SOURCE SEPARATION IN THE PRESENCE OF SIDE INFORMATION: NECESSARY AND SUFFICIENT CONDITIONS FOR RELIABLE DE-MIXING

Zahra Sabetsarvestani[†] *Francesco Renna* [⋆] *Franz Kiraly* [†] *Miguel R. D. Rodrigues* [†]

[†] University College London, U.K.
[⋆] Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal

ABSTRACT

This paper puts forth new recovery guarantees for the source separation problem in the presence of side information, where one observes the linear superposition of two source signals plus two additional signals that are correlated with the mixed ones. By positing that the individual components of the mixed signals as well as the corresponding side information signals follow a joint Gaussian mixture model, we characterise necessary and sufficient conditions for reliable separation in the asymptotic regime of low-noise as a function of the geometry of the underlying signals and their interaction. In particular, we show that if the subspaces spanned by the innovation components of the source signals with respect to the side information signals have zero intersection, provided that we observe a certain number of measurements from the mixture, then we can reliably separate the sources, otherwise we cannot. We also provide a number of numerical results on synthetic data that validate our theoretical findings.

1. INTRODUCTION

Source separation is the task that involves unmixing a mixture of signals into its constituents. It arises in many applications such as audio source separation [1], multiuser digital communication systems [2], cancer genetics [3], and more. The current literature is mostly focused on blind source separation (BSS), where one aims to recover the unobserved source signals given their linear mixtures. The BSS problem is intrinsically an ill-defined inverse problem, which requires some prior knowledge or additional assumptions in order to be solved.

Several methods have attempted to address the BSS problem by imposing different constraints on the source signals. A widely used technique for BSS is independent component analysis (ICA), where the source signals are separated by minimizing the mutual information between the sources under the assumption that the mixture components are non-Gaussian and statistically independent [4]. Another commonly used constraint in BSS is sparsity [5], where the BSS is solved by exploiting the fact that signals can often be described as linear combinations of a few atoms from a dictionary. Morphological component analysis [6], which has been developed to address the BSS problem, exploits both sparsity and morphological diversity.

There are various scenarios however where one can leverage additional information to aid source separation [7–10]. This framework is known as informed source separation and is mostly applied in audio source separation applications, where the audio objects are known in the encoding stage and a small amount of side information is transmitted to the decoder along with the mixture [11]. More recently, [10] proposed a novel coupled dictionary approach to exploit side information in separating X-ray images. The proposed approach couples the two image modalities, X-ray and RGB, by using a coupled dictionary learning algorithm. The beneficial use of side information in inverse problems has been well studied in previous works [12, 13], where it has also been theoretically and empirically proven that, in the presence of side information, fewer measurements are necessary and sufficient for perfect recovery.

This paper also studies source separation in the presence of side information. However, we stress that the primary aim of this paper is not to compare the performance of the proposed separation algorithm with the current state-of-the-art, but rather to analytically characterize the identifiability conditions of the source separation problem without and in the presence of side information in a general framework, a case which has never been treated in the literature, to the best of our knowledge.

In contrast with other works [7–10], we assume a joint Gaussian mixture model (GMM) to relate the individual components of the mixture to the side information. The GMM can be seen as the Bayesian counterpart of the union-of-subspaces model where each subspace corresponds to the image of the (possibly low rank) covariance matrix of each Gaussian component within the GMM [13, 14].The choice of this model is motivated by its ability to provide state-of-the-art results in different applications such as image processing [15,16], video compression [17] and dictionary learning [14, 18]. Moreover,

This work was in part supported by EPSRC under grant agreement EP/R032785/1.The work of F. Renna was funded by the FCT grant SFRH/BPD/118714/2016. F. Kiraly acknowledges support by the Alan Turing Institute, under EPSRC grant EP/N510129/1

using GMMs enables us to provide both necessary and sufficient conditions for the separation error to approach zero in the asymptotic regime of low noise.

We adopt the following notation throughout the paper: The identity matrix of dimension $n \times n$ is denoted by $\mathbf{I}_n$. The operators rank and Moore-Penrose pseudoinverse are denoted by $\text{rank}(\cdot)$ and $(\cdot)^\dagger$ respectively. $\text{Im}(\cdot)$ denotes the image of a matrix and $\dim(\cdot)$ denotes the dimension of a linear subspace. The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbb{E}(\cdot)$ denotes the expectation operator.

