Chapter 3

Blind and semi-blind source separation

Erkki Oja, Juha Karhunen, Alexander Ilin, Zhirong Yang, Jaakko Luttinen, He Zhang, Jarkko Ylipaavalniemi, Tele Hao

3.1 Introduction

Erkki Oja

What is Blind and Semi-blind Source Separation? Blind source separation (BSS) is a class of computational data analysis techniques for revealing hidden factors, that underlie sets of measurements or signals. BSS assumes a statistical model whereby the observed multivariate data, typically given as a large database of samples, are assumed to be linear or nonlinear mixtures of some unknown latent variables. The mixing coefficients are also unknown.

By BSS, these latent variables, also called sources or factors, can be found. Thus BSS can be seen as an extension to the classical methods of Principal Component Analysis and Factor Analysis. BSS is a much richer class of techniques, however, capable of finding the sources when the classical methods, implicitly or explicitly based on Gaussian models, fail completely.

In many cases, the measurements are given as a set of parallel signals or time series. Typical examples are mixtures of simultaneous sounds or human voices that have been picked up by several microphones, brain signal measurements from multiple EEG sensors, several radio signals arriving at a portable phone, or multiple parallel time series obtained from some industrial process. But BSS has other applications as well: it turns out that clustering, or finding mutually similar subsets of the dataset, can also be addressed as a linear source separation problem, when suitable constraints are added to the model. This also applies to the related problem of graph partitioning.

Perhaps the best known single methodology in BSS is Independent Component Analysis (ICA), in which the latent variables are nongaussian and mutually independent. However, also other criteria than independence can be used for finding the sources. One such simple criterion is the non-negativity of the sources. Sometimes more prior information about the sources is available or is induced into the model, such as the form of their probability densities, their spectral contents, etc. Then the term "blind" is often replaced by "semiblind".

Our earlier contributions in ICA research. In our ICA research group, the research stems from some early work on on-line PCA, nonlinear PCA, and separation, that we were involved with in the 80's and early 90's. Since mid-90's, our ICA group grew considerably. This earlier work has been reported in the previous Triennial and Biennial reports of our laboratory from 1994 to 2009 [1]. A notable achievement from that period was the textbook "Independent Component Analysis" by A. Hyvärinen, J. Karhunen, and E. Oja [2]. It has been very well received in the research community; according to the latest publisher's report, over 6000 copies had been sold by August, 2011. The book has been extensively cited in the ICA literature and seems to have evolved into the standard text on the subject worldwide. In Google Scholar, the de facto standard for citations in the ICT field, the book has received over 6700 citations (April 2012). In 2005, the Japanese translation of the book appeared (Tokyo Denki University Press), and in 2007, the Chinese translation (Publishing House of Electronics Industry).

Another tangible contribution has been the public domain FastICA software package [3]. This is one of the few most popular ICA algorithms used by the practitioners and a standard benchmark in algorithmic comparisons in ICA literature.

In the reporting period 2010 - 2011, ICA/BSS research stayed as one of the core projects in the laboratory, with the pure ICA theory waning and being replaced by several new directions in blind and semiblind source separation. In this Chapter, we present two such novel directions.

Section 3.2 introduces some theoretical advances on Nonnegative Matrix Factorization undertaken during the reporting period, especially on the new Projective Nonnegative Matrix Factorization (PNMF) principle, which is a principled way to perform approximate nonnegative Principal Component Analysis.

Section 3.3 introduces novel results in finding independent and dependent sources from two related data sets. It is based on a combination of Canonical Correlation Analysis and ICA.

Quite another way to formulate the BSS problem is Bayesian analysis. This is covered in the separate Chapter **??**.

References

[1] Triennial and Biennial reports of CIS and AIRC. `http://www.cis.hut.fi/research/reports/`.

[2] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. J. Wiley, 2001.

[3] The FastICA software package. `http://www.cis.hut.fi/projects/ica/fastica/`.

