

Classification and Reconstruction of Compressed GMM Signals with Side Information

Francesco Renna*, Liming Wang[†], Xin Yuan[†], Jianbo Yang[‡], Galen Reeves[†], Robert Calderbank[†],
Lawrence Carin[†] and Miguel R. D. Rodrigues*

*Department of Electronic and Electrical Engineering, University College London, UK

[†]Department of Electrical and Computer Engineering, Duke University, NC, USA

[‡]Institute for Infocomm Research, Singapore

Email: {f.renna, m.rodrigues}@ucl.ac.uk, yang-j@i2r.a-star.edu.sg,
{liming.w, xin.yuan, galen.reeves, robert.calderbank, lcarin}@duke.edu

Abstract—This paper offers a characterization of performance limits for classification and reconstruction of high-dimensional signals from noisy compressive measurements, in the presence of side information. We assume the signal of interest and the side information signal are drawn from a correlated mixture of distributions/components, where each component associated with a specific class label follows a Gaussian mixture model (GMM).

We provide sharp sufficient and/or necessary conditions for the phase transition of the misclassification probability and the reconstruction error in the low-noise regime. These conditions, which are reminiscent of the well-known Slepian-Wolf and Wyner-Ziv conditions, are a function of the number of measurements taken from the signal of interest, the number of measurements taken from the side information signal, and the geometry of these signals and their interplay.

Index Terms—Classification, reconstruction, Gaussian mixture models, side information.

I. INTRODUCTION

Compressive sensing (CS) is a signal acquisition paradigm that offers the means to simultaneously sense and compress a signal without any or with minimal loss of information [1], [2]. In particular, this emerging paradigm shows that it is possible to perfectly reconstruct an n -dimensional s -sparse signal (sparse in some orthonormal dictionary or frame) with overwhelming probability with only $\mathcal{O}(s \log(n/s))$ linear random measurements or projections. The signal recovery is performed using tractable ℓ_1 minimization methods [3] or iterative methods, like greedy matching pursuit [4]. Generalizations of the CS paradigm to settings including other signal processing operations in the compressive domain, such as detection and classification, have also become popular recently [5].

However, it is often the case that one is also presented at the decoder with additional information – known as *side information* – in the form of another signal that exhibits some correlation with the signal of interest. This paper is concerned

with the impact of side information in the classification and reconstruction of high-dimensional signals from compressive measurements.

The problem of reconstruction of a sparse signal in the presence of partial information about the desired signal, using reconstruction algorithms akin to those from CS, has been recently considered [6]–[10]. For example, [6] considers the reconstruction of a signal by leveraging partial information at the decoder about the support of the signal; [7] considers the reconstruction of the signal by using an additional noisy version of the signal at the decoder. In [8] the authors take the side information to be associated with the previous scans of a certain subject in dynamic tomographic imaging. In this case, ℓ_1 -norm based minimization is used for recovery, by adding an additional term that accounts for the distance between the recovered image and the side information snapshot. A similar approach has been adopted recently in [9], that is shown to require a smaller number of measurements than traditional CS in recovering magnetic resonance images. A theoretical analysis of the number of measurements sufficient for reliable recovery with high probability in the presence of side information for both ℓ_1/ℓ_1 and mixed ℓ_1/ℓ_2 reconstruction strategies is provided in [10].

This paper studies the impact of side information on the *classification* and *reconstruction* of a high-dimensional signal from noisy, compressive measurements, by assuming that both the signal of interest and the side information are drawn from a joint Gaussian mixture model (GMM). There are multiple reasons for adopting a GMM representation, which can be seen as a union of (linear or affine) subspaces, where each subspace is associated with the translation of the image of the (possibly low-rank) covariance matrix of each Gaussian component within the GMM. In fact, low-rank GMM priors have been shown to approximate signals in compact manifolds [11] and have been shown to provide state-of-the-art results in practical problems in image processing [12], dictionary learning [11], image classification [13] and video compression [14]. Of particular relevance, the adoption of GMM priors also offers an opportunity to analyze phase transitions in the classification

The work of F. Renna was supported by Project I-CITY - ICT for Future Health/Faculdade de Engenharia da Universidade do Porto, NORTE-07-0124-FEDER-000068, funded by the Fundo Europeu de Desenvolvimento Regional (FEDER) through the Programa Operacional do Norte (ON2) and by national funds, through FCT/MEC (PIDDAC). The work of M. Rodrigues was supported by the EPSRC grant EP/K033166/1. The work of R. Calderbank was supported in part by AFOSR under grant FA 9550-13-1-0076 and by NGA under grant HM017713-1-0006.

or reconstruction error [15]–[17]¹.