2. PROBLEM STATEMENT

We study a source separation problem where one aims to decompose a mixture of two signals into its constituents from a set of linear compressive measurements. In particular, we consider a standard linear mixing model:

$$\mathbf{v} = \boldsymbol{\Phi}(\mathbf{x}_1 + \mathbf{x}_2) + \mathbf{n}, \tag{1}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n_x}$ represents the measurement matrix, drawn from a rotationally invariant distribution, $\mathbf{n}$ is a zero-mean white Gaussian noise with the variance σ^2, i.e., $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m)$ The source signals $\mathbf{x}_1$, $\mathbf{x}_2$ are random vectors taking values in $\mathbb{R}^{n_x}$. In this model m represent the number of random measurements extracted from the linear mixture, where we assume $m < n_x$.

We also assume that the decoder has access to further noisy measurement vectors $\mathbf{y}_1$ and $\mathbf{y}_2$, where $\mathbf{y}_1 \in \mathbb{R}^{n_{y_1}}$ is correlated with $\mathbf{x}_1$, $\mathbf{y}_2 \in \mathbb{R}^{n_{y_2}}$ is correlated with $\mathbf{x}_2$ and $(\mathbf{x}_1, \mathbf{y_1})$ and $(\mathbf{x}_2, \mathbf{y_2})$ are statistically independent.

We further assume that the source signals and side-information signals follow a joint GMM. In particular, we consider the sets of labels $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{S}_1, \mathcal{S}_2$, where $\mathcal{C}_1$ and $\mathcal{C}_2$ are associated with the source signals $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, and $\mathcal{S}_1$ and $\mathcal{S}_2$ correspond to the side information signals $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively. Then, we assume that $\mathbf{x}_1$ and $\mathbf{y}_1$ follow a joint Gaussian mixture distribution, given by:

$$p(\mathbf{x}_1, \mathbf{y}_1) = \sum_{k_1=1}^{K_1} \sum_{j_1=1}^{J_1} p_{C_1,S_1}(C_1 = j_1, S_1 = k_1) \\ p(\mathbf{x}_1, \mathbf{y}_1 | C_1 = j_1, S_1 = k_1), \tag{2}$$

where $C_1 \in \mathcal{C}_1 = \{1, \dots, J_1\}$ and $S_1 \in \mathcal{S}_1 = \{1, \dots, K_1\}$. In this model, the probability distribution of $(\mathbf{x}_1, \mathbf{y}_1)$ conditioned on $(C_1, S_1) = (j_1, k_1)$ is a zero-mean [1] multivariate Gaussian,

$$\mathbf{x}_1, \mathbf{y}_1 | C_1 = j_1, S_1 = k_1 \sim \mathcal{N}(0, \bar{\boldsymbol{\Sigma}}_{x_1 y_1}^{(j_1 k_1)}), \tag{3}$$

where $\bar{\boldsymbol{\Sigma}}_{x_1 y_1}^{(j_1 k_1)} = \begin{bmatrix} \boldsymbol{\Sigma}_{x_1}^{(j_1 k_1)} & \boldsymbol{\Sigma}_{x_1 y_1}^{(j_1 k_1)} \\ \boldsymbol{\Sigma}_{y_1 x_1}^{(j_1 k_1)} & \boldsymbol{\Sigma}_{y_1}^{(j_1 k_1)} \end{bmatrix}$.

[1] In this work, we concentrate on zero mean signals, but the results can be also generalized from zero to non-zero mean signals.

