). However, the plots of such spectra can be misleading in this context because the scales are often normalized or shown in relative dB. The Fourier transform is very widely used in geophysics, for 1D functions, and for 2D and 3D data ensembles when it is also known as plane wave decomposition.

In a 1D Fourier transform, the orthogonal principal components are pre-determined in that they are a set of sinusoids whose form is independent of the original input traces that we are decomposing. Several other types of transform with pre-determined basis functions also exist, such as the Walsh transform which decomposes a trace into forms of square waves (e.g., Beauchamp, 1975). Such square-wave decompositions are of use in archaeology, as reconstruction of maps of near-surface magnetometer or resistivity data tend to emphasize linear and right-angled (corner) features associated with foundations of buried building remains. For a 2D or higher dimensional ensemble decomposition, we can either decompose into suite of pre-determined (deterministic) functions, as with the Fourier and Walsh transforms, or alternatively we can decompose into a suite of orthogonal functions derived from the input data themselves, as with the Karhunen–Loève (KL) transform and other singular value decomposition (SVD) methods. In the geological literature, this latter approach when used with certain other assumptions is called 'factor analysis' (e.g., Davis, 1973).

Non-orthogonal transforms

Orthogonality is a nice property to have when attempting to separate different parts of a digitally represented object. When the principal components show no shared variance, they are said to be orthogonal, but there are some transforms that decompose the input data into principal components that do exhibit shared variance, and these transforms are said to be oblique (e.g., Brown, 2009). There are also other transforms, such as the Radon transform, which are likewise not inherently orthogonal and produce non-orthogonal basis functions, but which can be used to reconstruct near-orthogonal subsets of the original data (Trad et al., 2003).

The mathematical form of the transforms

Mathematically, the transforms discussed here are very different from each other. I will not dwell on these differences, but rather on the common features that make them useful. For example, a Fourier transform for continuous data is expressed as an adjoint integral pair: the exact forward and inverse transforms. The KL transform pair is expressed in terms of a projection matrix and its transpose, whereas the Radon transform for discrete data can be written as a summation of a limited subset of transform domain traces.

In simple terms, these three transform pairs can be written as follows. For the Fourier transform, we sum over a range of sinusoids so as to build-up a representation of an arbitrary input signal $f(t)$ in terms of the sinusoids $F(f)$:

$$f(t) = \int_{-\infty}^{\infty} F(f) \exp(i2\pi ft) df \quad (1)$$

$$F(f) = \int_{-\infty}^{\infty} f(t) \exp(-i2\pi ft) dt \quad (2)$$

Remember that the complex exponential term represents both the frequency and phase delay of the sinusoid:

$$\exp(i2\pi ft) = \cos(2\pi ft) + i\sin(2\pi ft) \quad (3)$$

The variables t and f could represent, say, time and frequency. This similar-looking pair of integrals constitutes the 'there-and-back-again' transform pair. Practically, for discretely sampled data, we replace an integral with a summation over a limited range of frequencies. This will be discussed in more detail later. The Fourier transform, especially in two or three dimensions, has a special place in seismic data processing due to the convenience of representing slopes measured in the (t,x) domain 2D data by corresponding inverse slopes in the (f,k) domain data. These slopes represent velocity information and so can be used by certain classes of filter for noise removal or multiple suppression.

For the KL transform, which does not decompose data into a predefined set of functions, we have the following transform pair:

$$x_i(t) = \sum_{k=1}^{n} w_{ik} P_k(t) \quad (4)$$

$$P_k(t) = \sum_{i=1}^{n} w_{ik} x_i(t) \quad (5)$$

The terms $w_{ik,}$ that weight the summation, the eigenvectors, can be determined from manipulations of the covariance matrix, looking for the most similar components of the input data. The best known way of doing this uses singular value decomposition (e.g., Strang, 1980). The transform traces $P_k(t)$ are the n principal components corresponding to the n input traces.

And finally, for the Radon transform expressed as a discrete summation, we have the following transform pair, cast in terms of velocity v, and offset h:

$$\underline{x}(h,t) = \sum_{v=v\min}^{v\max} r\left(v, \tau = \sqrt{t^2 - h^2/v^2}\right) \quad (6)$$

$$r(v,t) = \sum_{h=h\min}^{h\max} x\left(h, t = \sqrt{\tau^2 + h^2/v^2}\right) \quad (7)$$

We write $\underline{x}(h,t)$ rather than $x(h,t)$ because we do not fully reconstruct $x(h,t)$ with this transform, but only the part of it encompassed by the range of values in the summation. The terms for t and τ, like $\sqrt{\tau^2 + h^2/v^2}$, are expressions representing the direction of the steering vector along which we integrate or sum. In this instance, the formulae represent a hyperbola, but can be replaced with that of a parabola or a straight line for the other forms of the Radon transform. Also, because we accept that this transform is not attempting to reconstruct the data exactly, we can formulate the Radon transform as a minimization problem to get the best fit to the data over the limited range of steering vectors being used. For example, the transform can be cast as a least-squares error minimization problem to find the best model describing the input data, x, over some user-specified range of steering vectors. This notion of a 'lossy' transform, which does not, and cannot, fully reconstruct the input data, is discussed later. In the various forms of the transform, the steering vectors are usually referred to as the p values (Figure 2).

Steering vector along which to compute similarity

The basis of all these transforms is the attempt to find similarity between each input trace, or group of traces, and the principal components we are decomposing into. However, when we compare traces, either through the covariance matrix computation or by comparing to a deterministic principal component such as a sinusoid, we also have the option of comparing in different directions.

