Soft-LOST: EM on a Mixture of Oriented Lines

Paul D. O'Grady and Barak A. Pearlmutter

Hamilton Institute
National University of Ireland Maynooth
Co. Kildare
Ireland
paul.ogrady@may.ie barak@cs.may.ie

Abstract. Robust clustering of data into overlapping linear subspaces is a common problem. Here we consider one-dimensional subspaces that cross the origin. This problem arises in blind source separation, where the subspaces correspond directly to columns of a mixing matrix. We present an algorithm that identifies these subspaces using an EM procedure, where the E-step calculates posterior probabilities assigning data points to lines and the M-step repositions the lines to match the points assigned to them. This method, combined with a transformation into a sparse domain and an L_1-norm optimisation, constitutes a blind source separation algorithm for the under-determined case.

1 Introduction

Mixtures of oriented lines arise in sparse separation when a set of observations from N sensors, $\mathbf{X} = (\mathbf{x}(1)|\cdots|\mathbf{x}(T))$, consist of a linear mixture of M source signals, $\mathbf{S} = (\mathbf{s}(1)|\cdots|\mathbf{s}(T))$, by way of an unknown linear mixing process characterised by the $N \times M$ mixing matrix $\mathbf{A}$ via $\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t)$. When $N = M$ the sources can be recovered by an unmixing matrix $\mathbf{W}$ where $\hat{\mathbf{s}}(t) = \mathbf{W}\,\mathbf{x}(t)$ and $\hat{\mathbf{s}}(t)$ holds the estimated sources at time t, with $\mathbf{W} = \mathbf{A}^{-1}$ up to permutation and scaling of its rows.

When the sources are sparse the mixtures have special structure corresponding to overlaid lines on a scatter plot. For sources of interest in practice (voice, music) a sparse representation can often be achieved by a transformation into a suitable basis such as such as a short-time Fourier, Gabor, or Wavelet basis. The line orientations correspond to the columns of the mixing matrix $\mathbf{A}$, so if the lines can be estimated from the data then an estimate of the mixing matrix can be trivially constructed.

An algorithm for identification of radial line orientation and line separation is presented in Section 2. The application of the algorithm to blind source separation (BSS) of speech signals in both the even-determined and under-determined cases, along with experimental results including empirical assessments of robustness to noise, are presented in Section 3.

2 Oriented Lines Separation

2.1 Determining Line Orientation Using Data Covariance

The orientation of a linear cloud of data corresponds to the principal eigenvector of its covariance matrix [1, pages 125-132]. In order to identify multiple lines within a scatter plot, we *soft assign* data into M classes corresponding to the elements of the mixture, represented by orientation vectors $\mathbf{v}_i$ (eq. 1). This calculation corresponds to the *Expectation* step of an EM algorithm [2]. The covariance matrix is then calculated for the data associated with each class (eq. 2) and the principal eigenvector of the matrix is used as the new line orientation vector estimate (eq. 4), in the *Maximisation* step of our EM algorithm. This process is iterated until convergence, at which point the estimated mixing matrix $\hat{\mathbf{A}}$ is constructed by adjoining the estimated line orientations to form the columns of the matrix (eq. 5). We initialised the line orientations uniformly by normalising samples from an N-dimensional zero-mean spherical Gaussian.

2.2 Data Point Separation

For the even-determined case ($N = M$) the estimated mixing matrix $\hat{\mathbf{A}}$ is square and the sensor data can be converted to sources using its inverse. When $N < M$, the under-determined case, $\mathbf{A}$ is not invertible so the sources need to be estimated by some other means. To this end, we assume the source coefficients are sparse. One appropriate technique is the *hard assignment* of coefficients to sources using a mask [3, 4]. Another is *soft assignment*, in which each coefficient is decomposed into more than one source. This is generally done by minimisation of the L_1-norm, which can be seen as a maximum likelihood reconstruction under the assumption that the coefficients are drawn from a distribution of the form $p(c) \propto \exp -|c|$, *i.e.* a Laplacian [5, 6].

2.3 Algorithm Summary

We present an algorithm called *Soft-LOST*, for *Line Orientation Separation Technique*. The "soft" indicates that data points are partially assigned to lines by weighted each line's proximity. (A discussion of hard and soft assignments is presented by Kearns et al. [7].) Separation is achieved by first using a soft line orientation estimation subroutine.

soft line orientation estimation

1. Randomly initialise the M line orientation vectors $\mathbf{v}_i$.
2. Partially assign each data point $\mathbf{d}_j$, where $\mathbf{d}_j = \mathbf{x}(j)$, to each line orientation vector using a soft data assignment

$$z_{ij} = \|\mathbf{d}_j - (\mathbf{v}_i \cdot \mathbf{d}_j)\,\mathbf{v}_i\|^2$$
$$\hat{z}_{ij} = \frac{e^{-\beta z_{ij}}}{\sum_{i'} e^{-\beta z_{i'j}}} \tag{1}$$

where β controls the softness of the boundaries between the regions attributed to each line and $\hat{z}_{ij}$ is the magnitude of the assignment of data point j to line i.