In the same way we also model the distribution of $(\mathbf{x}_2, \mathbf{y}_2)$ using the GMM model given by:

$$p(\mathbf{x}_2, \mathbf{y}_2) = \sum_{k_2=1}^{K_2} \sum_{j_2=1}^{J_2} p_{C_2,S_2}(C_2 = j_2, S_2 = k_2) \\ p(\mathbf{x}_2, \mathbf{y}_2 | C_2 = j_2, S_2 = k_2), \tag{4}$$

where $C_2 \in \mathcal{C}_2 = \{1, \dots, J_2\}$ and $S_2 \in \mathcal{S}_2 = \{1, \dots, K_2\}$, and accordingly,

$$\mathbf{x}_2, \mathbf{y}_2 | C_2 = j_2, S_2 = k_2 \sim \mathcal{N}(0, \bar{\boldsymbol{\Sigma}}_{x_2 y_2}^{(j_2 k_2)}), \tag{5}$$

where, $\bar{\boldsymbol{\Sigma}}_{x_2 y_2}^{(j_2 k_2)} = \begin{bmatrix} \boldsymbol{\Sigma}_{x_2}^{(j_2 k_2)} & \boldsymbol{\Sigma}_{x_2 y_2}^{(j_2 k_2)} \\ \boldsymbol{\Sigma}_{y_2 x_2}^{(j_2 k_2)} & \boldsymbol{\Sigma}_{y_2}^{(j_2 k_2)} \end{bmatrix}$. Note that, conditioned on the class labels $(C_1, C_2, S_1, S_2) = (j_1 j_2 k_1 k_2) \in \mathcal{L} = \mathcal{C}_1 \times \mathcal{C}_2 \times \mathcal{S}_1 \times \mathcal{S}_2$, the joint distribution of two source signals $\left[\mathbf{x}_1^T \mathbf{x}_2^T\right]^T$ and two side-information signals $\left[\mathbf{y}_1^T \mathbf{y}_2^T\right]^T$ follows a zero mean Gaussian distribution $\mathcal{N}(0, \boldsymbol{\Sigma}_{xy}^{(j_1, j_2, k_1, k_2)})$. with covariance:

$$\boldsymbol{\Sigma}_{xy}^{(j_1 j_2 k_1 k_2)} = \begin{bmatrix} \boldsymbol{\Sigma}_{x_1}^{(j_1 k_1)} & \mathbf{0} & \boldsymbol{\Sigma}_{x_1 y_1}^{j_1 k_1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{x_2}^{(j_2 k_2)} & \mathbf{0} & \boldsymbol{\Sigma}_{x_2 y_2}^{(j_2 k_2)} \\ \boldsymbol{\Sigma}_{y_1 x_1}^{(j_1 k_1)} & \mathbf{0} & \boldsymbol{\Sigma}_{y_1}^{(j_1 k_1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{y_2 x_2}^{(j_2 k_2)} & \mathbf{0} & \boldsymbol{\Sigma}_{y_2}^{(j_2 k_2)} \end{bmatrix}.$$

It will be convenient to re-write the measurement model as follows:

$$\mathbf{w} = \boldsymbol{\Phi}_0 \mathbf{s} + \mathbf{n}_0, \tag{6}$$

where $\mathbf{n}_0 \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{m + n_{y_1} + n_{y_2}})$ and,

$$\boldsymbol{\Phi}_0 = \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{\Phi} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{n_{y_1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 & \mathbf{I}_{n_{y_2}} \end{bmatrix} \mathbf{s} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \mathbf{w} = \begin{bmatrix} \mathbf{v} \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}. \tag{7}$$

Here, we have assumed that the side information signals are contaminated with additive Gaussian noise $\mathcal{N}(0, \sigma^2 \mathbf{I}_{n_{y_1} + n_{y_2}})$. The goal of source separation in this case is to recover $\mathbf{x} = \left[\mathbf{x}_1^T \mathbf{x}_2^T\right]^T$ given $\mathbf{w}$ and separation performance is quantified via the reconstruction error:

$$\text{MMSE}_{x|w}(\sigma^2) = \mathbb{E}\left[\|\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{w}]\|^2\right]. \tag{8}$$

In particular, we will be providing necessary and sufficient conditions for the reconstruction error to approach zero as $\sigma^2 \to 0$. Such conditions are expressed in terms of the number of measurements m and quantities that are related to the geometry of the signals and the measurement matrix. In particular we will be using the following quantities:

- $r_{x_i}^{(j_i k_i)} = \text{rank}(\boldsymbol{\Sigma}_{x_i}^{(j_i k_i)})$ represents the dimension of the subspace spanned by the source signal $\mathbf{x}_i$ and $r_{y_i}^{(j_i k_i)} = \text{rank}(\boldsymbol{\Sigma}_{y_i}^{(j_i k_i)})$ represents the dimension of the subspace spanned by the side information signal $\mathbf{y}_i$, identified by class labels $C_i = j_i$ and $S_i = k_i$.

