AN INFORMATION MAXIMIZATION BASED BLIND SOURCE SEPARATION APPROACH FOR DEPENDENT AND INDEPENDENT SOURCES

Alper T. Erdogan

EE Department and KUIS AI Lab, Koc University

ABSTRACT

We introduce a new information maximization (infomax) approach for the blind source separation problem. The proposed framework provides an information-theoretic perspective for determinant maximization-based structured matrix factorization methods such as nonnegative and polytopic matrix factorization. For this purpose, we use an alternative joint entropy measure based on the log-determinant of covariance, which we refer to as log-determinant (LD) entropy. The corresponding (LD) mutual information between two vectors reflects a level of their correlation. We pose the infomax BSS criterion as the maximization of the LD-mutual information between the input and output of the separator under the constraint that the output vectors lie in a presumed domain set. In contrast to the ICA infomax approach, the proposed information maximization approach can separate both dependent and independent sources. Furthermore, we can provide a finite sample guarantee for the perfect separation condition in the noiseless case.

Index Terms— Blind Source Separation, Information Maximization, Dependent Source Separation, Polytopic Matrix Factorization, Independent Component Analysis

1. INTRODUCTION

Blind source separation (BSS) has long been a central problem of interest in both the signal processing and machine learning communities [1, 2, 3]. Solution approaches proposed for this problem exploit side information regarding the data model to extract latent components from their mixtures.

Independent component analysis (ICA) has been a key framework that utilizes the assumption about the mutual independence of sources. Among several alternative ICA algorithmic approaches, the infomax principle provided an information-theoretic perspective to the BSS problem [4]. It aims to maximize the information transfer from input to output while constraining the output to satisfy the mutual independence assumption.

As an alternative to ICA, we can list deterministic approaches that utilize particular domain structures of the sources. For example, nonnegative matrix factorization (NMF) [5, 6, 7, 8] exploits the assumption that the sources are located in the nonnegative orthant. Antisparse bounded component analysis (BCA) algorithms make use of the sources' ℓ_∞-norm ball membership feature [9, 10, 11]. Sparse component analysis (SCA) algorithms can be based on the assumption that the sources are from the ℓ_1-norm ball[12, 13, 14]. Recently, polytopic matrix factorization (PMF) has been introduced as a new framework for solving inverse linear problems based on the knowledge that source vectors are drawn from polytopes [15, 16]. A common separation criterion employed by all these deterministic frameworks is the maximization of the determinant of the sample correlation matrix of source estimates.

This article introduces a new information maximization perspective for BSS approaches based on determinant maximization. The proposed framework differs from the ICA-infomax approach from [4] in two major ways:

(i) it does not necessarily assume that sources are mutually independent, and

(ii) instead of Shannon-mutual information, it uses an alternative, the so-called log-determinant (LD)-mutual information measure (see Section 4).

Regarding item (ii) above, for a vector with the covariance matrix $\mathbf{R_x}$, we use a joint entropy measure (LD-entropy) based on the log-determinant of the covariance matrix, which disregards the higher order statistics about the random vector. As we will show in Section 4, the corresponding mutual information reflects correlation rather than dependence between its arguments. Furthermore, it is based on a linear inference model with the minimum mean square error (MMSE) criterion. The LD-entropy measure has recently been proposed as an estimator of Shannon entropy for the information bottleneck analysis of deep neural networks [17]. However, we do not use it as an approximation tool; instead, we use it as a standalone measure relevant to the linear inference setting of the BSS problem. The resulting optimization objective function resembles but is different from those used in determinant maximization-based matrix factorization approaches [7, 8, 16].

The following is the article's organization: Section 2 provides the BSS setup used throughout the article. Although we

use PMF as a reference approach, it can be adapted to structured matrix factorization approaches such as NMF. Section 3 summarizes results relevant to BSS from PMF framework. In Section 4, we introduce the LD-entropy and LD-mutual information measures and their deterministic counterparts. We define the LD-infomax criterion for BSS in Section 5 and provide a corresponding algorithm in Section 6. Finally, in Section 7, we provide an example that illustrates the LD-infomax algorithm's dependent source separation capability.

