Iterative Singular Valued Decomposition

A set of data vectors $\{\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3, \cdots, \boldsymbol{d}_m\}$, $\boldsymbol{d}_i \in \mathbb{R}^n$ include information of significance to a physical property being measured by an experiment. As a consequence of the measurement process, a variation from the true values due to random noise is included in these data vectors. Each data vector $\boldsymbol{d}_i$ additionally may be identical to another data vector $\boldsymbol{d}_j$ within the data set in terms of true signal, but differ from each other only in terms of a random noise component.

If random noise were the only difference between these two data vectors $\boldsymbol{d}_i$ and $\boldsymbol{d}_j$ then plotting these two vectors as a set of points in a two dimensional plane, where the coordinates for these 2 D points are defined to be the pair-wise n-dimensional coordinates of $\boldsymbol{d}_i$ and $\boldsymbol{d}_j$, would result in a scatter plot with a distribution of 2D points about a line following the signal variation in terms of magnitude for the n-dimensional coordinates of $\boldsymbol{d}_i$ and $\boldsymbol{d}_j$. If the variation in the magnitude for true signal exceeds the variation in these data due to random noise, a line of best fit to these 2D points would provide, from these two data vectors, an estimate for the true signal underlying $\boldsymbol{d}_i$ and $\boldsymbol{d}_j$.

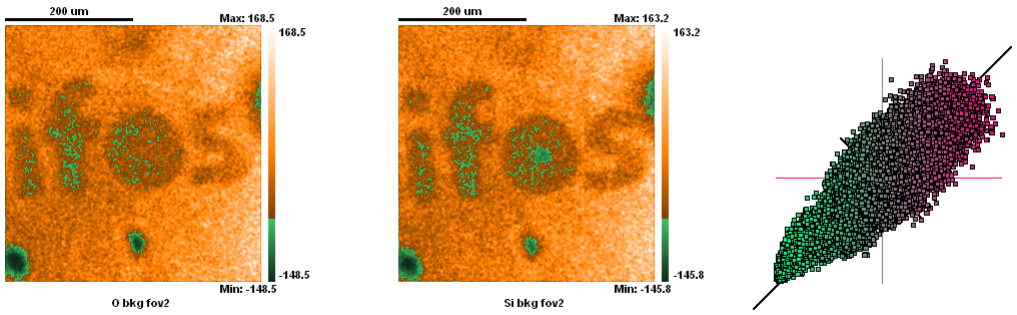

Figure 1: A scatter plot created from two data vectors corresponding to two very similar independent measurements. These images have been offset to centre the scatter plot distribution of 2D points collected from the correspond pixels in each image.

An example of data vectors are images. The two images in Figure 1 plotted as a scatter plot yields a distribution typical of two very similar images acquired from the same sample as separate measurements. These two images are representative of background signal in an XPS experiment and are therefore expected to be similar. Since the two images are reasonably similar, the scatter plot in Figure 1 results in a line of best fit which is essentially 45^{o} to the x-axis.

The line of best fit displayed with the scatter plot in Figure 1 is determined by minimising the sum of squared perpendicular distances from each point in the 2D scatter plot to the line drawn. Given the line of best fit, these images are transformed to the images in Figure 2, which are computed by rotating the line of best fit to align with the x-axis.

The transformed images are the most basic example of how experimental data are transformed by Principal Component Analysis (PCA) by determining eigenvectors for a 2×2 covariance matrix formed from the raw images treated as n-dimensional vectors. Normalised eigenvectors of the resulting covariance matrix are used to construct an orthogonal 2×2 matrix which when applied to the raw image vectors preserves distances between points in the scatter plot and aligns the direction

of greatest variation in these scatter points with the x-axis. The reason these eigenvectors of a covariance matrix serve to transform these data is a consequence of calculus appied to minimising the sum of squared perpendicular distances of the scatter points to a line.