Figure 1 shows a 2D gather with a number of trajectories indicated on it. This 2D ensemble of seismic traces could be a shot gather, an NMO-corrected gather, a seismic section, or a digitized picture of a face rasterized with, say, a vertical column vectors: the nature of the 2D object is not inherently important. If we compared traces to each other by simply looking laterally, then we are making comparisons along a horizontal trajectory, such as line a) in Figure 1. Conversely, if we compared traces by moving along the linear dipping line b), we would obtain another result. These directions

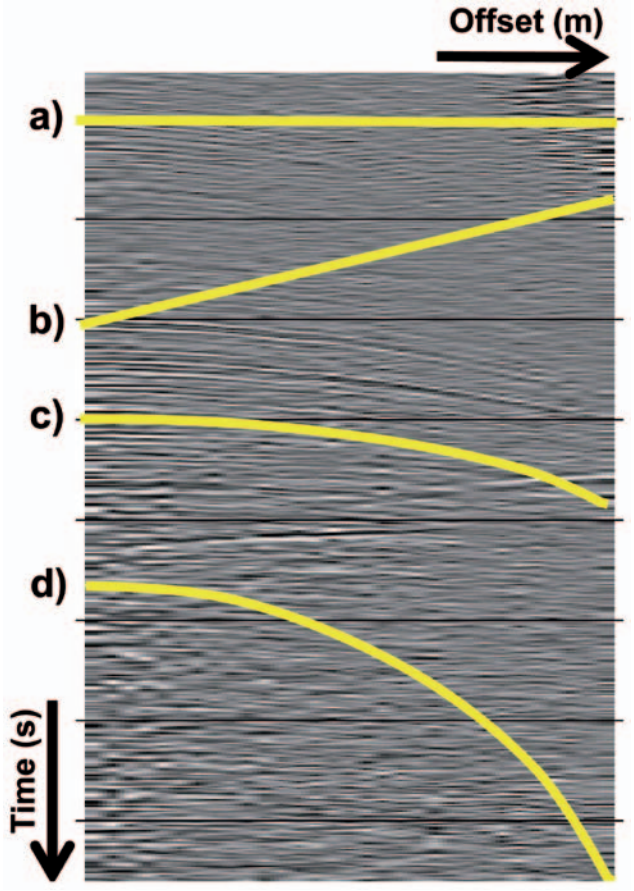

Figure 1 *Examples of possible trajectories along which to compute trace-to-trace similarity: the steering vectors.*

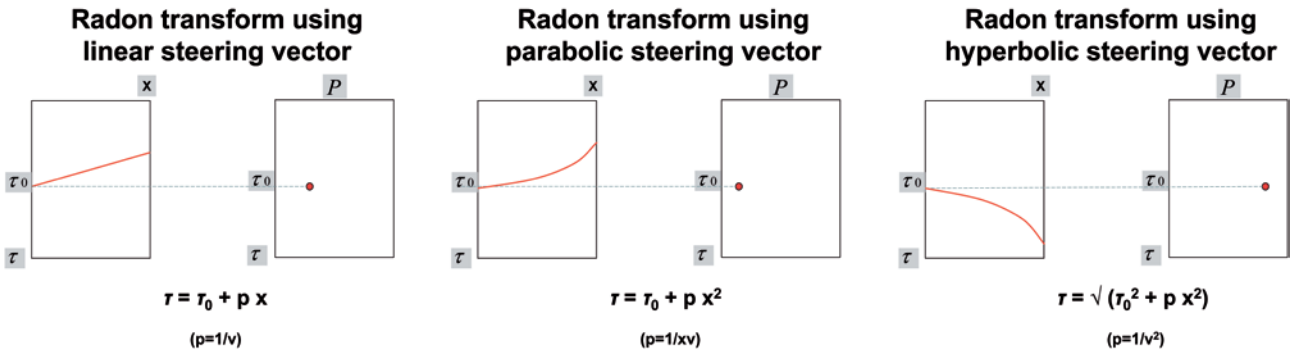

Figure 2 A line in the input (*t*, *x*) domain maps to a point in the Radon transform (*τ*, *p*) domain. The transform can be computed for various trajectories, the steering vectors, along which trace-to-trace similarity is computed. The meaning of p is slightly different in each case.

of investigation and comparison are referred to as the steering vector directions. We could also look along curves rather than straight lines, such as along the parabola c) or the hyperbola d). We could also compute the similarity measure for a large suite of steering vectors, rather than just one. For example, the Radon transform (e.g., Sacchi and Ulrych, 1995), which computes similarity by summing (integrating) along steering vectors as opposed to assessing similarity via analysis of a covariance matrix, is deployed in this way. This summation can be performed for a set of linear trajectories, yielding the linear Radon (τ-p) decomposition (Stoffa et al., 1981); parabolic trajectories, yielding the parabolic Radon transform (Chapman, 1981; Hampson, 1986); and hyperbolic trajectories, yielding the hyperbolic Radon transform which, in a simple form, is used for velocity analysis and is then called a 'velocity spectrum' (Taner and Koehler, 1969; Yilmaz, 1989).

The representation of a single steering vector is a point in the transform domain: each line, whether it be linear, parabolic of hyperbolic, maps to a single point in the basis space (Figure 2). It should also be noted that the steering vector for the Radon transforms emanates from offset zero, In other words, the apex of the parabola or hyperbola is at zero offset. So, for shifted apex events, the basic Radon transform is inappropriate.

For those transforms that determine their basis functions by use of a covariance matrix, then the analogous situation is one in which the covariance matrix is formed by overlapping, multiplying, and summing elements in the objects under consideration after shifting the individual traces or objects with respect to each other. We could compute the single, most similar, principal component for a set of steering vectors, as done for the Radon transforms, or we could compute all principal components for a single steering vector, as done in the KL transform (e.g., Hemon and Mace, 1978; Jones, 1985), or all principal components for all steering vectors such as in the slant-KL transform (Jones and Levy, 1987).

In the case of having computed all principal components, the weighting vectors telling us how much each principal component contributes to each trace is called the eigenvector. In the case of the Radon transform, where we only compute a single principal component for each steering vector, the contribution of that principal component to each input trace is a constant, so in a data reconstruction using this one steering vector, there would be no lateral variation in amplitude in the contribution from this single principal component. To reconstruct any real AVO behaviour, we would have to rely on the superposition of different contributions to reconstitute lateral amplitude variation. Hence, we could think of the KL and other eigenvector decomposition transforms as providing all the orthogonal principal components for a single steering vector direction, whereas the Radon transforms provide just the first, non-orthogonal, principal component for a range of steering vector directions.

An example of a KL transform is shown in Figure 3: nearly flat lying events representing geology (a) are added to steeply dipping noise (b) to produce a 'contaminated' seismic section (c). After decomposition into orthogonal principal components, two reconstructions are performed: Figure 3d shows a data reconstruction using the first five principal components, which in this case carry 75% of the total input energy, and Figure 3e shows the misfit reconstruction, using the remaining principal components. Given that the steering vector was for flat-lying events, the reconstruction shown in Figure 3d is a good representation of the near-flat-lying geology, and the remainder resembles the dipping noise.

