Contrast Functions for Independent Subspace Analysis

Jason A. Palmer and Scott Makeig

Swartz Center for Computational Neuroscience
University of California San Diego, La Jolla, CA 92093
{jason,scott}@sccn.ucsd.edu

Abstract. We consider the Independent Subspace Analysis problem from the point of view of contrast functions, showing that contrast functions are able to partially solve the ISA problem. That is, basic ICA can solve the ISA problem up to within-subspace separation/analysis. We define sub- and super-Gaussian subspaces and extend to ISA a previous result on freedom of ICA from local optima. We also consider new types of dependent densities that satisfy or violate the entropy power inequality (EPI) condition.

1 Introduction

The mutual information minimization approach to blind source separation has proved very effective at separating linear mixtures of independent, nongaussian sources [1]. This approach is equivalent to a maximum likelihood approach in which the source density models are adapted as well [2]. In general, however, some "sources" sources may exhibit mutual dependence, e.g. in signal power, leading to what has been variously called Multidimensional ICA [3], independent subspace analysis [4], and independent vector analysis [5]; or the dependent subspaces may not be further decomposable into unique components, as is the case with non-Gaussian subspaces with radial symmetry.

In a foundational paper on ICA [6], Comon defined the *contrast functions* to be those (statistical) functions of which are capable of separating or extracting independent sources from a linear mixture. This definition is actually very similar to the idea expressed by P. Huber in his work on *projection pursuit* [7, 8]. Basic ICA, i.e. the maximization of a contrast function, is often found to successfully separate sources of the variance dependence type, with the subspace dependence structure ascertainable after the separation. The good performance of basic ICA in the dependent subspace context has led to the conjecture that the minimization of mutual information of the output is able to perform separation of certain dependent sources as well [3].

Theis has considered the ISA problem in a number of articles, proposing the joint block-diagonalization approach [9], considering conditions for separation of non-Gaussian subspaces from Gaussian subspaces in complete mixtures [10], and using the autocorrelation structure of temporally correlated sources or subspaces to perform ISA [11]. Gutch [12] defined the concept of "irreducible subspaces" to be those containing no extractable Gaussian component, and showed that the solution to the ISA problem is unique in this case.

Castella and Comon [13] have also investigated ISA with known dependent subspace structure, and determined specific cases in which cumulant-based contrasts preserve separability and when they fail. Here again the emphasis is on separability of

dependent sources (dependent component analysis) rather than the separation of dependent subspaces from one another.

Szabo [14] has shown using the entropy power inequality (EPI), that dependent sources can be separated by minimum mutual information as long as all one-dimensional projections of the dependent sources satisfy the EPI. Szabo's emphasis is on the complete solution to the dependent component analysis problem, the EPI sufficient condition guaranteeing this possibility (in the case of non-radial symmetry). The EPI approach is shown to be successful when the EPI condition is satisfied by the sources, without requiring prior knowledge of the subspace structure or dimensions.

We show here that dependent subspaces can be separated, i.e. the pairwise mutual information can be block-diagonalized, in a more general setting than that considered in [14]. We take the fact that basic ICA can perform ISA (without necessarily further analyzing the subspaces) as significant in and of itself, since it shows that independent subspaces can be separated from one another as a preliminary processing step, with further analysis of the subspaces themselves carried out subsequently.

This result is significant because it immediately provides an answer for the common criticism of ICA-based methods as being naively misspecified, potentially calling in to question the validity of the results. Essentially we are expanding the concept of a source component to be a potentially multidimensional subspace, with the new "ICA model" that is to be presupposed in the often encountered linear model being that *subspaces* of components are independent. Thus we generalize basic ICA in which all subspaces are one-dimensional, and guarantee the ability of ICA approaches to extract independent sources even in the context of interfering dependent subspaces, as well as guaranteeing that estimated dependent subspaces contain all information pertaining to the subspace that is present in the data.

We also define sub- and super-Gaussian subspaces to be those in which all univariate projections are sub- or super-Gaussian in the Benveniste sense, and show using a previous result [15] that ISA of strongly super-Gaussian subspaces is free of local optima.

