2003-01-01

Independent Subspace Analysis using Locally Linear Embedding

Derry Fitzgerald
Dublin Institute of Technology

Eugene Coyle
Dublin Institute of Technology, Eugene.Coyle@dit.ie

Recommended Citation

Fitzgerald, D. & Coyle, E. (2003) Independent subspace analysis using locally linear embedding. Proceedings of the *Digital Audio Effects Conference* (DAFX03), pp.13-17, London, England, 2003.

INDEPENDENT SUBSPACE ANALYSIS USING LOCALLY LINEAR EMBEDDING

Derry FitzGerald, Eugene Coyle

Dublin Institute of Technology,
Rathmines Rd., Dublin, Ireland
derry.fitzgerald@dit.ie

Bob Lawlor

Department of Electronic Engineering,
National University of Ireland, Maynooth
rlawlor@eeng.may.ie

ABSTRACT

While Independent Subspace Analysis provides a means of blindly separating sound sources from a single channel signal, it does have a number of problems. In particular the amount of information required for separation of sources varies with the signal. This is as a result of the variance-based nature of Principal Component Analysis, which is used for dimensional reduction in the Independent Subspace Analysis algorithm. In an attempt to overcome this problem the use of a non-variance based dimensional reduction method, Locally Linear Embedding, is proposed. Locally Linear Embedding is a geometry based dimensional reduction technique. The use of this approach is demonstrated by its application to single channel source separation, and its merits discussed.

1. INDEPENDENT SUBSPACE ANALYSIS

Independent Subspace Analysis (ISA) provides a means of blind sound source separation from single channel mixtures [1]. ISA represents sound sources as low dimensional subspaces in the time-frequency plane. The single channel mixture is assumed to result from the sum of a number of unknown independent sources. The single channel mixture is converted to a time-frequency representation such as a spectrogram by means of carrying out a Short Time Fourier Transform on the signal and retaining only the magnitude values.

The resulting spectrogram is then assumed to result from the superposition of l unknown independent spectrograms. Further each independent spectrogram is assumed to be represented as the outer product of an invariant frequency basis function f_j and a corresponding time basis function t_j. This yields:

$$\mathbf{Y} = \sum_{j=1}^{l} Y_j = \sum_{j=1}^{l} f_j t_j^T \qquad (1)$$

One way to achieve the decomposition of a spectrogram into a sum of outer products as in eqn. (1) is the use of Principal Component Analysis (PCA). Also known as the Karhunen-Loeve Transform, PCA transforms a set of correlated variables into a number of uncorrelated or orthogonal variables that are termed principal components. The first principal component contains the largest amount of the total variance as possible, and each successive principal component contains as much of the total remaining variance as possible.

As a result of this property one of the uses of PCA is as a method of dimensional reduction, by discarding components that contribute minimal variance to the overall data. Therefore carrying out PCA on a spectrogram and discarding components of low variance will result in a set of low dimensional subspaces that represent aspects of the original spectrogram. However the components returned by PCA are only orthogonal, and are not statistically independent.

In order to achieve statistically independent basis functions a further technique, Independent Component Analysis (ICA) is performed on the components retained from PCA. Independent Component Analysis (ICA) attempts to separate a set of observed signals that are composed of linear mixtures of a number of independent non-gaussian sources into a set of signals that contain the independent sources [2] [3]. The combination of PCA for dimensional reduction followed by ICA to achieve independent basis functions results in the technique known as Independent Subspace Analysis (ISA). The method can be viewed as a two-step process, firstly dimensional reduction and secondly obtaining independent components from the reduced dimensional data. Once the independent basis functions have been obtained they can then be combined to resynthesise the independent sources as described in [1].

However there are a number of problems with ISA. Of particular interest is that the number of basis functions required to identify the sources was found to vary from signal to signal, depending on the relative amplitudes of the sources. Using the threshold method described in [1] was found to be unreliable in determining the required number of basis functions. This indeterminacy is as a result of the variance-based nature of the PCA stage of the algorithm. This inherently biases the analysis towards the loudest sounds in the overall spectrogram, which will account for the largest amounts of variance. As a result sources with low amplitude relative to other sources in the spectrogram will require larger numbers of components to be retained from the dimensional reduction stage before these low amplitude sources can be recognised.