2. SOURCE SEPARATION SETUP

We assume the following generative data model for observations throughout the article:

Sources: There are r real valued sources with N samples each, which form $\mathcal{S}_g = \{\mathbf{s}_g(1), \ldots, \mathbf{s}_g(N)\} \subset \mathbb{R}^r$. We also define the matrix for the generative source samples as

$$\mathbf{S}_g = \begin{bmatrix} \mathbf{s}_g(1) & \mathbf{s}_g(2) & \ldots & \mathbf{s}_g(N) \end{bmatrix} \in \mathbb{R}^{r \times N}.$$

We further assume that $\mathcal{S}_g$ is a subset of a polytope $\mathcal{P}$. The choice of $\mathcal{P}$ reflects the assumptions on sources and their mutual relationships [15, 16]. For example, defining $\mathcal{B}_p = \{\mathbf{x} | \|\mathbf{x}\|_p \le 1\}$ as the ℓ_p-norm ball:

- $\mathcal{P} = \mathcal{B}_1$ ($\mathcal{P} = \mathcal{B}_\infty$) corresponds to sparse (anti-sparse) sources, and

- $\mathcal{P} = \mathcal{B}_{1,+} = \mathcal{B}_1 \cap \mathbb{R}_+^r$ ($\mathcal{P} = \mathcal{B}_{\infty,+} = \mathcal{B}_\infty \cap \mathbb{R}_+^r$) is for sparse (anti-sparse) nonnegative sources.

In addition to these polytopes corresponding to globally defined source attributes, we can also define polytopes with heterogeneous source properties, such as the example provided in [16]:

$$\mathcal{P}_{\text{ex}} = \left\{ \mathbf{s} \in \mathbb{R}^3 \; \middle| \; \begin{array}{l} s_1, s_2 \in [-1, 1], s_3 \in [0, 1], \\ \left\| \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \right\|_1 \le 1, \left\| \begin{bmatrix} s_2 \\ s_3 \end{bmatrix} \right\|_1 \le 1 \end{array} \right\},$$

where s_1, s_2 are signed, s_3 is nonnegative and the sparsity is defined between s_1, s_2 and s_2, s_3. The PMF framework offers infinite choices for $\mathcal{P}$, which provides a high degree of freedom in generating diverse sets of source features and relationships. For a polytope to lead to a "solvable" BSS problem, it needs to be "identifiable" in the sense that only linear transformations that map $\mathcal{P}$ to itself are permutation or diagonal scaling operations [16].

To guarantee the identifiability of the original sources, the PMF framework in [16] proposed a sufficiently scattered condition for the source samples in $\mathcal{S}_g$ based on the maximum volume inscribed ellipsoid (MVIE) of $\mathcal{P}$ [18]. This condition requires that the source samples' convex hull, $\text{conv}(\mathcal{S}_g)$, forms a better approximation of the polytope than its MVIE, as formalized by the following definition:

Definition 1. *Sufficiently Scattered Set of $\mathcal{P}$: $\mathcal{S}_g$ is called a sufficiently scattered set of $\mathcal{P}$ if*

(PMF.SS.i) $\mathcal{P} \supseteq \text{conv}(\mathcal{S}_g) \supset \mathcal{E}_\mathcal{P}$, and

(PMF.SS.ii) $\text{conv}(\mathbf{S})^{*,\mathbf{g}_\mathcal{P}} \cap bd(\mathcal{E}_\mathcal{P}^{*,\mathbf{g}_\mathcal{P}}) = ext(\mathcal{P}^{*,\mathbf{g}_\mathcal{P}})$,

where $\mathcal{E}_\mathcal{P}$ is the MVIE of $\mathcal{P}$, centered at $\mathbf{g}_\mathcal{P}$, $bd(\cdot)$ and $ext(\cdot)$ stand for boundary and vertices respectively.

The notation $C^{*,\mathbf{g}}$ in the (PMF.SS.ii) above stands for the polar of the set $C \subset \mathbb{R}^r$ relative to a point $\mathbf{g}$ as defined by:

$$C^{*,\mathbf{g}} = \{\mathbf{x} \in \mathbb{R}^r | \langle \mathbf{x}, \mathbf{y} - \mathbf{g} \rangle \le 1 \; \forall \mathbf{y} \in C\}.$$

In Definition 1, (PMF.SS.i) ensures that $\text{conv}(\mathcal{S}_g)$ contains $\mathcal{E}_\mathcal{P}$ while (PMF.SS.ii) restricts how tight the boundary of $\text{conv}(\mathcal{S}_g)$ encloses $\mathcal{E}_\mathcal{P}$ by restricting its tangency points.