Our main contribution is the analysis of phase transition regions associated to the classification and reconstruction error in the presence of side information at the decoder.

In the remainder, we denote matrices with boldface uppercase letters (\mathbf{X}) and column vectors with boldface lower-case letters (\mathbf{x}). The symbols \mathbf{I}_n and $\mathbf{0}_{m \times n}$ represent the identity matrix of dimension $n \times n$ and the all-zero-entries matrix of dimension $m \times n$, respectively (subscripts will be dropped when the dimensions are clear from the context). $\text{Im}(\cdot)$ denotes the (column) image of a matrix. $\mathbb{E}[\cdot]$ represents the expectation operator.

For reasons of space, we relegate the mathematical proofs of our results to the extended version of the work [18].

II. MODEL

We consider both the classification and reconstruction of a high-dimensional signal from noisy compressive measurements in the presence of side information, as shown in Fig. 1. In particular, we assume that the decoder has access to a set of noisy linear measurements $\mathbf{y}_1 \in \mathbb{R}^{m_1}$ associated with the desired signal $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ given by:

$$\mathbf{y}_1 = \Phi_1 \mathbf{x}_1 + \mathbf{w}_1, \quad (1)$$

where $\Phi_1 \in \mathbb{R}^{m_1 \times n_1}$ is the projection kernel² and $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \cdot \sigma^2)$ is additive Gaussian noise that models noise introduced by the measurement process. We also assume that the decoder has access to another set of linear measurements $\mathbf{y}_2 \in \mathbb{R}^{m_2}$ – called *side information* – associated with another signal $\mathbf{x}_2 \in \mathbb{R}^{n_2}$ given by:

$$\mathbf{y}_2 = \Phi_2 \mathbf{x}_2 + \mathbf{w}_2, \quad (2)$$

where $\Phi_2 \in \mathbb{R}^{m_2 \times n_2}$ is the projection kernel associated with the side information and $\mathbf{w}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \cdot \sigma^2)$ is Gaussian additive noise, which is assumed, for simplicity, to have the same covariance as the noise \mathbf{w}_1 . We focus on random projection kernels, where both matrices Φ_1 and Φ_2 are assumed to be drawn from left rotation-invariant distributions³. For the sake of compact notation, we also use the symbol

$$\Phi = \begin{bmatrix} \Phi_1 & \mathbf{0} \\ \mathbf{0} & \Phi_2 \end{bmatrix}. \quad (3)$$

We consider underlying class labels $C_1 \in \{1, \dots, K_1\}$ and $C_2 \in \{1, \dots, K_2\}$, where C_1 is associated with the signal of interest \mathbf{x}_1 and C_2 is associated with the side information signal \mathbf{x}_2 . We assume that \mathbf{x}_1 and \mathbf{x}_2 , conditioned on the

¹In the low-noise regime, classification and reconstruction depend only on the dimensions of subspaces in the model and their intersections. We are extending this form of analysis.

²In the remainder of the paper, we will use interchangeably the terms projection/measurement/sensing kernel or matrix.

³A random matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to be (left or right) rotation-invariant if the joint probability density function (pdf) of its entries $p(\mathbf{A})$ satisfies $p(\Theta \mathbf{A}) = p(\mathbf{A})$, or $p(\mathbf{A} \Psi) = p(\mathbf{A})$, respectively, for any orthogonal matrix Θ or Ψ . A special case of (left and right) rotation-invariant random matrices is represented by matrices with independent identically distributed (i.i.d.), zero-mean Gaussian entries with fixed variance.