Techniques to overcome this limitation, such as the use of sub-band preprocessing and the use of prior subspaces for sources known to be present in the mixture signal, have been proposed in [4] and [5]. However these techniques make use of prior knowledge about the sources of interest in the signal and this information may not always be available. In such cases where prior knowledge is not available and where some of the sources are known to have lower amplitudes relative to other sources in the mixture signal it can be seen that using a dimensional reduction technique that is not variance-based could potentially improve the robustness of the ISA method. One such dimensional reduction technique is Locally Linear Embedding.

2. LOCALLY LINEAR EMBEDDING

Locally linear embedding (LLE) is a technique for dimensional reduction based on simple geometric intuitions [6] [7]. LLE attempts to obtain a low dimensional mapping for high dimensional data with the property that nearby points in the high dimensional space remain nearby and are similarly co-located with respect to each other in the low dimensional space. In other words the mapping attempts to preserve the local configurations of nearest neighbours.

The data is assumed to consist of N real-valued vectors X_i of dimensionality D. These vectors are taken as samples of the underlying manifold. Provided that the underlying manifold is well sampled then each vector and its nearest neighbours can be assumed to lie on or close to a locally linear piece of the underlying manifold. These pieces of the manifold are then characterised by the use of linear coefficients that reconstruct each vector from its nearest neighbours. In the simplest case K nearest neighbours are identified per vector as measured using Euclidean distance, though the use of other distance metrics is possible. Reconstruction errors are then measured by:

$$\varepsilon(W) = \sum_i \left| X_i - \sum_j W_{ij} X_j \right|^2 \qquad (2)$$

where the weights W_{ij} contain the contribution of the jth vector to the reconstruction of the ith vector. To obtain the W_{ij} the above cost function is minimised subject to two constraints, a sparseness constraint and an invariance constraint. The sparseness constraint is that each vector X_i can only be reconstructed from its K nearest neighbours, in effect forcing $W_{ij} = 0$ if X_j is not one of the nearest neighbours. The invariance constraint is that the rows of the weights matrix are constrained to sum to one, ie $\Sigma_j W_{ij} = 1$. The optimal weights can then be found by solving a set of constrained least squares problems.

An important property of these constrained weights is that for any given vector they are invariant to rotations, rescalings and translations of that vector and its K nearest neighbours. The invariance to rotations and scalings comes from the form of eqn (2) and the invariance to translation is enforced by the constraint that the rows of the weights matrix sum to one. As a result of this the weights characterise intrinsic geometric properties of each neighbourhood as opposed to properties that depend on a particular frame of reference.

The data is then assumed to be on or near a smoothly varying non-linear manifold, with the dimensionality of the manifold being $d<<D$. It is then assumed that there exists a linear mapping, consisting of a translation, rotation and rescaling, which maps the high dimensional neighbourhoods to global coordinates on the underlying manifold. As the reconstruction weights W_{ij} are invariant to translation, rotation and rescaling their characterisation of local geometry in the original data can be expected to be equally valid for local pieces of the underlying manifold. In other words the weights W_{ij} that reconstruct the original vectors X_i of dimensionality D can also be used to reconstruct the underlying manifold in d dimensions.

The next and final step in LLE is then to map the high dimensional inputs X_i to a low dimensional output R_i which represent the underlying manifold. This is done by finding the d dimensional coordinates of each R_i to minimise the embedding cost function:

$$\Phi(R) = \sum_i \left| R_i - \sum_j W_{ij} R_j \right| \qquad (3)$$

As can be seen the cost function is very similar to that of eqn. (1), and is again based on locally linear reconstruction errors. However in this case the weights W_{ij} are fixed and the outputs R_i are optimised. As can be seen the embedding is calculated directly from the W_{ij} without reference to the original inputs X_i and as a result the embedding is performed only with reference to the geometric information encoded in the W_{ij}. In effect the algorithm finds low dimensional outputs R_i that can be reconstructed from the same weights W_{ij} as the original high dimensional data X_i .

The embedding cost function is optimised by solving a sparse N x N eigenvalue problem which is a global operation over all the data points. This contrasts with the fact that the reconstruction weights are calculated from the local neighbourhood of each input. This is how the algorithm attempts to discover global structure, it attempts to integrate information from overlapping local neighbourhoods. Like PCA the resultant outputs R_i are orthogonal to each other. This is achieved in solving the eigenvalue problem. As a result of this LLE shares the property with PCA that only as many outputs R_i as required need be calculated.