Reduction of dimensionality

Consider a graph of two variables in (x,y) coordinates, as shown in Figure 4a, with a scatter of values which lie clustered in the (x,y) plane. To adequately describe this data distribution, we only need these two axes, or dimensions, x and y. Now consider the addition of a third axis, or dimension, z, as shown in Figure 4b. This additional axis is superfluous in describing the data distribution. However, within this three-axis framework, if we were now to rotate the cloud of values onto new axes (x',y',z'), it would indeed require all three axes to describe the newly rotated data distribution (Figure 4c). It can now be seen that by analysing the 3D data distribution shown in Figure 4c, we can find a new projection onto the original 2D axes that adequately describe the data scatter. The procedures to find the minimum number

of dimensions to adequately describe a data distribution are central to orthogonal transforms, and use techniques such as singular value decomposition (SVD). It turns out that the eigenvalues associated with a description of a given data distribution tend to be very small, and thus can be ignored, for any unnecessary dimensions being used in an initial data description. In the context of say, seismic data traces, the scatter of points in Figure 4c could represent some trace attribute plotted versus both shot and receiver locations, whereas plotting the attribute against just midpoint might have been adequate.

In another context, if we took a CMP gather of 100 traces with normal moveout corrections applied, where the NMO had not perfectly flattened all events in the gather, that contained a varied and complex family of reflection events forming a suite of varied and different residual moveout trajectories in the gather, we could perform a parabolic Radon transform to produce 100 p-value traces in the transform domain. In doing this, we would not have reduced the dimensionality of the problem. However, if we only have, say, seven unique parabolic trajectories in the original NMO's CMP gather, such as those shown in Figure 5, then ideally we could adequately describe the data with only seven p values, thus reducing the dimensionality of the problem. In practice, due to the smearing in the transform and band-limited nature of the data, we would need more. Reducing the number of p values too much would not be really reducing the dimensionality of the problem, as we would no longer be adequately representing the data. Application to real marine data is an effective way of removing multiples, as shown in Figure 6.

Lossy versus lossless transforms

If you perform a forward transform, and immediately follow up with the inverse transform, then we can ask whether we recover what we started with or not. If we reconstruct the input data faithfully, within the limits of numerical accuracy, then this is a lossless transform. Examples are the Fourier and KL transforms. However, if we do not recover the input

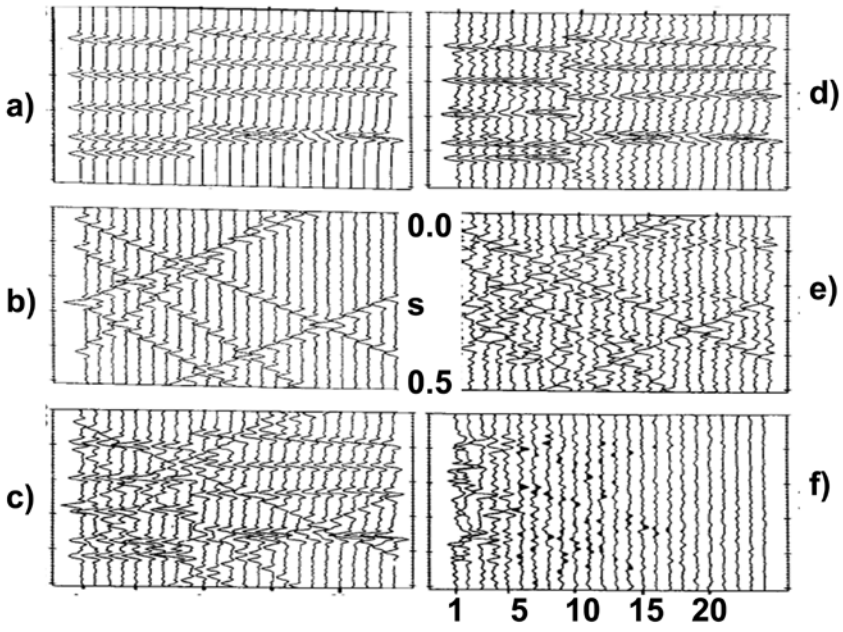

Figure 3 *(a) Flat-lying 'geology' signal. (b) Dipping 'noise'. (c) Sum of (a) and (b) to create the input data. (d) KL reconstruction of the signal using the first five principal components. (e) KL misfit reconstruction of the noise, formed from principal components 6-24. (f) The principal components. From Jones and Levy (1987).*

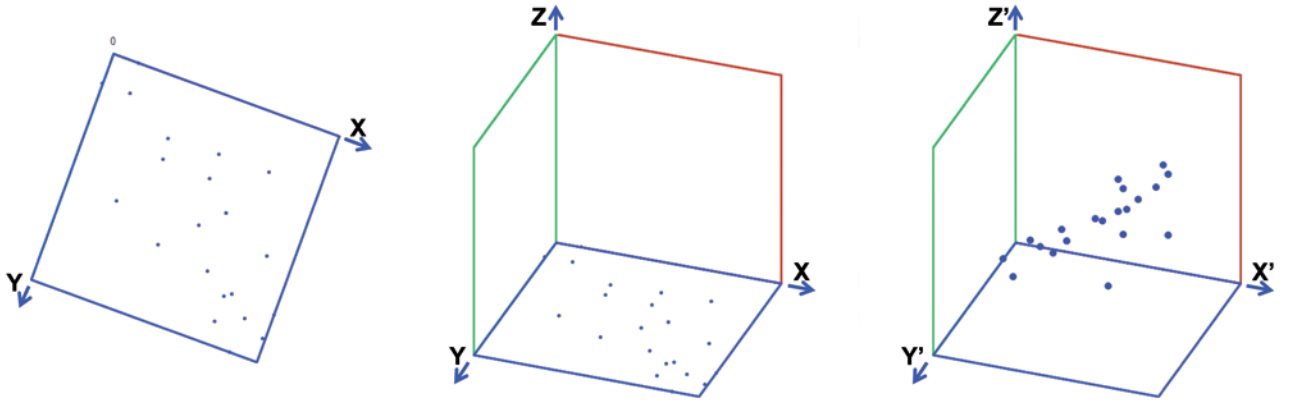

Figure 4 *The same scatter of data points which lie in a 2D plane, represented in (a) a 2D space, (b) a 3D space where the z-axis is superfluous, and (c) a 3D space with rotated coordinates.*