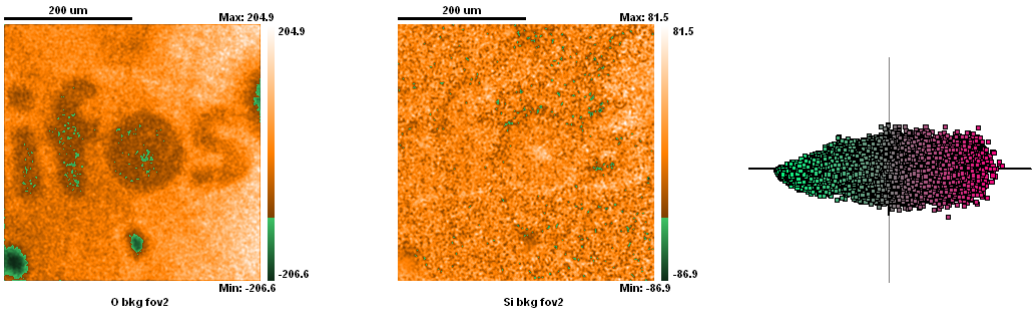

Figure 2: Transformed images corresponding to the raw image data in Figure 1.

This particular example is well suited to PCA. The suitability of these two images is evidenced by the scatter plot in Figure 1, which has the appearance of an elliptical distribution. That is, there are obvious perpendicular lines which could be drawn by eye and, in qualitative terms, intensities within the original images can be easily partitioned into two physically meaningful images, namely, one image containing signal and the other image (Figure 2) mostly representative of random noise from the measurement process.

If it is assumed the steps leading to the transformed images in Figure 2 have indeed partitioned true signal from noise, then to remove noise from both of the original images in Figure 1 each of these two original images when projected onto the signal image in Figure 2 yield noise reduced images.

This rather simplistic analysis for two similar images can be generalised for problems where many images are available. The case in point is where a set of XPS images are acquired from a sample over a range of binding energies. For such a data set, XPS intensities will respond to changes in composition and chemistry, therefore it is possible to obtain images without the similarities of the images in Figure 1. However, a sequence of energy incremented image measurements will result in a significant proportion of images associated with background signal which suggests a large proportion of these images have the potential for similar appearance and hence if analysed on a pair-wise basis, for example, could yield a reduced number of images partitioned into signal and noise. Although plausible on physical grounds, mathematically each transformation equivalent to Figure 2 does not collect all information into one image alone, but leaves subtle information within the lesser image. For example, the pixels at the centre of the O in the characters IFOS can be seen in Figure 2 to be of significance compared to the surrounding O shadow in the second image assumed to be representative of noise. These pixel intensities at the centre of the O are indeed significant and represent a location on the sample where silicon is expected. Neglecting these pixels as part of the noise would result in loss of elemental information from the sample.

To prevent the loss of information within an image set, the steps leading to the result in Figure 2 can be generalised by performing a singular valued decomposition (SVD) to compute the transformation required to minimise the sum of squared distances from a line for multidimensional distributions constructed from coordinates from all images in a data set at once.

While computing the eigenvalues and vectors of a 2×2 matrix is simple, for an $n \times n$ matrix the problem requires more sophisticated algorithms.

The problem stated mathematically is as follows.

Given a set of data vectors $\{\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3, \cdots, \boldsymbol{d}_m\}$, express these vectors using a corresponding set of abstract vectors (abstract factors) $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_m\}$ where the relationship between data vectors and abstract vectors is described by a singular valued decomposition for a data matrix $\boldsymbol{D}$ into three matrices $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{W}$ in the form

$$\boldsymbol{D} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^T$$

where

$\boldsymbol{d}_i \in \mathbb{R}^n$, $\boldsymbol{u}_i \in \mathbb{R}^n$, $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3, \cdots, \boldsymbol{d}_m]$ and $\boldsymbol{U} = [\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \hat{\boldsymbol{u}}_3, \cdots, \hat{\boldsymbol{u}}_m]$

$\boldsymbol{W}$ is a diagonal matrix with diagonal matrix elements equal to the square root eigenvalues of the covariance matrix

$$\boldsymbol{Z} = \boldsymbol{D}^T\boldsymbol{D}$$

and $\boldsymbol{V}$ is the matrix formed from the normalised eigenvectors of $\boldsymbol{Z}$ ordered with respect to the eigenvalues by magnitude. The square root eigenvalues appear ordered by size along the diagonal of $\boldsymbol{W}$.