3. Determine the new line orientation estimate by calculating the principal eigenvector of the covariance matrix. The covariance matrix expression (with zero mean) and assignment weightings are combined as follows:

$$\mathbf{\Sigma}_i = \frac{\sum_j \hat{z}_{ij} \mathbf{d}_j \mathbf{d}_j^T}{\sum_j \hat{z}_{ij}} \tag{2}$$

where $\mathbf{\Sigma}_i$ is the covariance of weighted data associated with line i. The eigenvector decomposition of $\mathbf{\Sigma}_i$ is expressed as:

$$\mathbf{\Sigma}_i = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{U}_i^{-1} \tag{3}$$

The matrix $\mathbf{U}_i$ contains the eigenvectors of $\mathbf{\Sigma}_i$ and the diagonal matrix $\mathbf{\Lambda}_i$ contains it's associated eigenvalues $\lambda_i \ldots \lambda_N$. The new line orientation vector estimate is the principal eigenvector of $\mathbf{\Sigma}_i$ which is expressed as

$$\mathbf{v}_i = \mathbf{u}_{\max} \tag{4}$$

where $\mathbf{u}_{\max}$ is the principal eigenvector, the eigenvector whose eigenvalue is $\lambda_{\max}$.

Return to step 2 and repeat until the $\mathbf{v}_i$ converge.

4. After convergence, adjoin the line orientations estimates to form the estimated mixing matrix.

$$\hat{\mathbf{A}} = [\mathbf{v}_1 | \cdots | \mathbf{v}_M] \tag{5}$$

Soft-LOST line separation algorithm

1. Perform *soft line orientation estimation* to calculate $\hat{\mathbf{A}}$.
2. For the even-determined case data points are assigned to line orientations using $\mathbf{s}(t) = \hat{\mathbf{A}}^{-1}\, \mathbf{x}(t)$. For the under-determined case, calculate coefficients $\mathbf{c}_j$ using linear programming for each data point j such that

$$\min_{\mathbf{c}_j} \|\mathbf{c}_j\|_1 \text{ subject to } \hat{\mathbf{A}}\mathbf{c_j} = \mathbf{d_j}$$

The resultant $\mathbf{c}_j$ coefficients, properly arranged, constitute the estimated linear subspaces, $\hat{\mathbf{S}} = [\mathbf{c}_1 | \cdots | \mathbf{c}_T]$.

3. The final result is a $M \times T$ matrix $\hat{\mathbf{S}}$ that contains the line orientation data sets in each row.

3 Experimental Results

The Soft-LOST algorithm was used for a blind source separation problem, where source attenuation vectors correspond to linear subspaces. The Soft-LOST solution to BSS is presented as follows

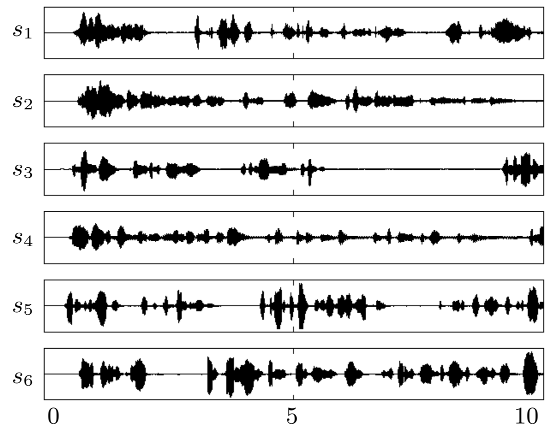

Fig. 1. Ten-second clips of six acoustic sources. Sound wave pressure is plotted against time, in seconds (see Appendix A.).

Soft-LOST for BSS

1. A $N \times T$ data matrix $\mathbf{X}(t)$ is composed of sensor observations of N instantaneous mixtures. The data is transformed into a sparse representation, $\mathbf{X}(t) \mapsto \mathbf{X}(\omega)$.
2. The Soft-LOST algorithm is performed on the data $\mathbf{X}(\omega)$. The algorithm estimates a mixing matrix, which in turn allows sources to be estimated from the mixtures via L_1-norm optimisation.
3. The resultant $M \times T$ matrix $\hat{\mathbf{S}}(\omega)$ contains in its rows the M estimated sources $\hat{\mathbf{s}}_1, \ldots, \hat{\mathbf{s}}_M$. These estimates are then transformed back into the time domain, $\hat{\mathbf{S}}(\omega) \mapsto \hat{\mathbf{S}}(t)$.