2   The ISA Problem

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$, be an invertible matrix consisting of m subspaces, and let $\mathbf{s} \in \mathbb{E}_2^n$ be a finite covariance random vector with m corresponding subvectors:

$$\mathbf{A} = [\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_m], \quad \mathbf{s}^T = [\mathbf{s}_1^T \, \mathbf{s}_2^T \, \cdots \, \mathbf{s}_m^T],$$

where $\mathbf{A}_j \in \mathbb{R}^{n \times d_j}$, $\sum_{j=1}^m d_j = n$, and the $\mathbf{s}_j \in \mathbb{E}_2^{d_j}$ are mutually independent, i.e. $p_\mathbf{s}(\mathbf{s}) = \prod_{j=1}^m p_{\mathbf{s}_j}(\mathbf{s}_j)$. Let $\mathbf{x}$ be the random vector given by,

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

so that $\mathbf{x} \in \mathbb{E}_2^n$.

The ultimate goal of ISA is to reproduce the source vectors, $\mathbf{s}_j$, $j = 1, \ldots, m$, given a set of observations $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$. That is, we would like to produce a matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$ such that $\mathbf{W}\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

However, ISA can be divided logically into two problems:

P1: Separate the independent subspaces from one another.
P2: Separate the dependent sources within each subspace.

Most of the work on the ISA problem has been concerned with solving both problem simultaneously. The ISA separation theory of Szabo gives an entropy condition on dependent sources that allows them to be separated using entropy contrasts. We take it, however, that the well known conjecture that basic ICA also performs ISA, is in fact largely concerned with **P1**, separating the subspaces, or in the particular case of extracting an independent source of interest from a mixture that includes interfering dependent source subspaces.

It should also be noted that **P2** of the ISA problem may not have a solution, as in the case of dependent source subspaces with spherically symmetric distributions. These sources are shown in [14] to satisfy the entropy constraint allowing "solution" of the ISA problem. However, as in the case of Gaussian sources in ICA, spherically symmetric subspaces can only be separated up to an arbitrary rotation. Therefore an ISA problem with spherically symmetric dependent subspaces only consists of **P1**.

3   Deflationary Contrast Functions, ICA, and ISA

In the deflationary approach to basic ICA (where $d_j = 1, j = 1, \ldots, m$,) the matrix $\mathbf{W}$ is constructed one row $\mathbf{w}_j^T$ at a time, i.e. the sources are estimated sequentially. This is usually done by sequentially determining maxima of a *contrast function*, $\Phi : \mathbb{E}_2 \to \mathbb{R}$,

$$\hat{\mathbf{w}}_j = \arg \max_{\mathbf{w}^T \mathbf{R}_{\mathbf{xx}} \mathbf{w} = 1} \Phi(\mathbf{w}^T \mathbf{x})$$

The process ensures successive estimates are unique by restricting $\hat{\mathbf{w}}_j^T \mathbf{R}_{\mathbf{xx}} \hat{\mathbf{w}}_{j'} = 0$ for all previously estimated $\hat{\mathbf{w}}_{j'}$.

3.1   Contrast functions

We define contrast functions as follows [6, 8].

Definition 1 *A **contrast function** is a functional, $\Phi : \mathbb{E}_2 \to \mathbb{R}$, defined on random variables, satisfying the condition,*

$$\Phi\big(\cos(\theta)X + \sin(\theta)Y\big) \leq \max\big(\Phi(X), \Phi(Y)\big), \ X, Y \, independent$$

*If the condition is satisfied for all $X, Y \in \mathcal{S} \subset \mathbb{E}_2$, and strictly satisfied only when θ is a multiple of $\pi/2$, then the contrast is said to **discriminate** over $\mathcal{S}$.*

Examples. Well known contrasts include the following:

1. Inverse entropy power. The entropy power functional, $N(X)$ is defined by

$$N(X) \triangleq (2\pi e)^{-1} \exp(2h(X))$$

where $h(X) \triangleq E\{-\log p_X(X)\}$. The entropy power satisfies, $N(aX) = a^2 N(X)$, and the entropy power inequality (EPI), $N(X + Y) \geq N(X) + N(Y)$. We thus have for the inverse,

$$
\begin{aligned}
N\big(\cos(\theta)X + \sin(\theta)Y\big)^{-1} &\leq \Big(N\big(\cos(\theta)X\big) + N\big(\sin(\theta)Y\big)\Big)^{-1} \\
&= \Big(\cos^2(\theta)N\big(X\big) + \sin^2(\theta)N\big(Y\big)\Big)^{-1} \\
&\leq \cos^2(\theta)\, N(X)^{-1} + \sin^2(\theta)\, N(Y)^{-1} \\
&\leq \max\big(N(X)^{-1}, N(Y)^{-1}\big)
\end{aligned}
$$

2. Fisher Information. The Fisher information $J(X)$ is given by,

$$
J(X) = E\left\{\frac{d^2}{dx^2} - \log p_X(x)\right\} = E\left\{\left(\frac{d}{dx}\log p_X(x)\right)^2\right\}
$$

The Fisher information satisfies, $a^2 J(aX) = J(X)$. It also satisfies an inequality related to the EPI [16]:

$$
\begin{aligned}
J\big(\cos(\theta)X + \sin(\theta)Y\big) &\leq \cos^2(\theta)J(X) + \sin^2(\theta)J(Y) \\
&\leq \max\big(J(X), J(Y)\big)
\end{aligned}
$$

3. Cumulant magnitude. The nth cumulant functional $\kappa_n(X)$ is defined by the nth coefficient of the Taylor expansion of the log characteristic function, $\log \varphi(t)$, where $\varphi(t) \triangleq E\{\exp(itX)\}$. Cumulants satisfy the property,

$$
\kappa_n\big(\cos(\theta)\, X + \sin(\theta)\, Y\big) = \cos^n(\theta)\, \kappa_n(X) + \sin^n(\theta)\, \kappa_n(Y)
$$

For n even, and X, Y of the same cumulant sign,

$$
\begin{aligned}
\big|\kappa_n\big(\cos(\theta)\, X + \sin(\theta)\, Y\big)\big| &= \cos^n(\theta)|\kappa_n(X)| + \sin^n(\theta)\,|\kappa_n(Y)| \\
&\leq \cos^2(\theta)\,|\kappa_n(X)| + \sin^2(\theta)\,|\kappa_n(Y)| \\
&\leq \max\big(|\kappa_n(X)|, |\kappa_n(Y)|\big)
\end{aligned}
$$

Thus even cumulant magnitude defines a contrast discriminating over sets of random variables with the same cumulant sign.

3.2　Contrasts and ISA

We now show that every contrast function is also a subspace contrast in the sense that deflationary ICA can be used to solve **P1**.

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be invertible, $\mathbf{A} = [\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_m]$, where $\mathbf{A}_j \in \mathbb{R}^{n \times d_j}$, $\sum_{j=1}^m d_j = n$. Let $\mathbf{s} \in \mathbb{E}_2^n$, $\mathbf{s}^T = [\mathbf{s}_1^T \mathbf{s}_2^T \cdots \mathbf{s}_m^T]$, with the $\mathbf{s}_j \in \mathbb{E}_2^{d_j}$ mutually independent. Let $\mathbf{x} = \mathbf{A}\mathbf{s}$.

We prove specifically that for a deflationary contrast Φ, if,

$$
\hat{\mathbf{w}}_{j'} = \arg \max_{\mathbf{w}^T \mathbf{R}_{\mathbf{xx}}\mathbf{w}=1} \Phi(\mathbf{w}^T \mathbf{x})
$$

then $\mathbf{w}^T \mathbf{A}_j \mathbf{A}_j^T \mathbf{w} = 0$ for all $j \neq j'$.

Theorem 1 *The deflationary contrast method solves* **P1**.