- $r_{x_iy_i}^{(j_ik_i)} = \text{rank}(\bar{\mathbf{\Sigma}}_{x_iy_i}^{(j_ik_i)})$ represents the dimension of the subspace spanned collectively by the source signal $\mathbf{x}_i$ and the side information signal $\mathbf{y}_i$, identified by the labels $C_i = j_i$ and $C_i = k_i$. The superscripts are dropped when the results hold for all possible choice of labels.

We are especially interested in the scenario where the covariance matrices $\mathbf{\Sigma}_{x_1}^{(j_1k_1)}$, $\mathbf{\Sigma}_{y_1}^{(j_1k_1)}$, $\mathbf{\Sigma}_{x_1y_1}^{(j_1k_1)}$, $\mathbf{\Sigma}_{x_2}^{(j_2k_2)}$, $\mathbf{\Sigma}_{y_2}^{(j_2k_2)}$ and $\mathbf{\Sigma}_{x_2y_2}^{(j_2k_2)}$ have low rank, strictly smaller than the ambient dimensions, and their associated images are subspaces drawn uniformly at the random from the corresponding Grassmann manifold[2].

3. SOURCE SEPARATION WITH SIDE INFORMATION

We now put forth necessary and sufficient conditions for reliable separation for two scenarios: i) source signals and side-information signals are drawn from a joint Gaussian distribution; ii) source signals and side-information signals follow the GMM described in Section 2 . It is clear that the former is a special case of latter for $J_1 = K_1 = J_2 = K_2 = 1$.

3.1. Gaussian Sources

We consider conditions for reliable separation of Gaussian sources in the presence of side information. We denote the MMSE associated to the separation of Gaussian signals with side information by $\text{MMSE}_{x|w}^{G}$. This leads to the following theorem.

Theorem 1. Consider the measurement model in (6), where $(\mathbf{x}_1, \mathbf{y}_1)$ and $(\mathbf{x}_2, \mathbf{y}_2)$ are drawn from joint Gaussian distributions described in (3) and (5), respectively. Then with probability 1, it holds:

$$\lim_{\sigma^2 \to 0} \text{MMSE}_{x|w}^{G}(\sigma^2) = 0 \iff$$
$$m \geq r_{x_1y_1} + r_{x_2y_2} - r_{y_1} - r_{y_2} \text{ and } \mathcal{D}_{x|y} = 0, \qquad (9)$$

where, $\mathcal{D}_{x|y} = \dim(\text{Im}(\mathbf{\Sigma}_{x_1|y_1}) \cap \text{Im}(\mathbf{\Sigma}_{x_2|y_2}))$ and $\mathbf{\Sigma}_{x_1|y_1} = \mathbf{\Sigma}_{x_1} - \mathbf{\Sigma}_{x_1y_1}\mathbf{\Sigma}_{y_1}^{\dagger}\mathbf{\Sigma}_{y_1x_1}$, $\mathbf{\Sigma}_{x_2|y_2} = \mathbf{\Sigma}_{x_2} - \mathbf{\Sigma}_{x_2y_2}\mathbf{\Sigma}_{y_2}^{\dagger}\mathbf{\Sigma}_{y_2x_2}$.

Remark 3.1. In the case where side information is not available,i.e., $\mathbf{y}_1 = \mathbf{y}_2 = 0$, irrespective of the number of measurements, a reliable separation of source signals is only feasible provided that the range spaces associated to the source signals have no overlap, i.e., $\mathcal{D}_x = \dim(\text{Im}(\mathbf{\Sigma}_{x_1}) \cap \text{Im}(\mathbf{\Sigma}_{x_2})) = 0$. Under this condition, $r_{x_1} + r_{x_2}$ measurements are necessary and sufficient to drive the MMSE to zero in the low-noise regime.