One of the most efficient algorithms for computing the desired abstract factors $\boldsymbol{U}$ is achieved by transforming the initial data vectors to an equivalent set of vectors with the same eigenvalues and eigenvectors by performing Householder transformations. These Householder transformations are designed to reduce the problem to a simpler eigenanalysis for a matrix in a standard form by, logically, performing a sequence of orthogonal transformations. Essentially, the data vectors are pre-conditioned by the Householder transformations before applying iteration to compute the eigenvalues and eigenvectors.

An alternative to making the investment required by Householder transformations is to guess at an eigenvector and commence NIPALS iterations to compute only a few of the eigenvectors and eigenvalues. The Householder approach is typically used when all the eigenvectors and eigenvalues are required and NIPALS is commonly used for situations where only a few eigenvectors are desired, that is, when the number of images far exceeds the number of images containing true signal.

An alternative to these two approaches is to iteratively compute, like NIPALS, one eigenvector at a time by applying a simple transformation equivalent to the steps described above for two images. The iteration scheme is based not on a 2×2 matrix but a 3×3 matrix. Taking three vectors at a time from the image set, these images are transformed to a new set of vectors which have the same eigenvalues and eigenvectors as the original set of three vectors.

The iterative scheme is as follows:

repeat “while current approximation to largest eigenvector has not converged”

loop i = m down to 3 do

> **"replace the vectors y_1, y_2 and y_i by transformed vectors corresponding to the eigenvectors of covariance matrix computed from y_1, y_2 and y_i in the order of magnitude of the eigenvalues."**

Following these steps causes the vector in $y_1$to converge to the eigenvector with the largest eigenvalue. At the same time as computing the largest eigenvector, the entire set of original vectors are moved towards an orthogonal set of vectors. Incrementing the lower limit of the loop allows the next eigenvector to be computed, thus eigenvectors can be computed one at a time. Like NIPALS only the work required is performed by iterative SVD. As a result data sets with large numbers of vectors are open to analysis which would otherwise not be possible if approached using an algorithm aimed at computing all the eigenvectors.

Random noise in experimental data generally provides a set of linearly independent vectors. One consequence of inherently linearly independent vectors is for such vectors computing the eigenvectors for a 3×3 matrix can be performed directly and, for data with experimental noise, only minor regard for floating point considerations is necessary. For precise numerical work, care must be taken to prevent numerical errors accumulating, but this statement is true for all numerical algorithm including Householder transformations.

Image Data Filtering

The images in Figure 1 when plotted as a scatter plot exhibit a distribution which is open to PCA. For pulse counted data, noise is expected to relate to the counts-per-bin distributed about the mean with a standard deviation equal to the square root of the counts-per-bin. Figure 1 represents images with a count rate for which this is valid, but for lower count rates or for images with artefacts of a detection system, certain pixels within an image distort the type of distribution seen in Figure 1 and prevent meaningful PCA. An approach developed for extreme examples of erroneous pixel intensities is based on filtering images using an averaging kernel in the following form. In CasaXPS the filter is referred to as the Outlier Filter.

The filter algorithm is as follows. An image, logically represented in Figure 3 as pixel boxes, is transformed by computing an average using intensities from surrounding (pink) pixels to the pixel under consideration (blue). The pixel under consideration is modified if the absolute intensity of the difference between the pixel intensity and the mean average is greater than the mean average multiply by a threshold value. If the pixel intensity is deemed to require modification, the value is replaced by the computed mean average from the surrounding pixels indicated as pink boxes in Figure 3.

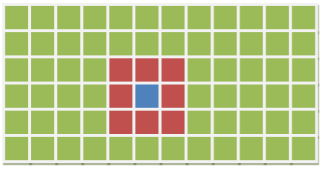

Figure 3: Outlier Filter illustration. Coloured squares represent pixel locations in a larger image.

Anomalous pixel intensities are altered by the Outlier Filter, while minimal adjustments are made to pixels within a range of intensities deemed to be reasonable. When applied to an image with low counting statistics or spikes in pixel intensities, the improvement in image quality can be measured by considering how successful PCA, when applied to these filtered images, distinguishes signal from noise. Moreover, spectra-at-pixels are significantly enhanced if spectra are computed from pre-filtered vectors then further enhanced by PCA noise reduction.