Proof. Let $y = \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{A} \mathbf{s} = \mathbf{c}^T \mathbf{s}$, where $\mathbf{c} \triangleq \mathbf{A}^T \mathbf{w}$. Since $\mathbf{R_{xx}} = E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{A}\mathbf{A}^T$, we have that $\mathbf{w}^T \mathbf{R_{xx}} \mathbf{w} = 1$ implies $\mathbf{c}^T \mathbf{c} = 1$.

We have $y = \mathbf{c}^T \mathbf{s} = \sum_{j=1}^m \mathbf{c}_j^T \mathbf{s}_j$. By the contrast function condition, we have,

$$\Phi(\mathbf{c}^T \mathbf{s}) = \Phi\left(\sum_{j=1}^m \mathbf{c}_j^T \mathbf{s}_j\right) = \Phi\left(\sum_{\|\mathbf{c}_j\| \neq 0} \|\mathbf{c}_j\| \frac{\mathbf{c}_j^T \mathbf{s}_j}{\|\mathbf{c}_j\|}\right) \leq \max_j \Phi\left(\frac{\mathbf{c}_j^T \mathbf{s}_j}{\|\mathbf{c}_j\|}\right)$$

with equality only if $\|\mathbf{c}_j\| = 0$ for all but the maximizing j. But this implies that $\mathbf{w}^T \mathbf{A}_j \mathbf{A}_j^T \mathbf{w} = 0$ for all but the maximizing j.

This theorem shows that the solution to a stage of the deflationary ICA process can only be a linear combination of sources from within one and only one of the dependent subspaces. Each subsequent source estimate will either be dependent with a previously estimated source (having positive mutual information) and be a linear combination only of sources in that subspace, or will be independent of previously estimated sources, beginning the estimate of (one direction in) a new independent subspace. At the end of the procedure, the matrix of pairwise mutual information values between the estimated sources will be a block diagonal permutation.

4 Sub- and Super-Gaussian Subspaces

In this section we define particular classes of dependent subspaces in terms of linear projections, and use a previously derived result on globally optimal ICA [15] to show that the solution to **P1** of the ISA problem is also free of local optima.

We first review the Benveniste definition of (strong) sub- and super-Gaussianity.

Definition 2 (Strongly Sub- and Super-Gaussian Random Variables) *Let X be a random variable with differentiable probability density function, $p_X(x)$. Define $f(x) \triangleq -\log p_X(x)$. Then p_X is a strongly super-Gaussian (sub-Gaussian) if $p_X(x)$ is symmetric about $x = 0$ and $f'(x)/x$ is strictly decreasing (increasing) on $x > 0$.*

We define sub- and super-Gaussian subspaces to be spaces of dependent random variables in which all linear projections are strongly sub- or super-Gaussian respectively.

Definition 3 (Sub- and Super-Gaussian Subspaces) *Let $\mathbf{x} \in \mathbb{R}^d$ be a non-Gaussian dependent random vector. Then $\mathbf{x}$ is a strongly super-Gaussian (sub-Gaussian) random vector if, for all $\mathbf{w} \in \mathbb{R}^d$, we have $y = \mathbf{w}^T \mathbf{x}$ strongly super-Gaussian (sub-Gaussian).*

In previous work [17], we have considered Generalized Gaussian scale mixtures as an example of a non-radially symmetric dependent subspace. As a more general GSM-based formulation dependent subspaces, let $\mathbf{s}$ have independent GSM components, i.e. $s_i = \xi^{1/2} z$ for a non-negative finite variance ξ_i and Gaussian z_i. Let,

$$\mathbf{y} = \eta^{1/2} \mathbf{s}$$

where a common nonnegative scalar $\eta^{1/2}$ multiplies each (independent) GSM component of $\mathbf{s} \in \mathbb{E}_2^d$ to form random vector $\mathbf{y}$ with dependent components. Then we have,

$$u = \mathbf{w}^T \mathbf{y} = \eta^{1/2} \sum_i w_i \xi_i^{1/2} z_i \overset{d}{=} \eta^{1/2} \big(\xi_1 w_1^2 + \cdots + \xi_d w_d^2 \big)^{1/2} z_1$$

so that u is also a GSM, and thus strongly super-Gaussian. Dependent GSM subspaces are thus strongly super-Gaussian as defined here.