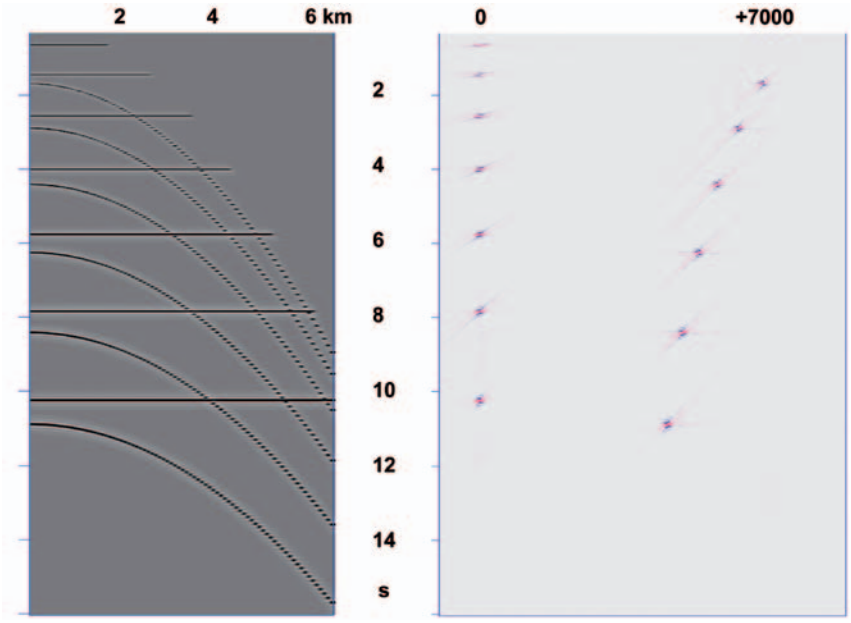

Figure 5 *Left: 13 moveout trajectories, of which seven are the same (flat), thus giving a total of seven different trajectories. The flat-lying events with zero residual moveout representing primaries; the six others represent multiples. Right: The parabolic Radon transform of these data. If we ignore the aspect of smearing in the transform, then we can see that the input data can be represented by just seven unique p value traces, rather than the original 100 input seismic traces: this constitutes data compression. These p traces are not orthogonal to each other, as p traces containing events with the same zero-offset time (τ) will give a non-zero correlation. However, a data reconstruction from events with well-separated p values will tend to be orthogonal. The p-axis has units of milliseconds moveout delay as measured on the far trace.*

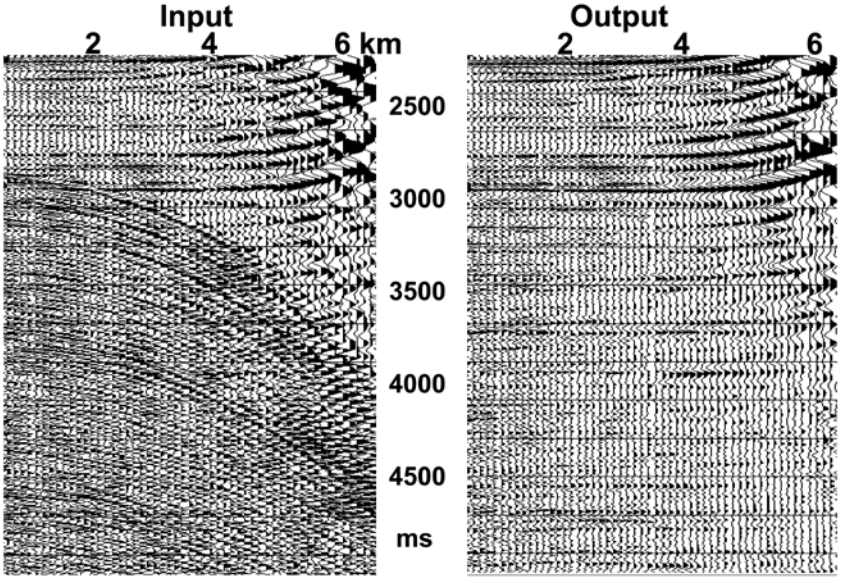

Figure 6 *Application of a parabolic Radon transform to a real normal-moveout corrected CMP gather from a deepwater area. The input CMP data exhibits far-offset aliased multiples. Application of a contemporary multiple suppression method using this technique is very effective. From Stewart et al. (2007).*

data, then it is a lossy transform. Examples are the Radon transforms. In a Radon transform, whether it be linear, parabolic, hyperbolic, or higher order 'anisotropic', we only investigate a limited, user-defined range of steering vectors, and we only estimate the commonest component for each of these steering vectors. Thus, the forward transform only investigates a limited part of the data space, and when we perform an inverse transform, we cannot fully reconstruct the data. It is in part for this reason, that we usually have the option to add back the non-modelled part of the input data during a Radon transform, especially if AVO analysis is later required because the offset-dependent amplitude variation might well be lost during the transform.

If we had a single flat-lying NMO-corrected event in a CMP gather, with zero-offset arrival time $t = \tau_0$, then ideally its parabolic Radon transform would comprise a point located at $(\tau_0, p = 0)$, and the inverse transform of this point would adequately reconstruct the flat event in the CMP gather. However, if we had AVO in the flat input CMP gather data, we could not adequately represent the data with a single point in transform space, but would need superposition of may such points to represent the AVO behaviour of the input data. Figure 7 demonstrates this concept: the flat-lying event has higher amplitude on the far traces; thus we cannot reconstruct it with a single p value, as other trajectories are required so as to constructively sum to build the greater amplitudes on the far offsets. If we computed all the principal components for this flat AVO event using, for example, a KL transform, then we could have reconstructed the AVO, but at a much higher cost compared to the single principal

component of the Radon transform. However, if we had the more realistic scenario, with many flat and non-flat events all with differing AVO behaviour, then even the KL transform would fail, unless we computed all principal components for all steering vectors. In this scenario, the Radon transform is the most cost-effective decomposition to address separation of differing steering vectors, i.e., segregating multiples from primaries, whilst maintaining a good representation of any AVO behaviour.

One-dimensional versus multi-dimensional transforms

As previously described, the Fourier, Radon and KL transforms can be thought of as methods which decompose each individual input trace in terms of the set of basis functions we are using, but each of the principal components is itself a 1D object. In this regard, it does not matter if we are inputting a 2D or 3D ensemble of traces. We are simply seeking similarity along the steering vector direction, or the content of the trace expressed in terms of basis vector contributions, e.g., sine waves in the case of the Fourier transform.

