Maximization of Component Disjointness: A Criterion for Blind Source Separation

Jörn Anemüller

Medical Physics Section
Dept. of Physics
Carl von Ossietzky University Oldenburg
26111 Oldenburg, Germany

Abstract. Blind source separation is commonly based on maximizing measures related to independence of estimated sources such as mutual statistical independence assuming non-Gaussian distributions, decorrelation at different time-lags assuming spectral differences or decorrelation assuming source non-stationarity.

Here, the use of an alternative model for source separation is explored which is based on the assumption that sources emit signal energy at mutually different times. In the limiting case, this corresponds to only a single source being "active" at each point in time, resulting in mutual disjointness of source signal supports and *negative* mutual correlations of source signal envelopes. This assumption will not be fulfilled perfectly for real signals, however, by maximizing disjointness of estimated sources (under a linear mixing/demixing model) we demonstrate that source separation is nevertheless achieved when this assumptions is only partially fulfilled.

The conceptual benefits of the disjointness assumption are that (1) in certain applications it may be desirable to explain observed data in terms of mutually disjoint "parts" and (2) the method presented here preserves the special physical information assigned to amplitude zero of a signal which corresponds to the absence of energy (rather than subtracting the signal mean prior to analysis which for non zero-mean sources destroys this information).

The method of *disjoint component analysis* (DCA) is derived and it is shown that its update equations bear remarkable similarities with maximum likelihood independent component analysis (ICA). Sources with systematically varied degrees of disjointness are constructed and processed by DCA and Infomax and Jade ICA. Results illustrate the behaviour of DCA and ICA under these regimes with two main results: (1) DCA leads to a higher degree of separation than ICA, (2) DCA performs particularly well on positive-valued sources as long as they are at least moderately disjoint, and (3) The performance peak of ICA for zero-mean sources is achieved when sources are disjoint (but not independent)[1].

1 Introduction

Representation of measured data in terms of a number of generating causes or underlying "sources" is an important problem that has gained widespread attention in recent

[1] This research was supported by the EC under the DIRAC integrated project IST-027787.

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 325–332, 2007.

years, either with the goal of extracting known-to-exist sources from measurements (blind source separation), or in order to find an efficient—possibly lower-dimensional—description of given data (exploratory data analysis).

We propose and investigate a novel technique, “disjoint component analysis" (DCA) that is based on the goal of extracting components with maximally disjoint support from given data, i.e., it is sought to describe the data in terms of components of which as few as possible should be activated at any single time (or sample) point. Ideally, only a single source process would account for a single sample of measured data. Since this goal is too strong for real-world data, we demonstrate that it can be significantly relaxed while still retaining the beneficial characteristics of the method.

Disjoint support between generating source processes may constitute a relevant general principle in domains where other assumptions, e.g., statistical independence and the implied effective physical separation of generating source processes, have to be postulated or justified post-hoc rather than deduced a-priori. In some cases such as communicating speakers or densely interconnected nervous cells in the brain, theoretical considerations argue in favor of dependencies between source processes. Even though such dependencies might turn out to be largely negligible in some domains, it does appear to be worthwhile to consider the implications of incorporating such dependencies into the models.

In the opposite direction (and with a different intention than ours), some authors have argued that sources that are often regarded as independent can effectively be modeled as being “w-disjoint orthogonal" [10]. We are demonstrating a close formal link between algorithms derived from striving for independent and disjoint representations, respectively, which may be seen as an indication that both notions may contain similarities that we have not yet fully appreciated.

In relation to existing techniques, DCA differs from ICA [2,4] since the disjointness assumption corresponds to a source model with *dependent* sources. Sparse-coding approaches [9], unlike DCA, impose a sparse prior on each source but do not incorporate a mutual disjointness of sources. Non-negative matrix factorisation approaches [6] and l1-norm minimization methods [5] aim to obtain a parts-based description of the data with fundamentally different algorithms than DCA.