Mixing: We assume that sources are mixed through linear mapping with matrix $\mathbf{H}_g \in \mathbb{R}^{M \times r}$, where we consider the (over)determined case, i.e., $M \ge r$. In the noiseless case, we can write the observations, i.e., the mixture vectors, as:

$$\mathbf{y}(k) = \mathbf{H}_g \mathbf{s}_g(k), \quad k \in \{1, \ldots, N\}. \tag{1}$$

We can write the corresponding matrix of mixtures as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}(1) & \mathbf{y}(2) & \ldots & \mathbf{y}(N) \end{bmatrix} = \mathbf{H}_g \mathbf{S}_g.$$

In BSS, the goal is to estimate $\mathbf{S}_g$ from $\mathbf{Y}$ when $\mathbf{H}_g$ is not known and no training information is available.

3. PMF AS A BSS APPROACH

In this section, we revisit the PMF approach introduced in [15, 16] which can be utilized to solve the BSS problem in Section 2. The PMF approach exploits the information that the original source samples are sufficiently scattered inside a polytope $\mathcal{P}$. The corresponding criterion is defined in terms of the following optimization problem:

$$\underset{\mathbf{H} \in \mathbb{R}^{M \times r}, \mathbf{S} \in \mathbb{R}^{r \times N}}{\text{maximize}} \quad \det(\mathbf{S}\mathbf{S}^T) \tag{2a}$$

$$\text{subject to} \quad \mathbf{Y} = \mathbf{H}\mathbf{S}, \tag{2b}$$

$$\mathbf{S}_{:,j} \in \mathcal{P}, \quad j = 1, \ldots, N. \tag{2c}$$

The objective function in (2a) is the scaled sample correlation matrix, which is a measure of the spread of the columns of $\mathbf{S}$ inside $\mathcal{P}$. According to [15, 16], if

I. $\mathcal{P}$ is an "identifiable" polytope, as defined in [16], and

II. the columns of $\mathbf{S}_g$ are a "sufficiently scattered set" in $\mathcal{P}$ as defined in Definition 1,

then any $\mathbf{S}_*$ of (2) can be written as: $\mathbf{S}_* = \mathbf{\Pi}\mathbf{D}\mathbf{S}_g$ where $\mathbf{D}$ is a diagonal matrix with ± 1 diagonal entries, corresponding to sign ambiguity, and $\mathbf{\Pi}$ is a permutation matrix, corresponding to permutation ambiguity. We can modify the

PMF optimization problem in (2) by replacing the objective with the log-determinant of the sample covariance matrix: $\hat{\mathbf{R}}_{\mathbf{s}} = \frac{1}{N}\mathbf{SS}^T - \frac{1}{N^2}\mathbf{S}\mathbf{1}\mathbf{1}^T\mathbf{S}^T$, as

$$\underset{\mathbf{H} \in \mathbb{R}^{M \times r},\, \mathbf{S} \in \mathbb{R}^{r \times N}}{\text{maximize}} \qquad \log\det(\hat{\mathbf{R}}_{\mathbf{s}}) \qquad (3a)$$

$$\text{subject to} \qquad \mathbf{Y} = \mathbf{HS}, \qquad (3b)$$

$$\mathbf{S}_{:,j} \in \mathcal{P}, \quad j = 1,\dots,N. \quad (3c)$$

In the case of noisy mixtures, we can remove the equality constraint in (3b) and replace the objective function with [7, 16]:

$$J(\mathbf{S}) = \frac{1}{2}\log\det(\hat{\mathbf{R}}_{\mathbf{s}}) - \gamma\|\mathbf{Y} - \mathbf{HS}\|_F^2. \qquad (4)$$

4. LOG-DETERMINANT ENTROPY AND INFORMATION MEASURE

We define the log-determinant (LD)-entropy of a random vector $\mathbf{x} \in \mathbb{R}^r$ with covariance $\mathbf{R}_{\mathbf{x}}$ as [17]

$$H_{LD}^{(\epsilon)}(\mathbf{x}) = \frac{1}{2}\log\det(\mathbf{R}_{\mathbf{x}} + \epsilon\mathbf{I}) + \frac{r}{2}\log(2\pi e), \qquad (5)$$

where $\epsilon > 0$ is a small value to ensure $\log\det$ term is lower bounded by a finite value. Note that $H_{LD}^{(0)}(\mathbf{x})$ is equal to the expression for Shannon differential entropy for a Gaussian vector, which constitutes an upper bound on the Shannon differential entropy for all random vectors with covariance $\mathbf{R}_{\mathbf{x}}$.