[2]Note that the assumption on the subspaces associated with covariance matrices is plausible as it reflects well the behaviour of many real data ensembles for various applications such as face recognition, digits classification and so on [19]. Moreover, it simplifies the statement of some of our theoretical results

Interestingly, Theorem 1 shows that the presence of side-information not only reduces the number of measurements, necessary and sufficient for reliable separation, but also relaxes the condition on the interaction between the subspaces associated with the two source signals. More intuitively, if the subspaces associated with the source signals intersect, we can still reliably separate the source signals in the presence of side information provided that such intersection is covered by some correlated portion of the side information (as suggested by the conditions in (9)).

Moreover under condition $\mathcal{D}_{x|y} = 0$, Theorem 1 shows that we can reliably separate the source signals in the presence of the side information provided that we observe at least $r_{xy} - r_{y_1} - r_{y_2}$ measurements extracted from the mixture, i.e., the dimension of the projected mixture is equal to or greater than the sum of the dimensions of two spaces spanned by the innovation components of the source signals with respect to side information signals.

3.2. GMM Sources

We now consider conditions for reliable separation of GMM sources, by focusing in scenarios where the side information is available.

The challenge relates to the absence of closed-form expressions for the MMSE in (8) associated with GMM sources. Therefore, to derive necessary conditions, we will be working with the following MMSE lower bound given by:

$$\text{MMSE}_{x|w}^{GM}(\sigma^2) = \mathbb{E}\left[\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{w})\|^2\right] =$$
$$\sum_{(j_1,k_1,j_2,k_2)\in\mathcal{L}} p_{C_1,S_1,C_2,S_2}(j_1,k_1,j_2,k_2)$$
$$\mathbb{E}\left[\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{w})\|^2 | C_1 = J_1, S_1 = K_1, C_2 = J_2, S_2 = K_2\right]$$
$$\geq \sum_{(j_1,k_1,j_2,k_2)\in\mathcal{L}} p_{C_1,S_1,C_2,S_2}(j_1,k_1,j_2,k_2)\text{MMSE}_{x|w}^{G}(\sigma^2),$$

where $\text{MMSE}_{x|w}^{G(j_1,k_1,j_2,k_2)}(\sigma^2)$ represents the MMSE associated with the recovery of the Gaussian component corresponding to the label (j_1, k_1, j_2, k_2) of the GMM signal given the measurements of the linear mixture and given the side information. This leads immediately to the following theorem.

Theorem 2. Consider the measurement model in (6), where the source signals $\mathbf{x}_1$, $\mathbf{x}_2$ and the side-information signals $\mathbf{y}_1$, $\mathbf{y}_2$ are drawn from the joint GMM distribution described in Section 2 and $(\mathbf{x}_1, \mathbf{y}_1)$ and $(\mathbf{x}_2, \mathbf{y}_2)$ are statistically independent. Then, with probability 1, it holds:

$$\lim_{\sigma^2 \to 0} \text{MMSE}_{x|w}^{GM}(\sigma^2) = 0 \Rightarrow$$
$$m \geq r_{x_1y_1}^{(j_1k_1)} +_{x_2y_2}^{(j_2k_2)} - r_{y_1}^{(j_1k_1)} - r_{y_2}^{(j_2k_2)} \text{ and } \mathcal{D}_{x|y}^{(j_1k_1j_2k_2)} = 0$$

[2]The inequality is the consequence of the optimality of the MMSE estimator for the Gaussian sources and side information.

$\forall(j_1k_1j_2k_2) \in \mathcal{L}$, where $\mathcal{D}_{x|y}^{(j_1k_1j_2k_2)} = \dim(\mathrm{Im}(\mathbf{\Sigma}_{x_1|y_1}^{(j_1k_1)}) \cap \mathrm{Im}(\mathbf{\Sigma}_{x_2|y_2}^{(j_2k_2)}))$ and $\mathbf{\Sigma}_{x_1|y_1}^{(j_1k_1)} = \mathbf{\Sigma}_{x_1}^{(j_1k_1)} - \mathbf{\Sigma}_{x_1y_1}^{(j_1k_1)} \mathbf{\Sigma}_{y_1}^{(j_1k_1)^\dagger} \mathbf{\Sigma}_{y_1x_1}^{(j_1k_1)}$, $\mathbf{\Sigma}_{x_2|y_2}^{(j_2k_2)} = \mathbf{\Sigma}_{x_2}^{(j_2k_2)} - \mathbf{\Sigma}_{x_2y_2}^{(j_2k_2)} \mathbf{\Sigma}_{y_2}^{(j_2k_2)^\dagger} \mathbf{\Sigma}_{y_2x_2}^{(j_2k_2)}$.