2 Disjoint Component Analysis

2.1 Derivation of Algorithm

Consider N observed signals $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ which are assumed to be generated from N underlying sources $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^T$ by multiplication with a mixing system $\mathbf{A}$ as

$$\mathbf{x}(t) = \mathbf{A}\,\mathbf{s}(t) \tag{1}$$

It is sought to linearly transform the observations by a matrix $\mathbf{W}$ to obtain output signals

$$\mathbf{y}(t) = \mathbf{W}\,\mathbf{x}(t) \tag{2}$$

with components $\mathbf{y}(t) = [y_1(t), \ldots, y_N(t)]^T$. When source reconstruction is desired, these should resemble the sources up to arbitrary rescaling and permutation. When an

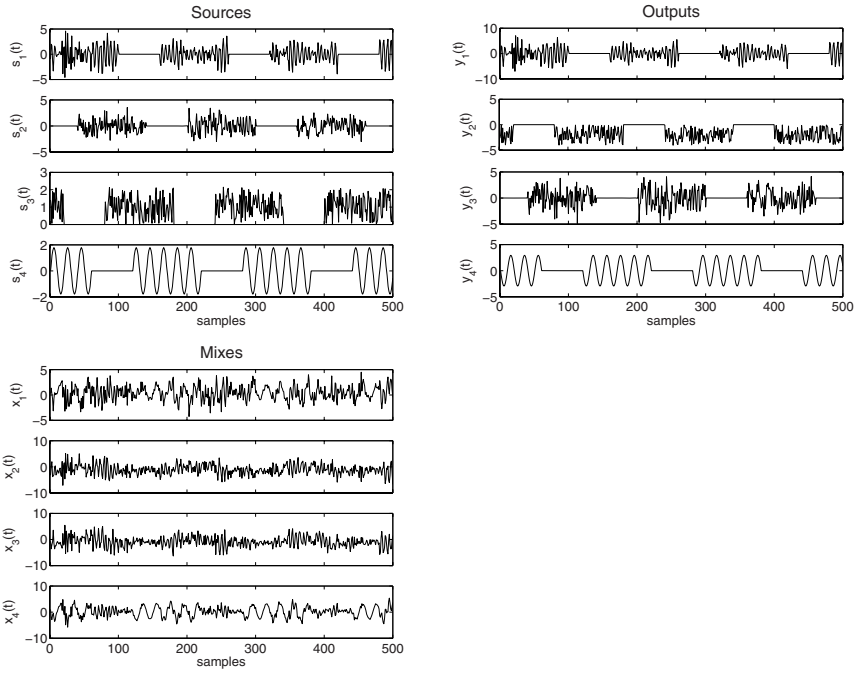

Fig. 1. Disjoint component analysis of four sources (top left) which are not strictly disjoint but exhibit significant overlap. Sources were mixed with a randomly chosen 4×4 mixing matrix to yield observation signals (bottom left) which were successfully separated into the original sources up to arbitrary permutation, rescaling and sign flip (top right) using DCA.

exploratory data analysis view is adopted, the output signals should convey a signal representation that is meaningful in some to-be-specified sense.

A central notion in our approach is the overlap between two output signals y_i and y_j which we define as[2]

$$o_{ij} = E(|y_i|\,|y_j|), \tag{3}$$

where $E(\cdot)$ denotes expectation and sample index t is omitted where convenient. With $o_{ij} \geq 0$ and $o_{ij} = 0$ if and only if $y_i(t)\,y_j(t) = 0$ for all t and $i \neq j$, two signals y_i and y_j have *disjoint support* if $o_{ij} = 0$. In this case, y_i and y_j are called *disjoint*, i.e., at most one of the signals is non-zero at any time.

For strictly disjoint source signals $\mathbf{s}(t)$ and a non-singular matrix $\mathbf{A}$, strictly disjoint outputs can be obtained that resemble the sources up to arbitrary permutation and rescaling. Note that in this case sources are not mutually *independent* but exhibit statistical *dependencies* through the negative correlations of their signal envelopes or signal power time-courses.

[2] Different definitions of the overlap, involving other non-linear functions of the output signals, are possible but beyond the scope of the present paper.