The joint LD-entropy of $\mathbf{x} \in \mathbb{R}^r$ and $\mathbf{y} \in \mathbb{R}^q$ can be written as:

$$H_{LD}^{(\epsilon)}(\mathbf{x},\mathbf{y}) = \frac{1}{2}\log\det(\mathbf{R}_{\left[\begin{smallmatrix}\mathbf{x}\\\mathbf{y}\end{smallmatrix}\right]} + \epsilon\mathbf{I}) + \frac{n+m}{2}\log(2\pi e),$$

$$= \frac{1}{2}\log\det\left(\begin{bmatrix}\mathbf{R}_{\mathbf{x}} + \epsilon\mathbf{I} & \mathbf{R}_{\mathbf{xy}}\\ \mathbf{R}_{\mathbf{xy}}^T & \mathbf{R}_{\mathbf{y}} + \epsilon\mathbf{I}\end{bmatrix}\right) + \frac{r+q}{2}\log(2\pi e),$$

$$= \frac{1}{2}\log\det(\mathbf{R}_{\mathbf{x}} + \epsilon\mathbf{I}) + \frac{r}{2}\log(2\pi e)$$

$$+ \frac{1}{2}\log\det(\mathbf{R}_{\mathbf{e}} + \epsilon\mathbf{I}) + \frac{q}{2}\log(2\pi e)$$

$$= H_{LD}^{(\epsilon)}(\mathbf{x}) + H_{LD}^{(\epsilon)}(\mathbf{y}|_L\mathbf{x}),$$

where $\mathbf{R}_{\mathbf{x}},\mathbf{R}_{\mathbf{y}}$ are the (auto)covariance matrices of $\mathbf{x}$ and $\mathbf{y}$ vectors respectively, $\mathbf{R}_{\mathbf{xy}}$ is the cross covariance matrix between $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{R}_{\mathbf{e}} = \mathbf{R}_{\mathbf{y}} - \mathbf{R}_{\mathbf{xy}}^T(\mathbf{R}_{\mathbf{x}} + \epsilon\mathbf{I})^{-1}\mathbf{R}_{\mathbf{xy}}$. Note that $H_{LD}^{(\epsilon)}(\mathbf{y}|_L\mathbf{x}) = \frac{1}{2}\log\det(\mathbf{R}_{\mathbf{e}} + \epsilon\mathbf{I}) + \frac{q}{2}\log(2\pi e)$ is not equal to the LD-entropy of $\mathbf{y}|\mathbf{x}$ which is expressed by $H_{LD}(\mathbf{y}|\mathbf{x}) = \frac{1}{2}\log\det(\mathbf{R}_{\mathbf{y}|\mathbf{x}}) + \frac{q}{2}\log(2\pi e)$. The notation $\mathbf{y}|_L\mathbf{x}$ signifies this difference.

If we consider the limit $\epsilon \to 0$, the expression $H_{LD}^{(\epsilon)}(\mathbf{y}|_L\mathbf{x})$ corresponds to the log-determinant of the error covariance of the best affine (MMSE) of $\mathbf{y}$ from $\mathbf{x}$. Therefore, $H_{LD}^{(\epsilon)}(\mathbf{y}|_L\mathbf{x})$ can be considered a measure of the uncertainty remaining

in $\mathbf{y}$ after linearly estimating it from $\mathbf{x}$ with respect to the MMSE criterion, for sufficiently small ϵ.