Theorem 2 *The ISA problem* **P1** *with strongly super-Gaussian dependent subspaces has no local optima when solved using a strongly super-Gaussian contrast.*

This follows from the theorem proved in [15].

5 Other Types of Norm Dependence

We finally consider random vectors with somewhat more general dependent densities to inquire as to which types of non-radially symmetric dependent subspaces violate the EPI condition of [14]. That is, what kinds of dependent sources are and are not separated by contrast functions in the solution of **P1**.

Consider a two dimensional dependent subspace with density,

$$p(x_1, x_2) = f\big(g(x_1) + g(x_2)\big)$$

Let $h_y(\theta)$ be the entropy of projections $y = \cos(\theta) x_1 + \sin(\theta) x_2$ as a function of θ.

Theorem 3 *Let f be decreasing, with $-\log f(\sqrt{x})$ concave. Let $g(\sqrt{x})$ be increasing and concave on $x \in (0, \infty)$, then for $\theta \in (0, \pi/4)$, we have,*

$$h_y'(\theta) \geq 0$$

This follows from a derivation similar to that in [15].

Definition 4 *A density, $p(x_1, \ldots, x_n)$, is said to be* sup-sup dependent *(respectively* sub-sub dependent*) if it is of the form,*

$$p(x_1, \ldots, x_n) = f\big(g_1(x_1) + \cdots + g_n(x_n)\big)$$

with f decreasing on $(0, \infty)$, $-\log f(\sqrt{y})$ concave (respectively convex), $g_i(x_i)$ nonnegative, symmetric, and increasing on $(0, \infty)$, and $g_i(\sqrt{x})$ concave (respectively convex) on $(0, \infty)$, for $i = 1, \ldots, n$. Sup-sub *and* sub-sup dependence *are defined by the concave-convex and convex-concave scenarios respectively.*

Corollary 1 *Sup-sup and sub-sub dependent densities satisfy the EPI condition of [14] and may thus be separated by contrast functions.*

If the convexity is not "homogeneous" but rather "conflicting" such that one of $-\log f$ and $-\log g$ is concave and one is convex, then we have,

$$h'_y(\theta) \leq 0, \quad \theta \in (0, \pi/4)$$

In Figure 1, we present some experiments to verify the theory of this section. We generate four sets of two-dimensional dependent sources, corresponding to the sup-sup, sub-sub, sub-sup, and sup-sub cases respectively. The "sup" density is Laplacian, i.e. $p(x) \propto \exp(-|x|)$, and the "sub" density is Generalized Gaussian with shape parameter 5, $p(x) \propto \exp(-|x|^5)$. The sup-sup data is generated by multiplying i.i.d. Laplacian samples by a common instance dependent scaling, which is Gamma distributed. This creates a supergaussian envelope dependence. The sub-sub data is generated by inducing a slight variance dependence on i.i.d. subgaussian data by multiplying it by a common random Gamma scaling that is tightly concentrated about unity. The sub-sup data is generated by multiplying uniform data over the diamond (Laplacian level curves) by a slight common scaling to induce a subgaussian envelope over Laplacian level curves. The sup-sub data is generated by multiplying i.i.d. uniform data by a strong scaling, to induce a supergaussian envelope on uniform (subgaussian) level curves. The "time series" are shown in the second row, shifted to improve visibility. The bottom row plots the entropy of projections as a function of the rotation angle for $\theta \in (0, \pi/2)$. Symmetry is expected about $\pi/4$, and deviation gives an idea of the noise in the empirical entropy calculation. Entropy is calculated by approximately integrating the histogram.

It can be seen that entropy increases with rotation for the sup-sup and sub-sub dependent sources, while it decreases for the sub-sup and sup-sub dependent sources, as predicted.