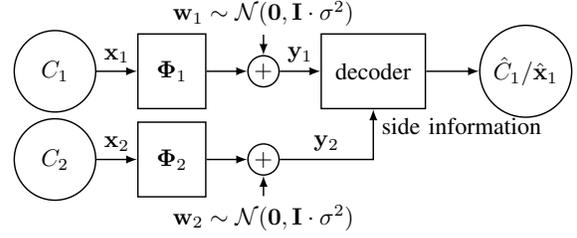


Fig. 1. Classification and reconstruction with side information. The user attempts to generate an estimate \hat{C}_1 of the index of the component from which the input signal \mathbf{x}_1 was drawn (classification) or it aims to generate an estimate $\hat{\mathbf{x}}_1$ of the input signal itself (reconstruction) on the basis of the observation of both measurement vectors \mathbf{y}_1 and \mathbf{y}_2 .

underlying class labels $C_1 = i$ and $C_2 = k$, are drawn from a joint distribution $p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k)$, with the class labels drawn with probability $p_{C_1, C_2}(i, k)$. We assume that the decoder, for both classification and reconstruction purposes, knows perfectly the joint probability mass function (pmf) $p_{C_1, C_2}(i, k)$ of the discrete random variables corresponding to the class labels of \mathbf{x}_1 and \mathbf{x}_2 , and the conditional distributions $p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k)$. For the problem of classification with side information, the objective is to estimate the value of the index C_1 that identifies the distribution/component from which \mathbf{x}_1 was drawn, on the basis of the observation of both vectors \mathbf{y}_1 and \mathbf{y}_2 . The minimum average error probability in classifying C_1 from \mathbf{y}_1 and \mathbf{y}_2 is achieved by the maximum *a posteriori* (MAP) classifier [19], given by

$$\hat{C}_1 = \arg \max_{i \in \{1, \dots, K_1\}} p(C_1 = i | \mathbf{y}_1, \mathbf{y}_2), \quad (4)$$

where $p(C_1 = i | \mathbf{y}_1, \mathbf{y}_2)$ is the *a posteriori* probability of class $C_1 = i$ conditioned on \mathbf{y}_1 and \mathbf{y}_2 .

For the problem of reconstruction with side information, the objective of the decoder is to estimate the signal \mathbf{x}_1 from the observation of \mathbf{y}_1 and \mathbf{y}_2 . In particular, we consider reconstruction obtained via the conditional mean estimator

$$\hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2) = \mathbb{E}[\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2] = \int \mathbf{x}_1 p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2) d\mathbf{x}_1, \quad (5)$$

where $p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2)$ is the *posterior* pdf of \mathbf{x}_1 given \mathbf{y}_1 and \mathbf{y}_2 , which minimizes the reconstruction error.

A. Signal, Side Information and Correlation Models

We adopt a multivariate Gaussian model for the distribution of \mathbf{x}_1 and \mathbf{x}_2 , conditioned on $(C_1, C_2) = (i, k)$, i.e.

$$p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(ik)}, \boldsymbol{\Sigma}_{\mathbf{x}}^{(ik)}), \quad (6)$$

where

$$\boldsymbol{\mu}_{\mathbf{x}}^{(ik)} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_1}^{(ik)} \\ \boldsymbol{\mu}_{\mathbf{x}_2}^{(ik)} \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\mathbf{x}}^{(ik)} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(ik)} & \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(ik)} \\ \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(ik)} & \boldsymbol{\Sigma}_{\mathbf{x}_2}^{(ik)} \end{bmatrix}, \quad (7)$$

so that $p(\mathbf{x}_1 | C_1 = i, C_2 = k) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_1}^{(ik)}, \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(ik)})$ and $p(\mathbf{x}_2 | C_1 = i, C_2 = k) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_2}^{(ik)}, \boldsymbol{\Sigma}_{\mathbf{x}_2}^{(ik)})$.

The motivation for this choice is associated by the fact that

this apparently simple model can accommodate a wide range of signal distributions. In fact, note that the joint pdf of \mathbf{x}_1 and \mathbf{x}_2 follows a GMM model, so that we can in principle approximate very complex distributions by incorporating additional terms in the decomposition [20]. Note also that the conditional marginal pdfs of \mathbf{x}_1 and \mathbf{x}_2 also follow GMM models.