The only parameters for the algorithm are chosing the number of dimensions d to represent the data, and the number of neighbours K for each data point. It has been observed in [7] that the results of LLE do not depend sensitively on the number of nearest neighbours, with the provisions that K must be greater than d and that too high a value for K invalidates the assumption that a vector and its neighbours can be modelled linearly.

LLE has proved sucessful in determining the underlying structure of high dimensional data in cases where PCA fails to obtain the underlying structure. LLE appeals to the underlying local geometry of the data presented to it to carry out dimensional reduction, whereas PCA carries out dimensional reduction with reference to the variance of the data. In some cases the geometric methods of LLE provide a more salient description of the data than a variance based approach such as PCA. It should also be noted that, like PCA, LLE also has other uses besides that of dimensional reduction. LLE can also be used as a clustering algorithm and for pattern matching.

3. ISA USING LLE FOR DIMENSIONAL REDUCTION

As noted previously PCA performs redundancy reduction based on variance. As a result, when attempting to separate sources from a spectrogram PCA is biased towards the loudest sources in the spectrogram and can recover a number of principal components from these sources before recovering a component that contains data from one of the lower amplitude sources present. This means that the number of components that needs to be retained to identify all the sources present varies with the relative amplitude of the sources. This can cause difficulties when attempting to separate sound sources which have much lower amplitudes than some of the other sources present in the mixture signal, for example hi-hats tend to be much lower in a mixture signal than either a snare drum or a kick drum.

LLE on the other hand determines components based on regions of similarity (or local neighbourhoods). Therefore LLE

should be less prone to variations in relative amplitude between sources in the mixture spectrogram, and the variation in the number of components required to identify sources should be less severe than that observed when using PCA.

Consider a spectrogram **Y** of size *n* x *m*, where *n* is the number of frequency channels and *m* is the number of time slices or frames. Then with regards to the LLE algorithm the dimensionality *D* of the data is given by *n* and the number of input vectors *N* is given by *m*. The outputs R_i are in this case taken to represent the evolution of similar neighbourhoods through the spectrogram. These similar neighbourhoods are made up of time slices that have similar frequency content, and so the outputs should capture events in the spectrogram that have similar frequency content. Alternatively, by transposing **Y** the neighbourhoods will then consist of frequency regions that have similar evolution through time, resulting in outputs that contain groups of frequencies that occur together, in other words frequency characteristics of a given source.

Figure 1 below shows the first three output vectors R_i obtained from carrying out LLE on a drum loop containing two occurrences each of a snare drum and kick drum, and eight occurrences of a hi-hat. The high hats occur at a lower amplitude relative to that of the snare and kick drums and so would be harder to detect using a variance based method such as PCA. In this case the spectrogram was orientated so that the outputs would capture time events in the spectrogram that have similar frequency content. The number of nearest neighbours *K* was set at 30 and *d* was chosen as 3.

As can be seen LLE has successfully captured the general characteristics of the drum loop, having prominent peaks in amplitude at the correct locations for each of the three drums. This compares favourably with the results obtained using PCA which are shown in Figure 2. While both snare and bass drum are clearly identified in the first two principal components, the hi-hats only show up as very small peaks in the third principal component and are not clearly defined. It can be seen that in this case LLE has more successfully captured information relating to the hi-hats, which were low in amplitude relative to the snare and bass drum.