3.2 Non-negative Low-Rank Learning

Zhirong Yang, He Zhang, Zhanxing Zhu, and Erkki Oja

Enforcing nonnegativity in linear factorizations [4] has proven to be a powerful principle for multivariate data analysis, especially sparse feature analysis, as shown by the well-known Nonnegative Matrix Factorization (NMF) algorithm by Lee and Seung [2]. Their method minimizes the difference between the data matrix $\mathbf{X}$ and its non-negative decomposition $\mathbf{WH}$. Yuan and Oja [11] proposed the Projective NMF (PNMF) method which replaces $\mathbf{H}$ in NMF with $\mathbf{W}^T\mathbf{X}$. Empirical results indicate that PNMF is able to produce more spatially localized, part-based representations of visual patterns.

Recently, we have extended and completed the preliminary work with the following new contributions [5]: (1) formal convergence analysis of the original PNMF algorithms, (2) PNMF with the orthonormality constraint, (3) nonlinear extension of PNMF, (4) comparison of PNMF with two classical and two recent algorithms [10, 1] for clustering, (5) a new application of PNMF for recovering the projection matrix in a nonnegative mixture model, (6) comparison of PNMF with the approach of discretizing eigenvectors, and (7) theoretical justification of moving a term in the generic multiplicative update rule. Our in-depth analysis shows that the PNMF replacement has positive consequences in sparseness of the approximation, orthogonality of the factorizing matrix, decreased computational complexity in learning, close equivalence to clustering, generalization of the approximation to new data without heavy re-computations, and easy extension to a nonlinear kernel method with wide applications for optimization problems. We have later demonstrated that combining orthogonality and negativity works well in graph partitioning [3].

In NMF, the matrix difference was originally measured by the Frobenius matrix norm or the unnormalized Kullback-Leibler divergence (I-divergence). Recently we have significantly extended NMF to a much larger variety of divergences with theoretically convergent algorithms. In [7], we have presented a generic principle for deriving multiplicative update rules, as well as a proof of the convergence of their objective function, that applies for a large variety of linear and quadratic NMF problems. The proposed principle only requires that the NMF approximation objective function can be written as a sum of a finite number of monomials, which is a mild assumption that holds for many commonly used approximation error measures. As a result, our method turns the derivation, which seemingly requires intense mathematical work, into a routine exercise that could be even readily automated using symbolic mathematics software. In our practice [8], both theoretical and practical advantages indicate that there would be good reasons to replace the I-divergence with normalized Kullback-Leibler for NMF and its variants. The PNMF method can also be generalized to the α-divergence family [6].

Automatic determination of the low-rank in NMF is a difficult problem. In [9], we have presented a new algorithm which can automatically determine the rank of the projection matrix in PNMF. By using Jeffrey's prior as the model prior, we have made our algorithm free of human tuning in finding algorithm parameters. Figure 3.1 visualizes the learned basis of the *Swimmer* dataset.

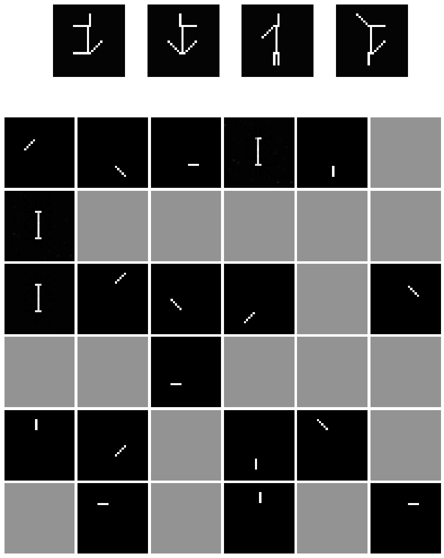


Figure 3.1: (Top) Some sample images of Swimmer dataset; (Bottom) 36 basis images of Swimmer dataset. The gray cells correspond to matrix columns whose L_2-norms are zero or very close to zero.