While it is not possible in general to linearly transform an arbitrary signal $\mathbf{x}(t)$ into a signal $\mathbf{y}(t)$ with only disjoint components, finding minimally overlapping outputs is a natural goal as it corresponds to a signal description in terms of processes out of which only a small number is active at any given time. A natural choice to obtain maximally disjoint, minimally overlapping output signals is minimization of the function

$$H = \frac{1}{2}\sum_{i\neq j} o_{ij} = \frac{1}{2}\sum_{i\neq j} E(|y_i|\,|y_j|) \tag{4}$$

The global minimum $H = 0$ is attained only for strictly disjoint signals where for all t any signal $y_i(t) \neq 0$ if and only if $y_j(t) = 0$ for all $j \neq i$. Substituting 2 into 4, the partial derivatives are given by

$$\frac{\partial H}{\partial w_{ij}} = E\Big(\mathrm{sign}(y_i)\, x_j \sum_{k\neq i} |y_k|\Big) \tag{5}$$

which in matrix notation is easily rewritten as

$$\nabla H = E\left(-\mathbf{y}\mathbf{x}^H + ||\mathbf{y}||_1 \mathrm{sign}(\mathbf{y})\mathbf{x}^H\right) \tag{6}$$

where $||\mathbf{y}||_1 = \sum_i |y_i|$ denotes the 1-norm of $\mathbf{y}$.

Right-multiplication with $\mathbf{W}^T\mathbf{W}$ yields an expression similar to the natural gradient ICA algorithm of [1],

$$\tilde{\nabla} H = E\left(-\mathbf{y}\mathbf{y}^H + ||\mathbf{y}||_1 \mathrm{sign}(\mathbf{y})\mathbf{y}^H\right)\mathbf{W}. \tag{7}$$

Gradients (6) and (7) are similar to the corresponding gradients derived from infomax or maximum-likelihood ICA with a sparse prior, however, we emphasize that the mean has not been removed from the output signals (i.e., source estimates) $\mathbf{y}(t)$.

Without constraints the gradients converge to the trivial solution $\mathbf{W} = \mathbf{0}$. To remove the scaling ambiguity each row $\mathbf{w}_i$ of matrix $\mathbf{W}$ is fixed to unit-norm $||\mathbf{w}_i||_2 = 1$. Hence, each row $\boldsymbol{\Delta}_i$ of ∇H is projected according to

$$\boldsymbol{\Delta}_i^{\perp} = \boldsymbol{\Delta}_i - (\boldsymbol{\Delta}_i^H \mathbf{w}_i)\,\mathbf{w}_i \tag{8}$$

resulting in the projected gradient matrix $\boldsymbol{\Delta}^{\perp}$ that is then used for gradient descent. The final update rule for matrix $\mathbf{W}$ with a step size of η is

$$\mathbf{W} \leftarrow \mathbf{W} - \eta\,\boldsymbol{\Delta}^{\perp} \tag{9}$$

for the ordinary gradient (6) and similarly for (7). Periodic row re-normalization of $\mathbf{W}$ is applied to keep it on the constraint manifold for non-infinitesimal η.

3 Evaluation

3.1 Synthetic Data Generation

Disjoint sources $s_i(t)$ are generated from mutually independent signals $\zeta_i(t)$ by multiplying them with disjoint masking functions $\mu_i(t) \in \{0, 1\}$ for all i, t and

$$s_i(t) = \mu_i(t)\,\zeta_i(t) \tag{10}$$

$$E(\mu_i\,\mu_j) = 0 \quad \text{if} \quad i \neq j \tag{11}$$

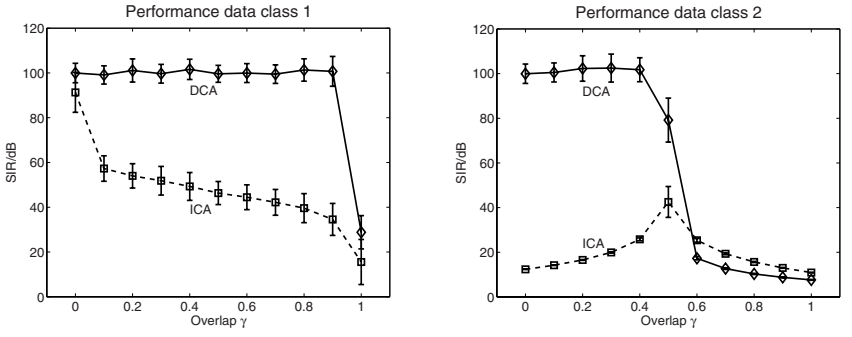

Fig. 2. Separation performance of DCA and ICA in terms of signal-to-interference ratio (SIR) in dB after separation. Performance is given for data class 1 (left panel, sources with positive and negative observation values) and data class 2 (right panel, sources with positive only observation values) as a function of overlap γ. A value of $\gamma = 0$ corresponds to strictly disjoint sources (statistical dependencies between sources through negative correlation of signal envelopes); $\gamma = 0.5$ corresponds to statistically independent sources; and $\gamma = 1.0$ corresponds to fully overlapping, not disjoint sources (statistical dependencies through positive correlation of signal envelopes). Mean and variance of performance for 100 separation runs, each with independently generated data, are given for each condition.