However, if we were working with a suite of digitized photographs of human faces, then looking from one column vector to the next in a given image would not be of much use, as there is no inherent similarity between one vertical slice of the photo and the next. However, for this 2D object, we could directly compare a set of 2D images (photographs of people's faces) by forming the covariance matrix between the suite of 2D images. In this way, we could find the 'common' image, and the subsequent next-most-common image contributions. These 2D principal component images are referred to as eigenfaces in the facial-recognition literature. Whereas the Radon and KL transforms discussed so far have one-dimensional principal components, i.e., a set of 1D principal components describing a collection of 1D traces comprising a single 2D object, those derived for a suite of 2D objects can be 2D. Similarly, if in medical imaging we compared a 3D volumetric representation of a human torso with many such volume images, then the principal components could be 3D objects. Such technology might be used to identify anomalous giblets in the torsos of the subject being studied.

In Figure 8 are shown three eigenfaces resulting from the 2D decomposition of a set of 600 facial images. The first, tenth, and 100th eigenimages from the set of 600 are shown. The eigenvalues of the decomposition (Figure 9) indicate that any of the input images could be reconstructed reasonably well from linear combinations of, perhaps, about 300 eigenimages. Deciding how many eigenvalues are significant is not simple, but typically we specify how much of the input image's energy we want to restore in the reconstruction, and sum over eigenimages so as to include this specified proportion of the total input energy (Jones, 1985). In this context, 'energy' is the sum of squares of the digital samples in the input image.

Figure 10 shows the reconstruction of one of the original 600 digital images in the set, using a sum of the first most energetic 40, 100, and 450 eigenimages, respectively. If all 600 eigenimages had been summed with their appropriate eigenvector weights, then this particular input image would have been perfectly reconstructed. Conversely, in Figure 11, we see a similar reconstruction using a sum of the first 450 eigenimages, but this time for an image that was not in the original set of 600 used to create the transform's basis set of principal component images. In this case, correlation of the new input image, in turn, with each of the 600 pre-computed eigenimages would have provided the weights (eigenvectors), but a perfect, or even a reasonable reconstruction using these weights in conjunction with the eigenimages is probably not possible, because this image never contributed to the computation of the basis set.

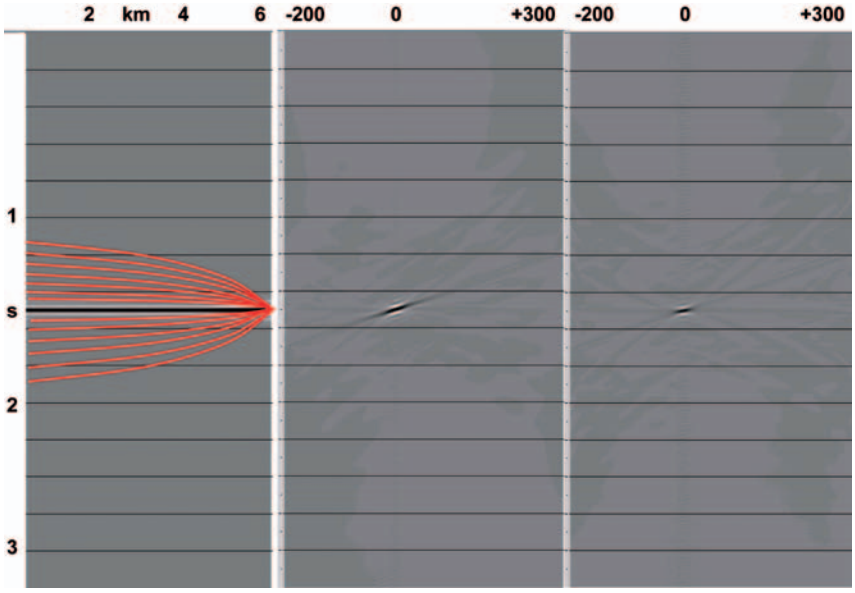

Figure 7 *AVO cannot be recovered using a single p value for a flat event, rather we need the superposition of many events to build-up the far trace amplitude increase, and we rely on destructive interference to cancel the unwanted contributions of the non-zero p trajectories at other offset locations. Left: A single flat-lying event with a three-fold amplitude increase on the far traces. Centre: Parabolic Radon transform of these data, showing a linear dipping feature which is the manifestation of the hypothetical red lines drawn on image at left: these red lies indicate the additional parabolic trajectories required to reconstitute the increase of amplitude seen on the far traces. Right: Radon transform result for a flat event without AVO. The scales have been adjusted to emphasize the features. Note that some required trajectories can emanate from negative times. The p-axis has units of milliseconds moveout delay as measured on the far trace.*

Transform artefacts: the consequences of sampling

In their underlying mathematical forms, the transforms we use are integrals over continuous (analogue) functions. If we discretely sample the data being transformed, then we need to assess the impact of this sampling on the underlying representation of the continuous function in transform space (e.g., Bracewell, 1978; Chapman, 1981; Kanasewich, 1981).

Earlier, I mentioned the mathematical form of the three types of transform that have been discussed here, the Fourier, KL, and Radon transforms, and although these can be written in integral form, in the discrete world of the digital computer we do use summation in place of integration. For example, whereas the Radon integral pair of the forward and inverse transforms represents infinitesimal summation along the path of the integral (the steering vector direction), in the digital representation of the problem, we have data on a regular grid of samples in, say, in time and offset, and it is highly unlikely that the line along which we want to sum will happen to fall nicely on the regularly gridded samples of the discrete data. Hence we have an immediate problem of not having data samples at the points in space where we want to sum them. As a consequence, we either have to accept degradation of the data during transformation, or we need to introduce some clever scheme of estimating what the data value would have been at a point on the integral path, based on what nearby actual data we have on the sampling grid. This problem is shown in Figure 12.

In addition, once we discretely sample a function, we can no longer capture information about frequencies above a certain value. The well-known Nyquist theorem relates the digital sampling of data to the maximum frequency that can be represented in a signal. This is given as the inverse of twice the sampling interval, L, so for 4 ms data sampling, the maximum frequency is $1/2L = 125$ Hz; or for offsets spatially sampled at 50m, the maximum wavenumber, or spatial frequency, that can be represented is 0.01 m^{-1}. In practice, the maximum frequency that can be adequately represented in terms of reliable amplitude behaviour is only about two-thirds of the Nyquist frequency.