In turn, to derive sufficient conditions, we will be considering an MMSE upper bound, $\mathrm{MSE}^{CS}(\sigma^2)$, associated with a specific two-step classify and separate (CS) decoder which operates in two main steps as follows:

- Classification step: First the label quadruple associated to the source signals and side information signals $(\hat{C}_1, \hat{C}_2, \hat{S}_1, \hat{S}_2)$ is estimated via a maximum *a posteriori* (MAP) classifier.
- Separation step: Second, the source signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are recovered via the Gaussian conditional mean estimator associated to class labels $(\hat{C}_1, \hat{C}_2, \hat{S}_1, \hat{S}_2)$.

This now leads to the following theorem.

Theorem 3. Consider the measurement model in (6), where the source signals $\mathbf{x}_1, \mathbf{x}_2$ and the side-information signals $\mathbf{y}_1$, $\mathbf{y}_2$ are drawn from the joint GMM distribution described in Section 2. Then, with probability 1, it holds $\forall(j_1k_1j_2k_2) \in \mathcal{L}$:

$$m > r_{xy}^{(j_1k_1j_2k_2)} - r_{y_1}^{(j_1k_1)} - r_{y_2}^{(j_2k_2)} \text{and } \mathcal{D}_{x|y}^{(j_1k_1j_2k_2)} = 0$$
$$\Rightarrow \lim_{\sigma^2 \to 0} \mathrm{MMSE}_{x|w}^{GM}(\sigma^2) = 0$$

The results from Theorems 2 and 3 state that the spaces spanned by conditional covariances, i.e., the space spanned by signal components which are not correlated with the side informations, ought to have no intersection for all possible label quadruples (C_1, S_1, C_2, S_2) for reliable separation. Moreover, the measurements extracted from the mixture should be enough to capture the components of source signals which are not correlated with the side informations for all Gaussian components. Notably, the provided conditions for reliable separations are shown to be tight, as the necessary conditions are only one measurement away from the sufficient conditions.

4. SIMULATION RESULTS

We now provide some numerical results to illustrate our theory. In particular, we will show how the interplay between the number of linear measurements from the mixture and the properties of the individual components of the mixture impacts on the quality of separation. In our simulations, we use the optimal conditional mean estimator to recover the individual signal components from the observations. We also use random measurement matrices whose entries are i.i.d., Gaussian random variables with zero mean and unit variance which have been normalized so that it holds $\mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{I}$. The low rank covariance matrices are generated by the product of a random matrix with its transpose where the entries of the matrix are i.i.d Gaussian random variables with zero mean and unit variance. Fig. 1 shows the MMSE associated with the separation of GMM signals with $J_1 = J_2 = K_1 = K_2 = 2$, and with dimensions $n_{x_1} = n_{x_2} = 10$, in the presence of side informations where $r_{x_1y_1} = r_{x_2y_2} = 3$, $r_{x_1} = r_{x_2} = 2$ and $r_{y_1} = r_{y_2} = 2$ for all class labels. We report the actual values of the $\mathrm{MMSE}_{x|w}^{GM}(\sigma^2)$, the lower bound and the upper bound $\mathrm{MSE}^{CS}(\sigma^2)$ associated to the classify and separate decoder. We observe that the reliable separation is achieved with $m = 3$ for the lower bound, the upper bound and the actual MMSE, although the curves offset are slightly different.

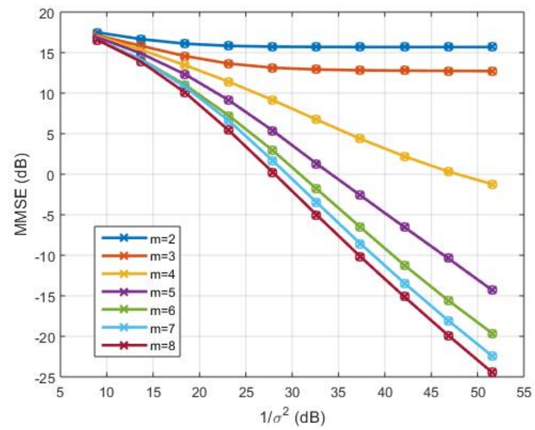

Fig. 1: MMSE associated to the separation of two GMM sources vs. $1/\sigma^2$ for different number of random measurements $m = 2$ to $m = 8$. The actual MMSE is represented by solid lines, the CS upper bound is represented by circled solid lines, and the lower bound by crossed solid lines.