These sources may then be used to generate observations by multiplication with a matrix $\mathbf{A}$ according to Eq. 1.

Strictly disjoint sources with zero overlap are not expected to be an appropriate model for real data. Hence, sources with variable masker overlap γ_{ij}, which may depend on the source pair (i, j),

$$\gamma_{ij} = E(\mu_i \mu_j) \,/\, E(\mu_i^2) \tag{12}$$

with $E(\mu_i^2) = \text{const}$ for all i are also generated. In the experiments reported below masker overlap γ_{ij} is chosen such that a value of $\gamma_{ij} = 1$ corresponds to a source pair (s_i, s_j) exhibiting mutual statistical dependence through maskers with *positive* correlation. The value $\gamma_{ij} = 0$ corresponds to strictly disjoint sources that exhibit mutual statistical dependence through maskers with *negative* correlation. Finally, a value of $\gamma_{ij} = 0.5$ coincides with statistically *independent* sources (s_i, s_j) because of uncorrelated maskers (and statistically independent $\zeta_i(t)$).

The signal generation scheme was inspired by a functional magnetic resonance imaging (fMRI) experiment design [3].

3.2 Separation of Synthetic Sources

Four sources were generated according to the scheme described above, mixed with a randomly chosen mixing matrix and processed with the natural gradient disjoint component analysis algorithm (Eq. 7) with regularization (Eq. 8). The underlying mutually independent signals $\zeta_i(t)$ were chosen as a speech signal (ζ_1), i.i.d. noise from a normal

distribution with zero-mean and unit-variance (ζ_2), i.i.d. noise from a uniform distribution on the interval $[0, 1]$ (ζ_3), and a sine wave (ζ_4). The maskers $\mu_i(t)$ were chosen such that $\gamma_{ij} = 0.6$ for source pairs $(1, 2)$, $(2, 3)$, $(3, 4)$, $(1, 4)$, and $\gamma_{ij} = 0.4$ for source pairs $(1, 3)$, $(2, 4)$. Source signals, observed (mixed) signals and output signals are displayed in Fig. 1, demonstrating that the algorithm performs successful separation even though sources are not strictly disjoint but show significant overlap. Similarly, the algorithm successfully separates mixtures of four strictly disjoint sources with $\gamma_{ij} = 0$ for all $i \neq j$ (data not shown here).

3.3 Variable Degree of Overlap

The goal of this experiment was to systematically study the influence of the degree of overlap on the performance of the disjoint component analysis algorithm. Results are reported for the gradient version of the algorithm (Eq. 6) with regularization (Eq. 8). Results for the natural gradient version are virtually identical and not reported separately.

Sources were generated based on two different underlying signal classes. In the first part of the experiment ("data class 1"), two sources s_1 and s_2 were generated from ζ_1 and ζ_2 that were drawn as i.i.d. signals from a zero-mean and unit-variance normal distribution, hence containing positive and negative values. In the second part of the experiment ("data class 2"), ζ_1 and ζ_2 were chosen to be i.i.d. signals from a uniform distribution on the interval $[0, 1]$, hence containing only positive values.

For both data sets the single overlap parameter γ was varied from 0 (no overlap, source dependence through negative masker correlation) via 0.5 (50% overlap, statistically independent sources) to 1.0 (full overlap, source dependence through positive masker correlation) in steps of 0.1.

Hence, 11 data set conditions were generated for each of the two data classes. For each condition, disjoint component analysis was performed on 100 individual datasets drawn independently according to the description above. This resulted in a total of 2200 datasets each with 10000 samples for each of the two sources.

Fig. 2 shows the results with mean and variance of signal separation in dB signal-to-interference ratio (SIR) after separation separately for data class 1 (left panel) and data class 2 (right panel). For data class 1 with sources that adopt positive and negative values, DCA separation performance shows no significant dependence on the overlap parameter γ except (as expected) for complete overlap at $\gamma = 1$ where the algorithm essentially attempts to separate two i.i.d. normally distributed sources which is ill-posed. In all other cases of data class 1, DCA separation is excellent with about 100 dB SIR.