Based on the LD-entropy definition in (5), we can provide the corresponding LD-mutual information measure as:

$$I^{(\epsilon)}(\mathbf{x},\mathbf{y}) = H^{(\epsilon)}(\mathbf{y}) - H_{LD}^{(\epsilon)}(\mathbf{y}|_L\mathbf{x}). \qquad (6)$$

We can show also show that:

$$I^{(\epsilon)}(\mathbf{x},\mathbf{y}) = H^{(\epsilon)}(\mathbf{x}) - H_{LD}^{(\epsilon)}(\mathbf{x}|_L\mathbf{y}) \qquad (7)$$

For LD-mutual information, the following property holds:

Lemma 1. *For non-degenerate random vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *with* $\mathbf{R}_{\mathbf{x}} \succ \mathbf{0}$ *and* $\mathbf{R}_{\mathbf{y}} \succ \mathbf{0}$ *respectively,* $I^{(\epsilon)}(\mathbf{x},\mathbf{y}) \geq 0$ *with equality iff they are mutually uncorrelated, i.e.* $\mathbf{R}_{\mathbf{xy}} = \mathbf{0}$.

In summary, LD-entropy and LD-mutual information act as relevant uncertainty and correlative information measures, respectively, when the inference is based on the linear MMSE estimation, for ϵ sufficiently small.

4.1. Deterministic Case

For a deterministic setting such as the BSS setup in Section 2, we can work with the deterministic counterparts of the LD-entropy and LD-mutual information provided above. For example, for a set of source samples $\{\mathbf{s}(1),\mathbf{s}(2),\dots,\mathbf{s}(N)\}$ with the sample covariance $\hat{\mathbf{R}}_{\mathbf{s}}$, the corresponding deterministic LD-entropy can be written as:

$$\hat{H}_{LD}^{(\epsilon)}(\mathbf{S}) = \frac{1}{2}\log\det(\hat{\mathbf{R}}_{\mathbf{s}} + \epsilon\mathbf{I}) + \frac{r}{2}\log(2\pi e). \qquad (8)$$

We can, therefore, consider that the PMF optimization objective in (3a) is essentially equivalent to $\hat{H}_{LD}^{(\epsilon)}(\mathbf{s})$. It is interesting that for a sufficiently scattered $\mathcal{S}_g$ of $\mathcal{P}$, no linear transformation that maps $\mathcal{S}_g$ into $\mathcal{P}$ generates a set with larger deterministic LD-entropy, which is implied by the identifiability condition of the PMF setting in (3) [16]. Therefore, one can consider a sufficiently scattered set $\mathcal{S}_g$ in $\mathcal{P}$ as a maximal deterministic LD-entropy set of $\mathcal{P}$ in this sense.

The deterministic LD-mutual information for $\mathbf{s}$ and $\mathbf{y}$ is expressed as:

$$\hat{I}^{(\epsilon)}(\mathbf{Y},\mathbf{S}) = \frac{1}{2}\log\det(\hat{\mathbf{R}}_{\mathbf{s}} + \epsilon\mathbf{I})$$

$$-\frac{1}{2}\log\det(\hat{\mathbf{R}}_{\mathbf{s}} - \hat{\mathbf{R}}_{\mathbf{sy}}(\epsilon\mathbf{I} + \hat{\mathbf{R}}_{\mathbf{y}})^{-1}\hat{\mathbf{R}}_{\mathbf{sy}}^T + \epsilon\mathbf{I}), \quad (9)$$

which is written in terms of the sample covariance matrices $\hat{\mathbf{R}}_{\mathbf{y}} = \frac{1}{N}\mathbf{YY}^T - \frac{1}{N^2}\mathbf{Y}\mathbf{1}\mathbf{1}^T\mathbf{Y}^T$, and $\hat{\mathbf{R}}_{\mathbf{sy}} = \frac{1}{N}\mathbf{SY}^T - \frac{1}{N^2}\mathbf{S}\mathbf{1}\mathbf{1}^T\mathbf{Y}^T$.