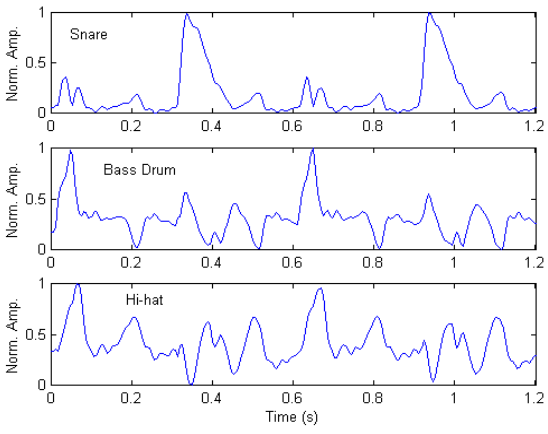

Figure 1:*First 3 components obtained using LLE (K=30)*

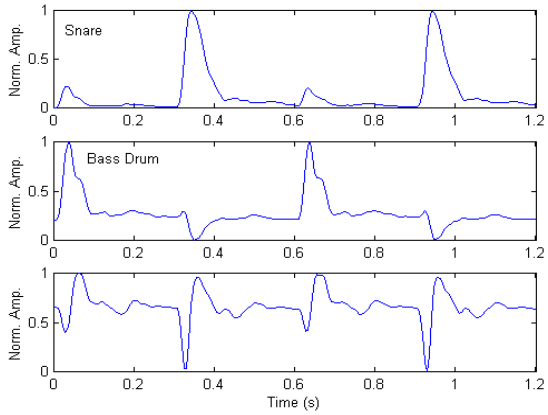

Figure 2: *First 3 components obtained using PCA*

As noted previously the results of LLE do not depend sensitively on the choice of the number of nearest neighbours. Choosing different values for *K* results in outputs that essentially capture the same information on the sources. Figure 3 shows the results obtained by carrying out LLE with $K = 50$ on the same drum loop as in Figure 1. The sources have been captured in the same order, and the same main peaks occur in each source but the overlap between the sources is different. In this case the snare vector shows little evidence of the hi-hats, which instead show up in the bass drum vector, and the hi-hat peaks are more consistent in their amplitudes in the hi-hat vector.

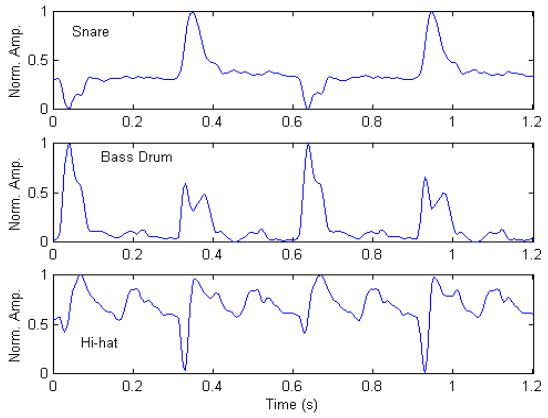

Figure 3:*First 3 components obtained using LLE (K=50)*

Despite capturing the overall structure of the sources, smaller peaks are still visible in the output LLE vectors where other drums occur. These peaks are possibly due to the fact that some of the neighbourhoods integrated in the final step of LLE may consist of neighbours that belong to more than one source, especially in cases where sources occur simultaneously. The vectors recovered are also not statistically independent.

Having achieved dimensional reduction using LLE the outputs from LLE can be passed to an ICA algorithm in a similar manner to the way the outputs from PCA are passed to an ICA algorithm in ISA, effectively performing ISA in the same manner as before but substituting LLE for PCA in the dimensional reduction step of the algorithm. This results in a set of

independent basis functions which can be resynthesised as described in [1]. Figure 4 shows the independent sources obtained if the R_i shown in Figure 1 are transformed using ICA. As can be seen improved separation of the sources has occurred, with noticeably clearer peaks for both the bass drum and hi-hats.

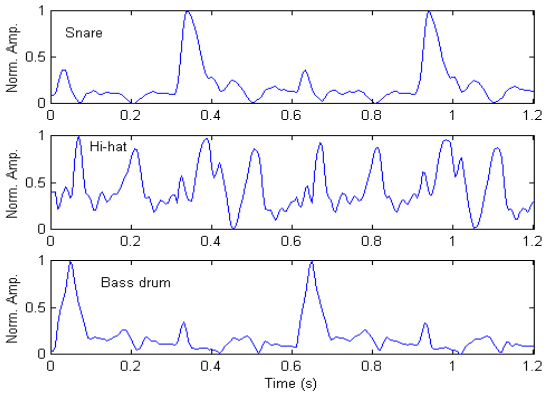

Figure 4: *Independent Components obtained from ICA of LLE outputs (K=30)*

When the R_i for K = 50 are transformed using ICA the sources are again recovered correctly. The independent components obtained are shown in Figure 5. However as can be seen in this case performing ICA has lead to a reduction in peak height for the hi-hats, with the dominant peaks in the hi-hat component being those of the snare drum. This occurs as a result of the two prominent local minima present in the LLE hi-hat vector. As ICA is invariant to scaling these two minima are regarded by the ICA algorithm to be as important as the peaks.