5. CONCLUSION

We proposed a novel framework to incorporate side information in the source separation problem by using a joint GMM for the source and side information signals. Our main contribution is the characterization of necessary and sufficient conditions for reliable separation of source signals from a linear mixture in the presence of side information. We proved analytically that the presence of side information not only reduces the number of measurements required for reliable separation of the source signals, but also alleviates the limitations of the source separation problem in terms of the geometrical properties of sources. We showed via a range of simulation results on synthetic data that our theory is aligned with actual simulation results.

6. REFERENCES

[1] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, “Underdetermined blind audio source separation using modal decomposition,” *EURASIP J. Audio Speech Music Process.*, vol. 2007, no. 1.

[2] M. Brandt-Pearce, “Signal separation using fractional sampling in multiuser communications,” *IEEE Transactions on Communications*, vol. 48, no. 2, pp. 242–251, Feb 2000.

[3] C. Yau, O. Papaspiliopoulos, G. O. Roberts, and C. Holmes, “Bayesian non-parametric hidden markov models with applications in genomics,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 73, no. 1, pp. 37–57, 2011.

[4] A. Hyvärinen and E. Oja, “Independent component analysis: Algorithms and applications,” *Neural Networks*, vol. 13, pp. 411–430, 2000.

[5] M. Zibulevsky and B. A. Pearlmutter, “Blind source separation by sparse decomposition in a signal dictionary,” *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[6] J. Bobin, J. L. Starck, J. Fadili, and Y. Moudden, “Sparsity and morphological diversity in blind source separation,” *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, Nov 2007.

[7] A. Liutkus, J. Pinel, R. Badeau, L. Girin, and G. Richard, “Informed source separation through spectrogram coding and data embedding,” *Signal Processing*, vol. 92, no. 8, pp. 1937 – 1949, 2012.

[8] M. Parvaix, L. Girin, and J. M. Brossier, “A watermarking-based method for informed source separation of audio signals with a single sensor,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1464–1475, Aug 2010.

[9] S. Gorlow and S. Marchand, “Informed separation of spatial images of stereo music recordings using second-order statistics,” in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept 2013, pp. 1–6.

[10] N. Deligiannis, J. F. C. Mota, B. Cornelis, M. R. D. Rodrigues, and I. Daubechies, “Multi-modal dictionary learning for image separation with application in art investigation,” *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 751–764, Feb 2017.

[11] M. Parvaix, L. Girin, and J. M. Brossier, “A watermarking-based method for informed source separation of audio signals with a single sensor,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1464–1475, Aug 2010.

[12] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, “Compressed sensing with prior information: Strategies, geometry, and bounds,” *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4472–4496, July 2017.

[13] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. R. D. Rodrigues, “Classification and reconstruction of high-dimensional signals from low-dimensional features in the presence of side information,” *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6459–6492, Nov 2016.

[14] F. Renna, R. Calderbank, L. Carin, and M. R. D. Rodrigues, “Reconstruction of signals drawn from a gaussian mixture via noisy compressive measurements,” *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2265–2277, May 2014.

[15] G. Yu and G. Sapiro, “Statistical compressed sensing of gaussian mixture models,” *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5842–5858, Dec 2011.

[16] G. Yu, G. Sapiro, and S. Mallat, “Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, May 2012.

[17] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, “Video compressive sensing using gaussian mixture models,” *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4863–4878, Nov 2014.

[18] M. Chen, J. Silva, J. Paisley, C. Wang, D. Dunson, and L. Carin, “Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds,” *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6140–6155, Dec 2010.

[19] M. Chen, W. Carson, M. Rodrigues, R. Calderbank, and L. Carin, “Communications inspired linear discriminant analysis,” in *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ser. ICML'12. USA: Omnipress, 2012, pp. 1507–1514. [Online]. Available: http://dl.acm.org/citation.cfm?id=3042573.3042766