5. LD-INFOMAX AS BSS CRITERION

To obtain a solution to the BSS setting in Section 2, based on the LD-mutual information measure defined in Section 4, we

propose the use of the following optimization problem:

$$\underset{\mathbf{S} \in \mathbb{R}^{r \times N}}{\text{maximize}} \quad \hat{I}^{(\epsilon)}(\mathbf{Y}, \mathbf{S}) \tag{10a}$$

$$\text{subject to} \quad \mathbf{S}_{:,j} \in \mathcal{P}, \quad j = 1, \dots, N. \tag{10b}$$

The proposed optimization aims to maximize the LD-mutual information between the output samples $\{\mathbf{s}(1), \dots, \mathbf{s}(N)\}$ and input mixture samples $\{\mathbf{y}(1), \dots \mathbf{y}(N)\}$ while the output samples are constrained to lie in $\mathcal{P}$. This approach exploits the assumption that the original source samples form a maximal LD-entropy set of $\mathcal{P}$. The objective function in (9) resembles the form in (4), of which the second term is replaced with $\frac{1}{2} \log \det(\hat{\mathbf{R}}_{\mathbf{e}} + \epsilon \mathbf{I})$, where $\hat{\mathbf{R}}_{\mathbf{e}} = \hat{\mathbf{R}}_{\mathbf{s}} - \hat{\mathbf{R}}_{\mathbf{sy}}(\epsilon \mathbf{I} + \hat{\mathbf{R}}_{\mathbf{y}})^{-1}\hat{\mathbf{R}}_{\mathbf{sy}}^T$.

In the noiseless BSS setting, when ϵ is selected to be sufficiently small, the second term in (9) is maximized when the relation between the input and output is linear. Furthermore, the first term in (9) approaches the objective function in (3a). Therefore, the optimization problem in (10) becomes equivalent to the PMF optimization in (3). If $\mathcal{S}_g$ is sufficiently scattered, i.e., a maximal LD-entropy set of $\mathcal{P}$, then it is possible to identify the original sources, with some permutation and sign ambiguity, from a finite number of samples.

6. ALGORITHM

As a simple approach, we consider a projected gradient search algorithm for the optimization problem in (10). The gradient of the objective function in (10a) is given by:

$$\nabla_{\mathbf{S}} \hat{I}^{(\epsilon)}(\mathbf{Y}, \mathbf{S}) = \frac{1}{N}(\hat{\mathbf{R}}_{\mathbf{s}} + \epsilon \mathbf{I})^{-1} \mathbf{S}(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T)$$
$$- \frac{1}{N}(\hat{\mathbf{R}}_{\mathbf{e}} + \epsilon \mathbf{I})^{-1}(\mathbf{S} - \hat{\mathbf{R}}_{\mathbf{sy}}(\hat{\mathbf{R}}_{\mathbf{y}} + \epsilon \mathbf{I})^{-1}\mathbf{Y})(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T).$$

We can express the corresponding algorithm update rule as:

$$\mathbf{S}^{(k+1)} = P_{\mathcal{P}}\left(\mathbf{S}^{(k)} + \mu^{(k)}\nabla_{\mathbf{S}}\hat{I}^{(\epsilon)}(\mathbf{Y}, \mathbf{S}^{(k)})\right), \tag{11}$$

where $P_{\mathcal{P}}(\cdot)$ is the operator for projection onto the polytope $\mathcal{P}$, and $\mu^{(k)}$ is the learning rate at iteration k.

7. NUMERICAL EXAMPLE

We consider a BSS scenario with $r = 5$ sources and $M = 8$ mixtures. The sources are nonnegative copula-t random variables with 4-degrees of freedom with a Toeplitz generator correlation matrix of which the first row is given by $\begin{bmatrix} 1 & \rho & \rho & \rho & \rho \end{bmatrix}$. Here, ρ is the parameter that controls the correlation between sources. Therefore, the columns of $\mathbf{S}_g$ are correlated samples in $\mathcal{B}_{1,+}$. The mixing matrix $\mathbf{H}$ is generated as a random 8×5 matrix with i.i.d. zero mean and unity variance Gaussian entries. The mixtures are corrupted by i.i.d. Gaussian noise. We used $N = 10000$ samples at each realization and chose $\epsilon = 1 \times 10^{-5}$.