This highlights the fact that while LLE itself is not particularly sensitive to the choice of K, using LLE as a substitute for PCA in the dimensional reduction step of Independent Subspace Analysis results in an increased sensitivity to the choice of K . Careful choice of K results in LLE vectors which give better separation when passed to the ICA step of ISA, though the required separation is still always achieved to some degree. Unfortunately at present there is no suitable method for choosing K for optimal performance with the ICA step and so this remains an issue for future research.

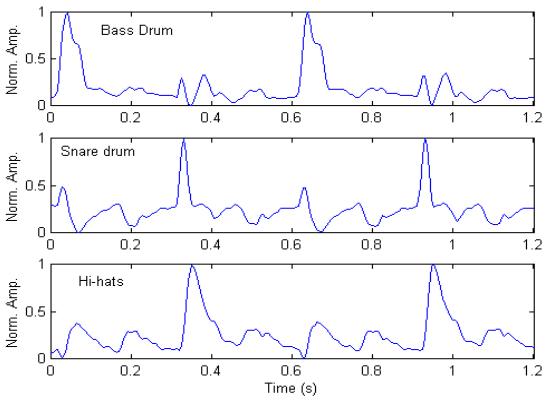

Figure 5: *Independent Components obtained from ICA of LLE outputs (K=50)*

In some cases the LLE algorithm can fail to characterise the sources sufficiently to allow extraction using ISA. One potential reason for this is that as mentioned previously some of the neighbourhoods embedded in the final step of LLE may contain neighbours from different sources. In some cases, if the nearest neighbours do not consistently come from the same source, whether as a result of similar frequency characteristics, or due to overlapping sources causing the occurrence of similar vectors, then the LLE algorithm will fail to characterise the sources adequately. This is as a result of the algorithm mapping faraway inputs to nearby outputs. This type of failure can also occur if the original data is too spare, noisy, or if there is not enough data to ensure that the underlying manifold is well-sampled [7]. However despite the fact that the algorithm will fail under certain conditions, LLE has shown itself to be capable of better characterising the sources present in a mixture signal in many circumstances where PCA fails to do so, and can often do so using fewer dimensions than PCA.

4. CONCLUSIONS

This paper has proposed the use of LLE as a potential means of overcoming some of the problems associated with the variance based PCA step of ISA. LLE is shown to be capable of characterising sources with fewer numbers of components than that required using PCA. This is due to the fact that LLE makes use of local geometry to embed high dimensional data in a low dimensional space. However in some cases LLE does fail to characterise the sources correctly due to too much overlap between integrated neighbourhoods. Despite this LLE has in many cases proved to be an improvement over PCA for dimensional reduction in the ISA algorithm and has proved to be a useful tool for attempting sound source separation.

5. REFERENCES

[1] M. Casey, & A. Westner, “Separation of Mixed Audio Sources By Independent Subspace Analysis” in Proc. Of ICMC 2000, pp. 154-161, Berlin, Germany.

[2] A. Hyvärinen and E. Oja. “Independent Component Analysis: Algorithms and Applications”. *Neural Networks*, 13(4-5): pp 411-430, 2000.

[3] J.F. Cardoso, “Blind Signal Separation: statistical Principles”, Proceedings of the IEEE, Vol.9, No. 10, pp. 2009-2025, Oct 1998,1

[4] D. FitzGerald, E. Coyle, B. Lawlor, “Sub-band Independent Subspace Analysis for Drum Transcription”, Proceedings of the. Digital Audio Effects Conference (DAFX02), Hamburg, pp. 65-69, 2002.

[5] D. FitzGerald, E. Coyle, B. Lawlor. “Prior Subspace Analysis for Drum Transcription”, 114th AES Convention Amsterdam March 22nd–25th 2003

[6] S.T Roweis & L.K. Saul. “Nonlinear dimensionality reduction by locally linear embedding” Science 290:2323-2326, 2000.

[7] L.K. Saul, & S.T. Roweis. “Think Globally, Fit Locally: Unsupervised Learning of Nonlinear Manifolds” Technical Report MS CIS-02-18, University of Pennsylvania, 2002.