3.1 Experimental Methods

The Signal-to-Noise Ratios of the estimated sources $\hat{\mathbf{s}}_i$ (in dB) are used to measure the performance of the algorithm, $\mathrm{SNR}_i = 20 \log_{10}(\|\mathbf{s}_i\| / \|\hat{\mathbf{s}}_i - \mathbf{s}_i\|)$.

Speech signals (see Figure 1 and Appendix A) were transformed using a 512-point windowed FFT and the real coefficients were used to create a scatter plot. The experiments were coded for Matlab 6.5.0 and run on a 3.06 GHz Intel Pentium-4 based computer with 768MB of RAM. Experiments for the under-determined case typically took 35 minutes while the tests for the even-determined case ran for less than six minutes, depending on the number of iterations required for convergence. For comparison, the potential performance given a perfect estimate of $\mathbf{A}$ was also evaluated. In these experiments the line orientation estimation phase is skipped and the L_1-norm minimisation phase is tested separately. In general the better defined the line orientations in the scatter plot, the more accurate the source estimates. Experiments were performed for

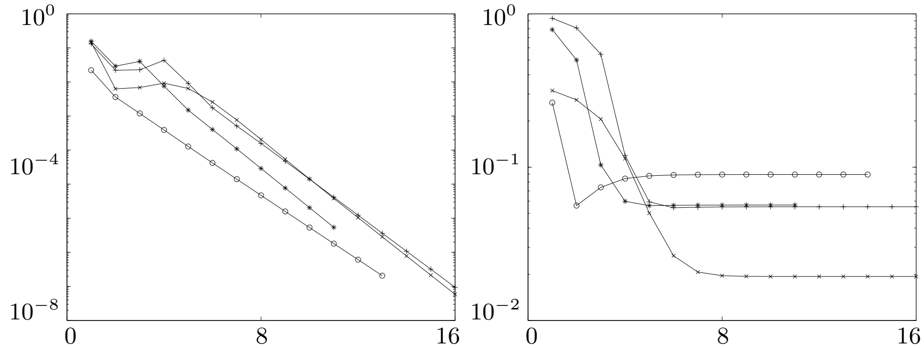

Fig. 2. Convergence plots of estimated mixing matrices for the following experiments; 5 mixtures 6 Sources ($*$), 5 mixtures 5 sources ($+$), 4 mixtures 5 sources ($\circ$) and 2 mixtures 3 sources ($\times$). On the left is the difference between consecutive estimates $\|\hat{\mathbf{A}}_l - \hat{\mathbf{A}}_{l-1}\|$, while the right is the difference between the true mixing matrix and the current estimate, $\|\mathbf{A}_{\text{orig}} - \hat{\mathbf{A}}_l\|$. The x axis of each plot is in units of algorithm iterations l.

a range of different values of N and M, and the parameter β was varied on an *ad hoc* basis.

3.2 Results

Results are presented for a total of 15 experiments. Data on the number of mixtures, sources used, and the value of the parameter β are contained in the tables of results. Results in Tables 1 and 2 demonstrate the effectiveness of the algorithm for the even-determined case. Experiments for testing line separation using L_1-norm minimisation were performed and their results are presented in Table 3. These experiments evaluate the effectiveness of the separation phase of the Soft-LOST algorithm in the under-determined case, and provide a benchmark for the subsequent experiments. Results for experiments that test both line orientation estimation and line separation in the under-determined case are presented in Tables 4 and 5. The Soft-LOST algorithm was tested for robustness to noise. Gaussian noise of various calibrated intensities was added to the signals in the experiments in Table 6. These results, when contrasted with those previously presented, measure the algorithm's empirical robustness to noise.

These experimental results demonstrate that the Soft-LOST algorithm is an effective technique for BSS in both the even-determined and under-determined cases, even in the presence of noise. A convergence plot is shown in Figure 2.

4 Conclusion

The results presented demonstrate that the identification of line orientations using a modified EM procedure is an effective method for determining the mixing matrix of a set of linear mixtures. It has been demonstrated that once the mixing matrix is found, sources can then be separated by minimising the L_1-norm between the data point being considered and the line orientations represented by the columns of the mixing matrix. The Soft-LOST algorithm provides a good solution to blind source separation of instantaneous mixtures even when there are fewer sensors than sources. The experiments presented are concerned with the specific problem of blind source separation of speech signals, however the results can be applied to any situation involving a mixture of oriented lines.