An abrupt termination of a function, i.e., stopping or starting your measurements after a certain time or distance, constitutes a step function, called a Heaviside function after the mathematician who characterized it. During an experimental procedure, we start the experiment, then we stop it: so we 'turn on,' then 'turn off' the measurement of the process. This constitutes multiplying the theoretical infinite version of the underlying process with two step functions, i.e., we multiply the underlying continuous data with zeroes before we start measuring and after we stop measuring. This double Heaviside function has a Fourier transform response in the frequency domain that is an oscillatory function known as a sinc function, and the effect of this truncation on the spectrum of the real underlying data response is imposed via a convolution of the data spectrum with the truncation operator spectrum (Bracewell, 1978), resulting in a spectrum with a ringy appearance even though the real, underlying data spectrum is smooth.

Figure 8 *A selection of the eigenimages (principal component images) from a 2D decomposition of a suite of 600 digital images of human faces. Shown from left to right are the first, tenth, and 100th eigenimages. Data courtesy of Neil Muller (Muller et al., 2004).*

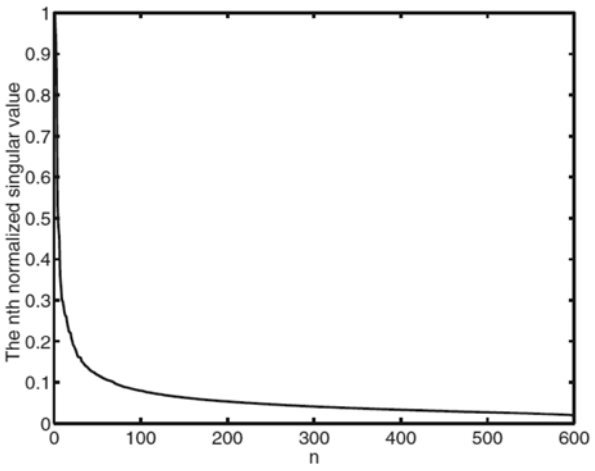

Figure 9 *The eigenvalue distribution (spectrum) for the 2D basis decomposition of 600 sample reference images of human faces.*

Figure 10 *Reconstruction of one of the original suite of 600 images, using summation over the first 40, 100, and 450 eigenimages, respectively. If all 600 eigenimages had been summed with their appropriate eigenvector weights, then this particular input image would have been perfectly reconstructed. Data courtesy of Neil Muller.*

Figure 13 is a cartoon representing the Fourier transform of a continuous time function, and also the Fourier transforms of a sampling grid and measurement truncation operators. In the input time domain, the sampling grid is essentially multiplied with the continuous time function to provide the digitized data that we work with on the com-

Figure 11 *Reconstruction for an image that was not in the original set of 600 used to create the transform's basis set of principal components, using summation of the first most energetic 450 eigenimages. Reconstruction of a reasonable image in this case is unlikely, as the underlying properties of the input image are not captured in the basis set. Data courtesy of Neil Muller.*

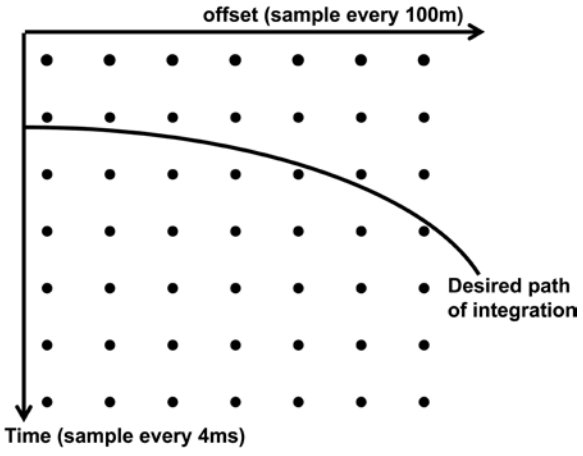

Figure 12 *Where we would like to have data, on the line-integral path, in order to sum it as part of the transform, and where we actually have discretely sampled values. This sampling problem leads to wavelet distortion if not adequately dealt with.*

puter. In addition, as we start and stop taking measurements, the underlying continuous function is truncated with step functions. In the Fourier transform domain, the transform of the truncation and sampling operators are convolved with the transform of the underlying continuous signal. Conventionally, we only plot the positive half of the first repetition of the total frequency-domain signal.

In the case of Radon transforms, whereas the continuous transform of the single parabolic trajectory in Figure 2 (centre) would have been a single well focussed point in the parabolic Radon domain, the effect of sampling acts on this point so as to smear it with the transform response of the sampling operation, which makes the point look like a tilted bow-tie. Here the sampling has been done both temporally and spatially to give, for example, a time sample every 4 ms and a spatially sampled trace every 100 m. The characteristics of the smearing operators can be understood by considering the contribution of a single input wavelet at zero-offset time τ_0 and offset x. Figure 14 shows the contribution to the transform of a single waveform at offsets x = 0, 3, and 6 km. The superposition of the responses for all input offset contributions produces the total transform response shown in Figure 15.

One way to circumvent the undesired ringing or other artefacts introduced by truncation operators is to modify the shape of the truncation operator so that the ringiness of its transform is less pronounced. It is for this reason that we taper data edges prior to using transforms: all the data edges for the Fourier transform and the far-offset edge for the Radon transform. In addition, it has also been common practice to interpolate the input data prior to transformation, so as to reduce the deleterious effects of sampling on the transform process.

Regularity of sampling

If data are regularly sampled, then this regularity can be exploited to speed up computation, such as with the fast Fourier transform (Cooley and Tukey, 1965). Consequently, most processing techniques are developed with the assumption that the input data are regularly sampled. However, most seismic data are acquired on a grid that is not truly regular, and which may also have extensive gaps requiring interpolation prior to subsequent processing. Hence much effort has gone into developing techniques to regularize irregularly and/or sparsely sampled data, e.g., the discrete and anti-leakage Fourier transforms (Sacchi and Ulrych, 1996; Duijndam and Schonewille, 1997; Xu et al., 2005).

Historical example of sampling pitfalls

The behaviour of digital data is influenced by the imprint of the sampling processes used. Hence an understanding of the results obtained in an experiment, using data sampled discretely over a fixed range of measurements, implicitly involves an understanding of the filter responses of the sampling and conditioning processes used in the experiment.