References

[1] I. Dhillon, Y. Guan, and B. Kulis. Kernel kmeans, spectral clustering and normalized cuts. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 551–556, Seattle, WA, USA, 2004.

[2] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[3] E. Oja and Z. Yang. Orthogonal nonnegative learning for sparse feature extraction and approximate combinatorial optimization. *Frontiers of Electrical and Electronic Engineering in China*, 5(3):261–273, 2010.

[4] P. Paatero P and U. Tapper. Positive matrix factorization: A nonnegative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.

[5] Z. Yang and E. Oja. Linear and nonlinear projective nonnegative matrix factorization. *IEEE Transaction on Neural Networks*, 21(5):734–749, 2010.

[6] Z. Yang and E. Oja. Projective nonnegative matrix factorization based on α-divergence. *Journal of Artificial Intelligence and Soft Computing Research*, 1(1):7–16, 2011.

[7] Z. Yang and E. Oja. Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization. *IEEE Transactions on Neural Networks*, 22(12):1878–1891, 2011.

[8] Z. Yang, H. Zhang, Z. Yuan, and E. Oja. Kullback-leibler divergence for nonnegative matrix factorization. In *Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN2011)*, pages 250–257, Espoo, Finland, 2011.

[9] Z. Yang, Z. Zhu, and E. Oja. Automatic rank determination in projective nonnegative matrix factorization. In *Proccedings of the 9th International Conference on Latent Variable Analysis and Signal Separation (LVA 2010)*, pages 514–521, St. Malo, France, 2010.

[10] S. X. Yu and J. Shi. Multiclass spectral clustering. In *Proceedings of the 9th IEEE International Conference on Computer Vision*, volume 2, pages 313–319, 2003.

[11] Z. Yuan and E. Oja. Projective nonnegative matrix factorization for image compression and feature extraction. In *Proceedings of 14th Scandinavian Conference on Image Analysis (SCIA)*, pages 333–342, Joensuu, Finland, 2005.

3.3 Finding dependent and independent sources from two related data sets

Juha Karhunen, Tele Hao, and Jarkko Ylipaavalniemi

We have considered in two papers [3, 4] extension of independent component analysis (ICA) and blind source separation (BSS) for separating mutually dependent and independent components from two different but related data sets. This problem is important in practice, because such data sets are commonplace in real-world applications. We propose a new method which first uses canonical correlation analysis (CCA) [2] for detecting subspaces of independent and dependent components. The data sets are then mapped onto these subspaces. Even plain CCA can provide a coarse separation in simple cases, and we can justify this. Better separation results are obtained by applying some suitable ICA or BSS method [1] to the mapped data sets. These methods can utilize somewhat different properties of the data such as non-Gaussianity, temporal correlatedness, or nonstationary depending on the characteristics of the data.