In the simulations, at each iteration, we calculated the mean square error $\text{MSE}^{(k)} = \frac{1}{N}\|\mathbf{S}^{(k)} - \mathbf{D}^{(k)}\mathbf{\Pi}^{(k)}\mathbf{S}_g\|_F^2$, where $\mathbf{D}^{(k)}$ and $\mathbf{\Pi}^{(k)}$ are diagonal and permutation matrices that minimize the error. We evaluated the signal-to-interference+noise-power-ratio (SINR) by calculating the ratio of the average source power to $\text{MSE}^{(k)}$. We used the
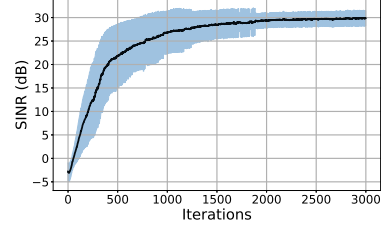

Fig. 1. Convergence of the LD-infomax algorithm.

step size rule with $\mu^{(k)} = \frac{200}{\sqrt{k+1}}$ for the LD-Infomax algorithm in Section 6. For a mixture SNR of 30dB and $\rho = 0.5$, Figure 1 shows mean SINR convergence curve, for 100 trials, with standard deviation envelope as a function of iterations. Although the iteration complexity of this batch algorithm is relatively high, it exhibits similar smooth convergence behavior for all the polytope examples in Section 2. We consider lower complexity implementations of LD-infomax as future work.

We next compared the proposed LD-infomax algorithm with the ICA-infomax algorithm (the implementation in Python MNE toolbox [19]) for different correlation parameter ρ values. Since the sources are correlated uniform random variables, we selected the number of sub-Gaussian parameters to be $r = 5$. The SINR vs. ρ curves for both algorithms in Figure 2 demonstrate that the ICA-infomax algorithm is prone to source correlations, whereas the LD-infomax algorithm is more immune. In fact, the LD-infomax algorithm can successfully separate correlated natural images from their mixtures[1]. We can attribute the slight degradation in the LD-infomax algorithm's performance to the increasing probability of sufficient scattering condition violation with the increasing correlation level.
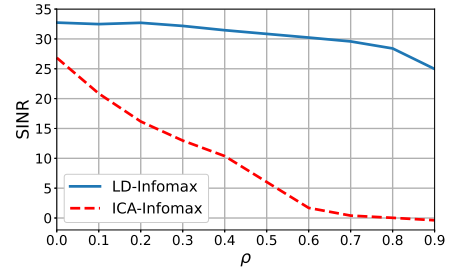

Fig. 2. SINR performance comparison of LD and ICA Infomax algorithms.

---

[1] We had to skip an illustrating example due to space limitations.

8. REFERENCES

[1] Pierre Comon and Christian Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic press, 2010.

[2] Pierre Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[3] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, "Independent component analysis, adaptive and learning systems for signal processing, communications, and control," *John Wiley & Sons, Inc*, vol. 1, pp. 11–14, 2001.

[4] Anthony J Bell and Terrence J Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[5] Pentti Paatero and Unto Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[6] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, October 1999.

[7] Xiao Fu, Kejun Huang, Bo Yang, Wing-Kin Ma, and Nicholas D Sidiropoulos, "Robust volume minimization-based matrix factorization for remote sensing and document clustering," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6254–6268, August 2016.

[8] Xiao Fu, Kejun Huang, Nicholas D Sidiropoulos, and Wing-Kin Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications.," *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 59–80, March 2019.

[9] Sergio Cruces, "Bounded component analysis of linear mixtures: A criterion of minimum convex perimeter," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2141–2154, January 2010.

[10] Alper T Erdogan, "A class of bounded component analysis algorithms for the separation of both independent and dependent sources," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5730–5743, August 2013.

[11] Huseyin A Inan and Alper T Erdogan, "A convolutive bounded component analysis framework for potentially nonstationary independent and/or dependent sources," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 18–30, November 2014.

[12] Michael Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer Science & Business Media, 2010.

[13] Pando Georgiev, Fabian Theis, and Andrzej Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 992–996, July 2005.

[14] Eren Babatas and Alper T Erdogan, "An algorithmic framework for sparse bounded component analysis," *IEEE Transactions on Signal Processing*, vol. 66, no. 19, pp. 5194–5205, August 2018.

[15] Gokcan Tatli and Alper T Erdogan, "Generalized polytopic matrix factorization," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, June 2021, pp. 3235–3239.

[16] Gokcan Tatli and Alper T. Erdogan, "Polytopic matrix factorization: Determinant maximization based criterion and identifiability," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5431–5447, 2021.

[17] Zhanghao Zhouyin and Ding Liu, "Understanding neural networks with logarithm determinant entropy estimator," *arXiv preprint arXiv:1401.3420*, 2021.

[18] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[19] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, no. 267, pp. 1–13, 2013.