The earliest example of misunderstanding these processes of which I am aware is that of Knott's (1897) analysis of periodicities in earthquake data brought about by lunar tides (Jones, 1980). If we take a time series that is infinitely long, it will have a certain Fourier frequency spectrum. However, if we truncate it to include only those measurements made over a fixed time period, such as one year, then the frequency spectrum will now be influenced by the truncation effects imposed on the time series, which is a sinc function with zeros at frequencies which are multiples of the inverse of the length of the truncation window. Knott (1897) interpreted the periodicity of the truncation operator's ringiness as being a tidal period imposed on the earthquake data by lunar tides, as the ringiness happened to coincide with the expected

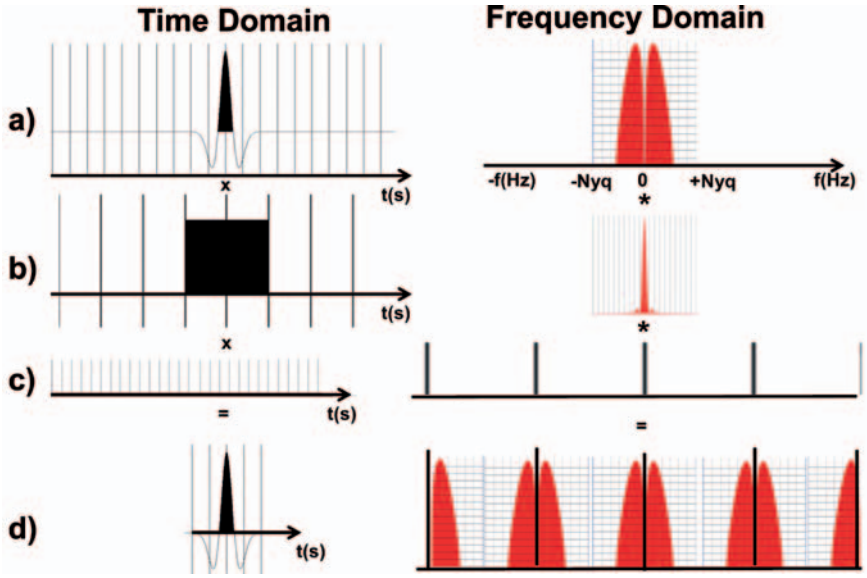

Figure 13 (a) The input analogue signal and its Fourier frequency spectrum. For the continuous signal, the frequency spectrum axis extends to ±∞, though the signal itself might be bandlimited, as shown here. (b) The data truncation operator is a 'boxcar' function whose spectrum is a sinc function. (c) Sampling the data discretely at intervals of, say, 4 ms is essentially multiplying the analogue single with a 4 ms sampling comb, with zeros between live points at 4 ms spacing. The spectrum of a comb with time sample spacing L seconds is a comb with frequency spacing 1/L Hz. (d) In the time domain, multiplying the analogue signal with the boxcar and the sampling comb corresponds to convolving their frequency domain counterparts, to produce the spectrum shown. Conventionally we only ever display the first 'repeat' of the frequency spectrum, out to a maximum frequency of ±1/2L, the Nyquist sampling theory limit. For example, for time domain data sampled at 4 ms, the Nyquist frequency will be 125 Hz.

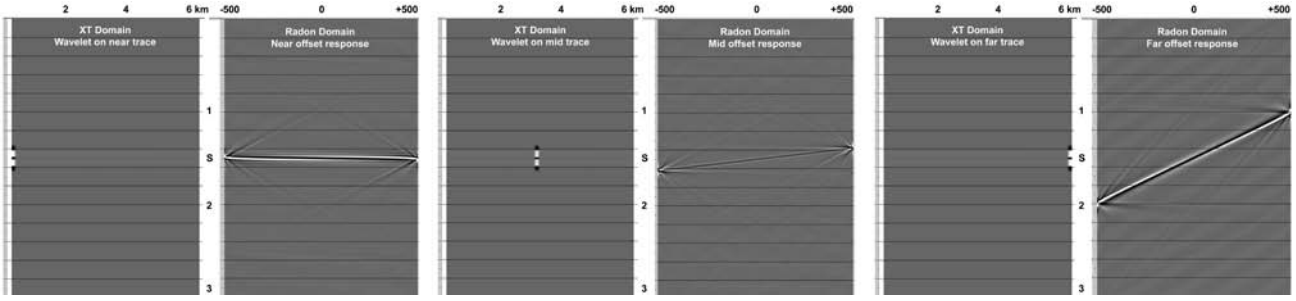

Figure 14 Left: A wavelet on the near trace can be thought of as being made up of many curves all intersecting at zero offset, but with differing dips; hence the transform shows all possible dips, within the range scanned, for a constant arrival time. Centre: A mid-offset wavelet is constituted by a limited range of curves spanning a small time range. Left: A far-offset wavelet is constituted with a wider range of dips and times. The p-axis has units of milliseconds moveout delay as measured on the far trace.

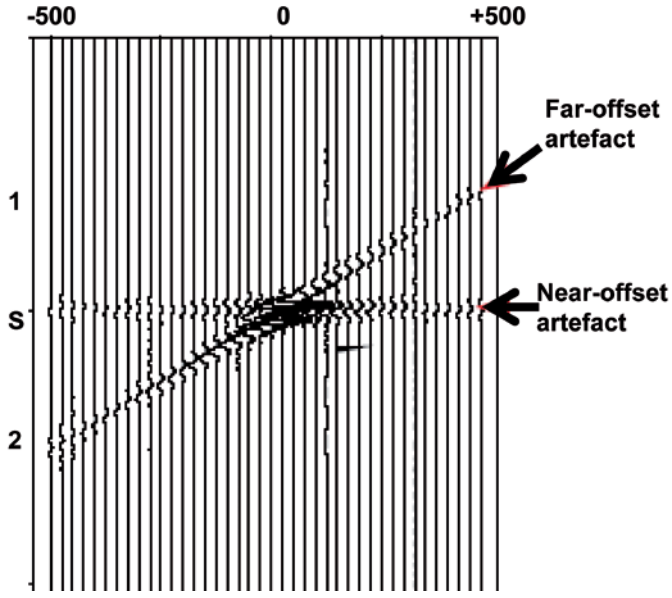

Figure 15 Boundary artefacts for the parabolic Radon transform. We do not see events between the near-offset and far-offset artefacts because the sidelobes of the wavelets tend to cancel, as they are not in-phase, leaving just the near and far boundary effects.

roughly 28-day period of the lunar cycle and its harmonics. Later the same year, Schuster (1897) pointed out the error of Knott's ways, showing that an otherwise perfectly flat spectrum will display these periodicities once the truncation operators are applied to it: in other words, Knott's (1897) analysis was biased by the effects of a filter response, which was perhaps masking any underlying meaningful data. However, Knott later justified his interpretation by noting that, to some extent, the truncation notches alone could not explain the observed periodicities.