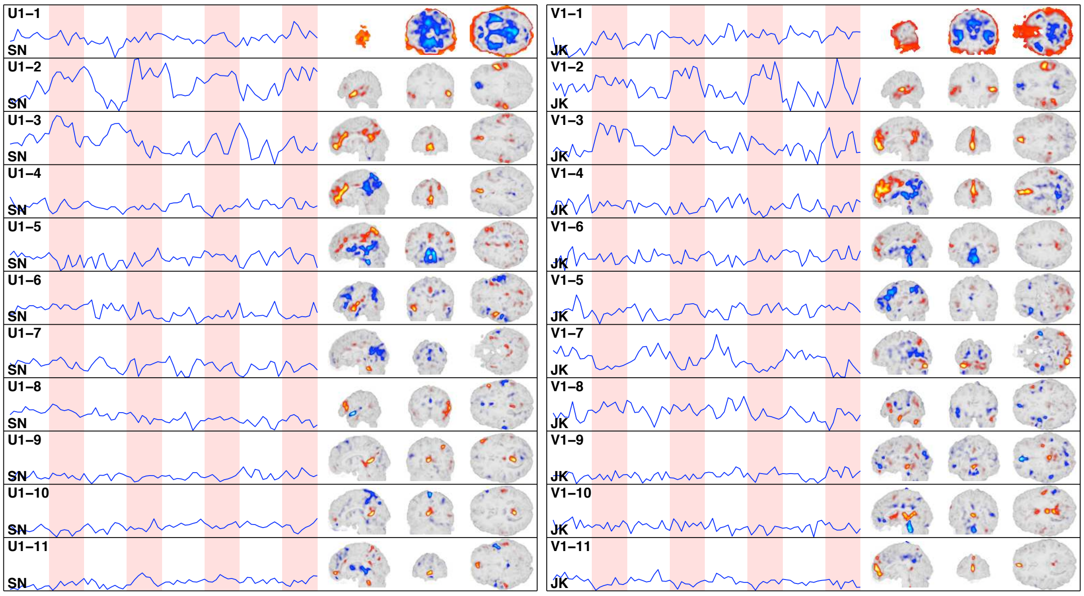
The proposed method is straightforward to implement and computationally not too demanding. CCA preprocessing improves often quite markedly the separation results of the chosen ICA or BSS method especially in difficult separation problems. Not only are the signal-to-noise ratios of the separated sources clearly higher, but CCA also helps a method to separate sources that it alone is not able to separate. In [3, 4], we present experimental results for several well-known ICA and BSS methods for synthetically constructed source signals [5] which are quite difficult to separate for most ICA and BSS methods. Furthermore, we have applied our method successfully to real-world robot grasping data in [3].

In [4], we tested the usefulness of our method with data taken from a functional magnetic resonance imaging (fMRI) study [6], where it is described in more detail. We used the measurements of two healthy adults while they were listening to spoken safety instructions in 30 s intervals, interleaved with 30 s resting periods. In these experiments we used slow feature analysis (SFA) [7] for post-processing the results given by CCA, because it gave better results than the most widely used standard ICA method FastICA [1].

Fig. 3.2 shows the results of applying our method to the two datasets and separating 11 components from the subspaces of dependent components. The consistency of the components across the subjects is quite good. The first component shows a global hemodynamic contrast, that may also be related to artifacts originating from smoothing the data in the standard preprocessing. The activity of the second component is focused on the primary auditory cortices. The third and fourth components show both positively and negatively task-related activity around the anterior and posterior cingulate gyrus. These first results are promising and in good agreement with the the ones reported in [6].

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, 2001.



(a) (b)

Figure 3.2: Experimental results with fMRI data. Each row shows one of the 11 separated components. The activation time-course with the stimulation blocks for reference, shown on the left, and the corresponding spatial pattern on three coincident slices, on the right. Components from (a) the first and (b) the second dataset.

[2] A. Rencher. *Methods of Multivariate Analysis, 2nd ed.*. Wiley, 2002.

[3] J. Karhunen and T. Hao. Finding dependent and independent components from two related data sets. In *Proc. of the IEEE Int. Joint Conf. on Neural Networks (IJCNN 2011)*, pp. 457–466, San Jose, California, USA, August 2011.

[4] J. Karhunen, T. Hao, and J. Ylipaavalniemi. A canonical correlation analysis based method for improving BSS of two related data sets. To appear in *Proc. of the 10th Int. Conf. on Latent Variable Analysis and Signal Separation (LVAICA 2012)*, Tel-Aviv, Israel, March 2012.

[5] A. Hyvärinen. A unifying model for blind separation of independent sources. *Signal Processing*, 85(7):1419–1427, 2005.

[6] J. Ylipaavalniemi et al. Analyzing consistency of independent components: An fMRI illustration. *NeuroImage*, 39:169–180, 2008.

[7] L. Wiskott and T. Sejnowski. Slow feature analysis: unsupervised learning of invariances. *Neural Computation*, 14:715–770, 2002.