By way of example involving spectra, consider a set of images measured from a silver sample using PE160, FOV1 on an Axis Nova XPS imaging instrument. The entire measurement required less than 34 minutes to acquire 1005 images at 1 eV step-size with a dwell-time per image of 2 seconds. Two images from this silver data set are used in Figure 4 to illustrate the outlier filter and the consequence of applying the outlier filter on scatter plots derived from raw and filtered images.

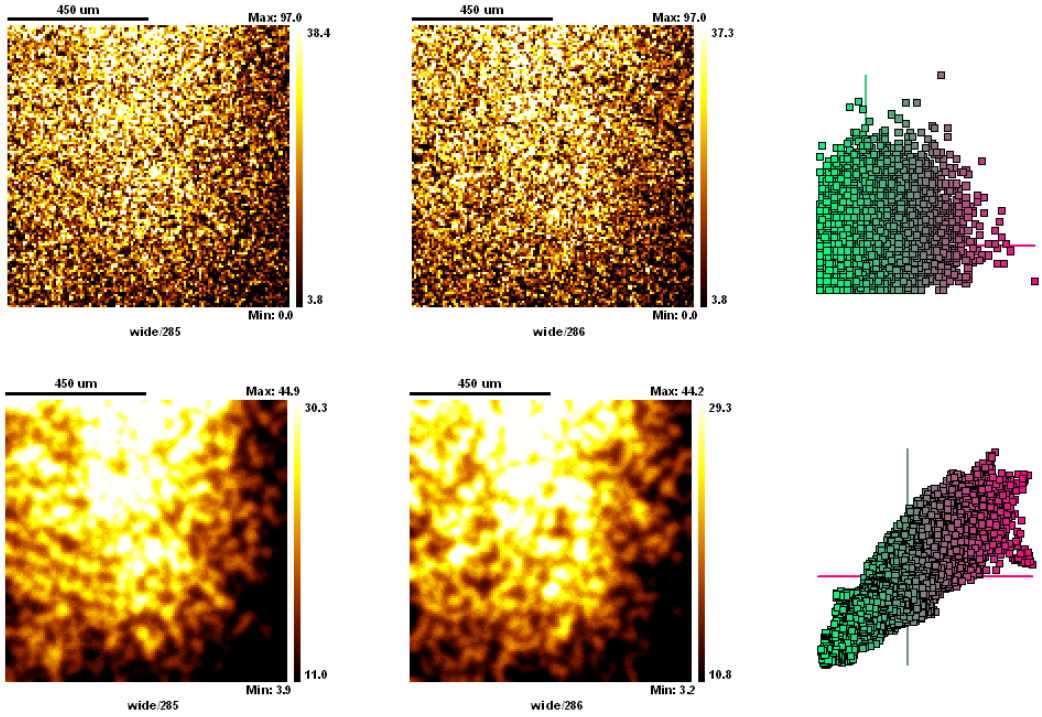

Figure 4: Two images acquired consecutively in a measurement from a silver sample differing in binding energy by 1 eV. The top row represents raw image data, while the bottom row is images processed from the top row by multiple applications of the outlier filter using a threshold of 5%.

Note how in Figure 4 the filtered images result in a scatter plot similar in nature to the scatter plot in Figure 1. By contrast, the raw images result in a distorted scatter plot without a clear principal axis. These images are symptomatic of when pixel intensities in raw images mask the structure clearly evident in both filtered images in Figure 4 and also the distribution of points in the corresponding scatter plot.

A further illustration of how images measured sequentially in energy can be enhanced using outlier filtering is by considering spectra-at-pixels. Figure 5 are three spectra from the same pixel location within an image, formed from the original 128×128 images by compressing the pixel resolution to 32×32 and displaying these spectra using raw compressed data (green), outlier filtered images (blue) and PCA enhanced filtered data using 14 abstract factors to reconstruct the image data set.

The expected ratio for the Ag 3d doublet peaks is 40:60. The expected ratio for these doublet peaks is computed using a component model constrained only by FWHM for these doublet components. The filtered and PCA enhanced spectra conform to Ag 3d doublet component peak ratio.