This work extends previous research in which we developed a modified k-means algorithm called *Hard-LOST* [8]. The Soft-LOST results presented here can be contrasted with those of Hard-LOST. In future work, we plan to modify the L_2 norm of the line distance calculation to use the covariance matrix of each line, and to partition the coefficients into classes exhibiting different noise levels to allow optimal combination of evidence using such a noise-sensitive measure.

Acknowledgements

Supported by Higher Education Authority of Ireland (An tÚdarás Um Ard-Oideachas), and Science Foundation Ireland grant 00/PI.1/C067.

References

[1] Juha Karhunen Aapo Hyvärinen and Erkki Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1976.

[3] S. T. Rickard and F. Dietrich. DOA estimation of many W-disjoint orthogonal sources from two mixtures using DUET. In *Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing (SSAP2000)*, pages 311–314, Pocono Manor, PA, August 2000.

[4] Sam T. Roweis. One microphone source separation. In *Advances in Neural Information Processing Systems 13*, pages 793–799. MIT Press, 2001.

[5] T.-W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing Letters*, 6(4):87–90, 1999.

[6] Michael Zibulevsky and Barak A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13(4):863–882, April 2001.

[7] M. Kearns, Y. Mansour, and A. Y. Ng. An information-theoretic analysis of hard and soft assignment methods for clustering. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 282–293, 1997.

[8] Paul D. O'Grady and Barak A. Pearlmutter. Hard-LOST: Modified k-means for oriented lines. In *Proceedings of the Irish Signals and Systems Conference*, Belfast, June 30–July 2 2004.

[9] Elise Paschen, Charles Osgood, and Rebekah Presson Mosby, editors. *Poetry Speaks: Hear Great Poets Read Their Work from Tennyson to Plath.* Sourcebooks Incorporated, 2001. ISBN 1570717206.

A Source Signals

The source signals were taken from *Poetry Speaks*, a commercial audio CD of poems read by their authors [9]. Audio CD data is recorded as uncompressed 44.1 kHz 16-bit stereo waveforms. Prior to further processing ten-second clips were extracted, the two signal channels were averaged, and the data was down-sampled to 8 kHz. The scale of the audio data is arbitrary, leading to the arbitrary units on auditory waveform samples throughout the manuscript.

s_1 *Coole Park and Ballylee*, by William Butler Yeats.
s_2 *The Lake Isle of Innisfree*, by William Butler Yeats.
s_3 *Among Those Killed in the Dawn Raid Was a Man Aged a Hundred*, by Dylan Thomas.
s_4 *Fern Hill*, by Dylan Thomas.
s_5 *Ave Maria*, by Frank O'Hara.
s_6 *Lana Turner Has Collapsed*, by Frank O'Hara.

Table 1. Two Mixtures and Two Sources

Mixtures	Sources	β	SNR (dB)	
2	s_1 s_2	1.5	35.28	43.90
2	s_3 s_4	1.5	41.24	63.32
2	s_5 s_6	1.5	40.30	39.17

Table 2. Five Mixtures and Five Sources

Mixtures	Sources	β	SNR (dB)		
5	s_1 s_2 s_3	6	27.76	24.06	28.31
	s_4 s_5		26.08	28.67	
5	s_1 s_2 s_3	6.6	27.95	24.15	28.41
	s_4 s_5		26.2	28.77	
5	s_1 s_2 s_3	5.5	27.54	23.96	28.18
	s_4 s_5		25.94	28.54	

Table 3. L_1-Norm and True Mixing Matrix

Mixtures	Sources	SNR (dB)
2	s_1 s_2 s_3	10.41 15.64 7.75
5	s_1 s_2 s_3	20.85 20.62 19.10
	s_4 s_5 s_6	17.08 21.93 48.96

Table 4. Two Mixtures and Three Sources

Mixtures	Sources	β	SNR (dB)
2	s_1 s_2 s_3	2	10.43 15.58 7.87
2	s_1 s_2 s_3	1.5	10.43 15.58 7.87

Table 5. Five Mixtures and Six Sources

Mixtures	Sources	β	SNR (dB)
5	s_1 s_2 s_3	6.5	20.17 19.85 18.88
	s_4 s_5 s_6		16.66 21.09 32.19
5	s_1 s_2 s_3	6	20.15 19.83 18.87
	s_4 s_5 s_6		16.65 21.08 32.21

Table 6. Additive Gaussian Noise

Mixtures	Sources	Noise (dB)	SNR (dB)
2	s_1 s_2	5	34.75 40.05
5	s_1 s_2	15	26.58 23.45
	s_3 s_4		27.02 25.41
	s_5		27.14
5	s_1 s_2	15	17.03 16.78
	s_3 s_4		15.79 13.82
	s_5 s_6		18.27 31.59