The notion of tidally triggered earthquake activity was not investigated again for almost 100 years, until the Apollo astronauts had installed geophones on the moon (e.g., Brzostowski and Brzostowski, 2009) and tidally triggered moonquake activity was observed (e.g., Goulty, 1979). This led to a reassessment of the earthquake catalogue, specifically for small magnitude quakes, and indeed tidally triggered earthquakes are clearly present (e.g., Datta and Kamal, 2011; Tanaka, 2012).

Conclusions

Orthogonality can be thought of as the absence of shared variance. Exploiting this property permits the separation of

different parts of an input signal, which is of great use in many data processing environments. Non-orthogonal transforms, although lacking this desired property, can also be of great use in cost-effectively removing unwanted components of an input signal. Awareness of the various types of transform, and their associated sampling–induced artefacts, is essential in understanding contemporary seismic data processing.

Acknowledgements

I thank Brian Russell for extensive proof reading and constructive feedback on improving this work, and to John Brittan and Joe Zhou for many helpful suggestions.

References

Beauchamp, K.G. [1975] *Walsh Functions and Their Applications*. Academic Press, London.

Bracewell, R.N. [1978] *The Fourier Transform and Its Applications*. McGraw-Hill, New York.

Brown, J.D. [2009] Choosing the right type of rotation in PCA and EFA. *Shiken: JALT Testing & Evaluation SIG Newsletter*, **13**(3), 20–25. http://jalt.org/test/PDF/Brown31.pdf.

Brzostowski, M. and Brzostowski, A. [2009] Archiving the Apollo active seismic data. *The Leading Edge,* **28**, 414–416.

Chapman, C.H. [1981] Generalized Radon transforms and slant stack. *Geophysical Journal of the Royal Astronomical Society*, **66**, 445–453.

Cooley, J.W. and Tukey, J.W. [1965] An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, **19**, 297–301.

Datta, A. and Kamal, [2011] Triggering of aftershocks of the Japan 2011 earthquake by Earth tides. http://arxiv.org/ftp/arxiv/papers/1106/1106.2938.pdf.

Davis, J.C. [1973] *Statistics and Data Analysis in Geology*. Wiley, New York.

Duijndam, A. and Schonewille, M. [1997] Nonuniform fast Fourier transform. *SEG Expanded Abstracts*, **16**, 1135–1138.

Easton, R.L. [2010] *Fourier Methods in Imaging*. Wiley, New York.

Goulty, N.R. [1979] Tidal triggering of deep moonquakes. *Physics of the Earth and Planetary Interiors*, **19**, 52–58.

Hampson, D. [1986] Inverse velocity stacking for multiple elimination. *SEG Expanded Abstracts*, **5**, 422–424.

Jones, I.F. [1980] *Analyses of Global Seismicity: Earthquake Migration and the Frequency-Magnitude Relation*. MSc Thesis, University of Western Ontario.

Jones, I.F. [1985] *Applications of the Karhunen- Loève Transformation in Reflection Seismology*. PhD Thesis, University of British Columbia.

Jones, I.F. and Levy, S. [1987] Signal-to-noise ratio enhancement in multichannel seismic data via the Karhunen- Loève transform. *Geophysical Prospecting*, **35**, 12–32.

Kanasewich, E.R. [1981] *Time Series Analysis in Geophysics*. University of Alberta Press, Edmonton.

Knott, C.G. [1897] On lunar periodicities in earthquake frequency. *Proceedings of the Royal Society*, **60**, 457–466.

Muller, N., Magaia, L.O. and Herbst, B.M. [2004] Singular value decomposition, eigenfaces, and 3D reconstructions. *SIAM Review*, **46**, 518–545.

Papadakis, P., Pratikakis, I., Perantonis, S. and Theoharis, T. [2007] Efficient 3D shape matching and retrieval using a concrete radialized spherical projection representation. *Pattern Recognition*, **40**, 2437–2452.

Sacchi, M.D. and Ulrych, T.J. [1995] High-resolution velocity gathers and offset space reconstruction. *Geophysics*, **60**, 1169–1177.

Sacchi, M.D. and Ulrych, T.J. [1996] Estimation of the discrete Fourier transform: a linear inversion approach. *Geophysics*, **61**, 1128–1136.

Schuster, A. [1897] On lunar and solar periodicities of earthquakes. *Proceedings of the Royal Society*, **61**, 455–464.

Stewart, P.G., Jones, I.F. and Hardy, P.B. [2007] Solutions for deep water imaging. *Geohorizons*, January, 8–22.

Stoffa, P. L., Buhl, P., Diebold, J.B. and Wenzel, F. [1981] Direct mapping of seismic data to the domain of intercept time and ray parameter–a plane-wave decomposition. *Geophysics*, **46**, 255–267.

Strang, G. [1980] *Linear Algebra and Its Applications*. Academic Press, New York.

Tanaka, S. [2012] Tidal triggering of earthquakes prior to the 2011 Tohoku-Oki earthquake (M_w 9.1). *Geophysical Research Letters*, **39**, doi:10.1029/2012GL051179.

Taner, M.T. and Koehler, F. [1969] Velocity spectra-digital computer derivation and applications of velocity functions. *Geophysics*, **41**, 441–463.

Trad, D., Ulrych, T. and Sacchi, M. [2003] Latest views of the sparse Radon transform. *Geophysics*, **68**, 386–399.

Xu, S., Zhang, Y., Pham, D. and Lambare, G. [2005] Anti-leakage Fourier transform for seismic data regularization. *Geophysics*, **70**, 87–95.

Yilmaz, O. [1989] Velocity-stack processing. *Geophysical Prospecting*, **37**, 357–382.

Received 18 June 2012; accepted 12 November 2012.
doi: 10.3997/1365-2397.2013002