References

1. D. T. Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Trans. Information Theory*, vol. 48, no. 7, pp. 1935–1946, 2002.
2. J.-F. Cardoso, "Infomax and maximum likelihood for source separation," *IEEE Letters on Signal Processing*, vol. 4, no. 4, pp. 112–114, 1997.
3. J.-F. Cardoso, "Multidimensional independent component analysis," in *Proceedings of the IEEE International Conference on Acoustics and Signal Processing (ICASSP 1998), Seattle, WA*, 1998, pp. 1941–1944.
4. A. Hyvärinen and P. O. Hoyer, "Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, vol. 12, pp. 1705–1720, 2000.
5. T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *Proceedings of the 6th International Conference on Independent Component Analysis*, J. Rosca et al., Ed. 2006, Lecture Notes in Computer Science, pp. 165–172, Springer-Verlag.
6. P. Comon, "Independent component analysis: a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
7. P. J. Huber, "Projection pursuit," *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, 1985.
8. D. T. Pham, "Contrast functions for blind separation and deconvolution of sources," Tech. Rep., Laboratoire de Modelisation et Calcul, CNRS, IMAG, 2001.
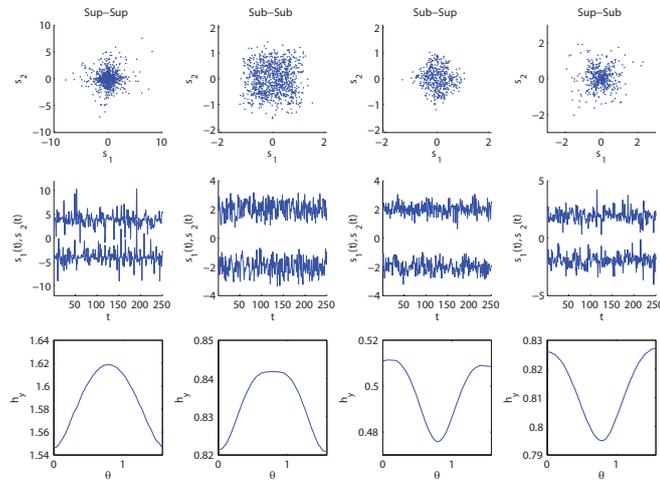
Fig. 1. Examples of dependent densities with various combinations of subgaussian and super-gaussian envelope and level curve function.

 9. Fabian J. Theis, "Blind signal separation into groups of dependent signals using joint block diagonalization," in *ISCAS (6)*, 2005, pp. 5878–5881.
10. F.J. Theis and M. Kawanabe, "Uniqueness of non-gaussian subspace analysis," in *Proc. ICA 2006*, Charleston, USA, 2006, pp. 917–925.
11. Fabian J. Theis, "Colored subspace analysis: Dimension reduction based on a signal's auto-correlation structure," *IEEE Trans. on Circuits and Systems*, vol. 57-I, no. 7, pp. 1463–1474, 2010.
12. Harold W. Gutch and Fabian J. Theis, "Independent subspace analysis is unique, given irreducibility," in *ICA*, 2007, pp. 49–56.
13. Marc Castella and Pierre Comon, "Blind separation of instantaneous mixtures of dependent sources," in *ICA*, 2007, pp. 9–16.
14. Zoltán Szabó, Barnabás Póczos, and András Lörincz, "Undercomplete blind subspace de-convolution via linear prediction," in *ECML*, 2007, pp. 740–747.
15. J. A. Palmer, K. Kreutz-Delgado, and S. Makeig, "Strong sub- and super-gaussianity," in *Proceedings of the 9th International Conference on Latent Variable Analysis and Independent Component Analysis*, Rémi Gribonval and Emmanuel Vincent, Eds. 2010, Lecture Notes in Computer Science, Springer-Verlag.
16. A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1501–1518, 1991.
17. J. A. Palmer, K. Kreutz-Delgado, B. D. Rao, and S. Makeig, "Modeling and estimation of dependent subspaces with non-radially symmetric and skewed densities," in *Proceedings of the 7th International Conference on Independent Component Analysis*, M. E. Davies, C. J. James, S. A. Abdallah, and M. D Plumbley, Eds. 2007, Lecture Notes in Computer Science, Springer-Verlag.