The results look different for data class 2 with positive only source values. Separation remains excellent for data sets with a small overlap ($0.0 \leq \gamma \leq 0.4$), with again about 100 dB SIR. In the case of independent sources at $\gamma = 0.5$, separation is still very good at 80 dB. Performance breaks down for large overlaps ($1.0 \geq \gamma \geq 0.6$), an effect which we attribute to the positivity of the sources.

3.4 Comparison with Independent Component Analysis

The same data generated for section 3.3 was re-analyzed with natural gradient infomax ICA [1,2] using the ICA toolbox [7,8] with logistic function non-linearity. For comparison, a simple gradient approach with fixed step size and sign function non-linearity

was also used and gave virtually identical results for data class 1. On data class 2, the fixed step gradient approach gave qualitatively similar results but was outperformed by the referenced ICA toolbox in terms of SIR separation performance. All source signals have been checked to have positive kurtosis. Processing of the same signal with the jade algorithm [4] gave virtually identical results.

Results in Fig. 2 show that in most cases ICA results in a poorer SIR than DCA. For data class 1, ICA shows excellent signal separation for strictly disjoint sources ($\gamma = 0.0$). Performance is significantly lower, though still good, for independent sources, which seems to stand in contradiction to the independence assumption. As expected, performance decreases towards sources with strong overlap ($\gamma = 1.0$).

For data class 2, ICA performs best when sources are independent ($\gamma = 0.5$) with a drop off in performance towards both lower and higher source overlaps, which is plausible due to ICA's independence assumption.

4 Conclusion

Disjoint component analysis (DCA) has been shown to yield good performance for strictly disjoint and moderately disjoint data sets. For data with high overlap between sources (weakly disjoint), performance depends on the specific type of data, with good performance for data sets with sources that take positive and negative observation values, and a break-down of performance in case of purely positive source data.

We have shown that under certain approximations DCA is closely related to independent component analysis (ICA), albeit both start from significantly different assumptions. The empirical algorithm evaluation showed a better separation performance for DCA than for ICA under most conditions. Interestingly, ICA produced the best performance not for statistically independent sources but for strictly disjoint ones (cf. also ICA 2006 oral presentation of I.C. Daubechies).

Results presented here appear to warrant a closer investigation of the differences and similarities of both algorithm classes. It would be desirable to gain experience with a wider range of synthetic and natural data than could be presented here. We are tempted to speculate that DCA might be appropriate in particular for analyzing data where the independence assumption is not strictly fulfilled, where a data representation in terms of disjoint components is preferable to independent components, and where signals are comprised of positive only measurement values. This could be the case, e.g., for brain signals such fMRI, for data from dialog speech signals, and for comparably short signal sequences where independence cannot be fully attained due to finite sample effects.

References

1. Amari, S.-I.: Natural gradient works efficiently in learning. Neural Computation 10, 251–276 (1998)
2. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159 (1995)
3. Benharrosh, M.S., Takerkart, S., Cohen, J.D., Daubechies, I.C., Richter, W.: Using ICA on fMRI: Does independence matter?. In: Human Brain Mapping, abstract no. 784 (2003)

4. Cardoso, J.-F., Souloumiac, A.: Blind beamforming for non Gaussian signals. IEE Proceedings–F 140, 362–370 (1993)
5. Donoho, D., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. Proc. Nat. Acad. Sci. 100, 2197–2202 (2003)
6. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative Matrix Factorization. Nature 40, 788–791 (1999)
7. Makeig, S., et al.: EEGLAB: ICA toolbox for psychophysical research, Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California, San Diego (2000), `http://www.sccn.ucsd.edu/eeglab`
8. Makeig, S., Bell, A.J., Jung, T.-P., Sejnowski, T.J.: Independent component analysis of electroencephalographic data. Advances in neural information processing system 8, 145–151 (1996)
9. Olshausen, B.A., Field, D.J.: Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Research 37, 3311–3325 (1997)
10. Rickard, S., Yilmaz, Z.: On the approximate w-disjoint orthogonality of speech. In: ICASSP '02, pp. I–529–I–532 (2002)