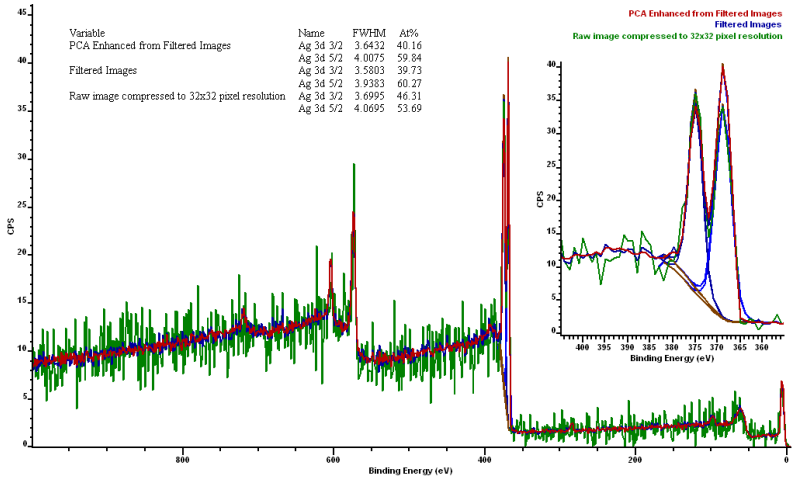

Figure 5: Spectra from the same pixel within a 32x32 image derived from three sets of images, namely, the raw unprocessed set of images (green), the same set of images after multiple applications of the Outlier Filter option (blue) and the PCA enhanced filtered data set using 14 abstract factors to reproduce the data set (red).

Since the spectra in Figure 5 are typical of spectra from the silver data set after compression to 32×32 images, these spectra are effectively parallel acquisition of survey spectra from a silver sample resulting in 1024 spectra for an acquisition time of 34 minutes, that is, less than 2 seconds per spectrum.

As a further illustration of the potential in these types of experiments data presented in Figure 6 are PE40 spectra measured via sets of images. Applying imaging techniques to inhomogeneous samples identifies areas of interest on the sample, and allows the retrospective interrogation of the surface using numerous spectra to guide the analysis. The spectra in Figure 6 are a small set of Ni 3p and Fe 3p peaks sub-sampled from 4096 spectra processed from the set of XPS images. A point of interest with these spectra in Figure 6 is as follows. A peak correlated with Fe 3p can be compared to Na KLL images to test the hypothesis Na 2s signal is the reason for this peak structure rather than a chemically shifted Fe or Ni peak. Correlating chemistry with spatial information is a powerful tool for understanding sample chemistry.

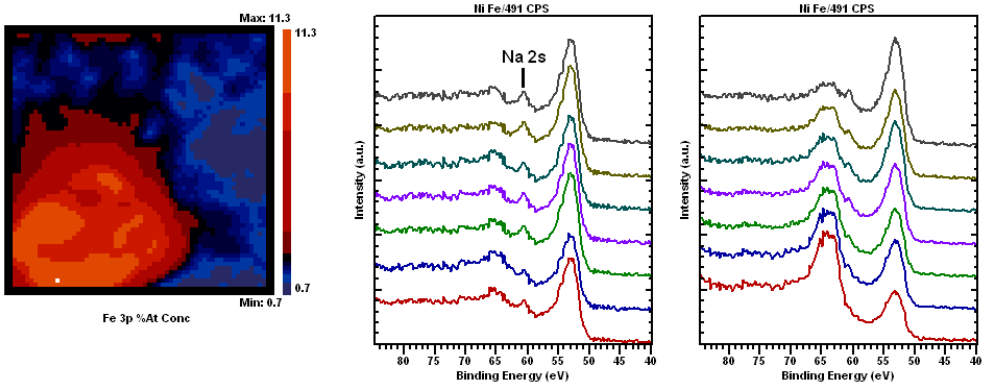

Figure 6: A Ni/Fe corrosion sample measured using 768 images PE40 yield 4096 spectra in parallel for each element. These Fe 3p, Ni 3p and Na 2s spectra are examples of closely located pixels from two separate locations on the sample.