Then, we define the ranks of the matrices appearing in the model in (7) as follows:

$$r_{\mathbf{x}_1}^{(ik)} = \text{rank}(\Sigma_{\mathbf{x}_1}^{(ik)}) \quad , \quad r_{\mathbf{x}_2}^{(ik)} = \text{rank}(\Sigma_{\mathbf{x}_2}^{(ik)}) \quad (8)$$

which represent the dimensions of the subspaces spanned by input signals \mathbf{x}_1 and side information signals \mathbf{x}_2 , respectively, drawn from the Gaussian distribution corresponding to the indices $C_1 = i, C_2 = k$;

$$r_{\mathbf{x}_1}^{(ik,j\ell)} = \text{rank}(\Sigma_{\mathbf{x}_1}^{(ik)} + \Sigma_{\mathbf{x}_1}^{(j\ell)}) \quad (9)$$

which represents the dimension of the sum of the subspaces spanned by input signals drawn from the Gaussian distribution corresponding to the indices $C_1 = i, C_2 = k$ and those from the Gaussian distribution corresponding to the indices $C_1 = j, C_2 = \ell$; analogously, we define

$$r_{\mathbf{x}_2}^{(ik,j\ell)} = \text{rank}(\Sigma_{\mathbf{x}_2}^{(ik)} + \Sigma_{\mathbf{x}_2}^{(j\ell)}); \quad (10)$$

finally, the corresponding dimensions spanned collectively by input and side information signals are given by

$$r_{\mathbf{x}}^{(ik)} = \text{rank}(\Sigma_{\mathbf{x}}^{(ik)}) \quad , \quad r_{\mathbf{x}}^{(ik,j\ell)} = \text{rank}(\Sigma_{\mathbf{x}}^{(ik)} + \Sigma_{\mathbf{x}}^{(j\ell)}). \quad (11)$$

We also define the ranks:

$$r^{(ik)} = \text{rank} \left(\Phi \Sigma_{\mathbf{x}}^{(ik)} \Phi^T \right) \quad (12)$$

$$r^{(ik,j\ell)} = \text{rank} \left(\Phi (\Sigma_{\mathbf{x}}^{(ik)} + \Sigma_{\mathbf{x}}^{(j\ell)}) \Phi^T \right), \quad (13)$$

that represent the dimension of the subspace spanned collectively by the projections of input signals and the projections of side information signals drawn from the Gaussian distribution identified by the component indices $C_1 = i, C_2 = k$, and the dimension associated to the sum of the subspaces corresponding to classes $C_1 = i, C_2 = k$ and $C_1 = j, C_2 = \ell$.

The quantities in (8)–(13), which provide a concise description of the geometry of the input source, the side information source, and the geometry of the interaction of such sources with the projections kernels, will be fundamental to determining the performance of the classification and reconstruction of high-dimensional signals from compressive measurements in the presence of side information.

III. CLASSIFICATION WITH SIDE INFORMATION

We first consider signal classification in the presence of side information. The basis of the analysis is an asymptotic characterization – in the limit of $\sigma^2 \rightarrow 0$ – of the behavior of an upper bound to the misclassification probability associated with the optimal MAP classifier (rather than the exact misclassification probability which is not tractable). In particular, via the Bhattacharyya bound [19] in conjunction with the union bound, the misclassification probability can be upper bounded

as follows:

$$\begin{aligned} \bar{P}_{\text{err}} = & \sum_{i=1}^{K_1} \sum_{\substack{j=1 \\ j \neq i}}^{K_1} p_{C_1}(i) \int \sqrt{\sum_{k,\ell=1}^{K_2} p_{C_2|C_1}(k|i) p_{C_2|C_1}(\ell|j)} \\ & \cdot \sqrt{p(\mathbf{y}_1, \mathbf{y}_2 | C_1 = i, C_2 = k)} \\ & \cdot \sqrt{p(\mathbf{y}_1, \mathbf{y}_2 | C_1 = j, C_2 = \ell)} d\mathbf{y}_1 d\mathbf{y}_2, \end{aligned} \quad (14)$$

where $p_{C_2|C_1}(k|i) = \frac{p_{C_1, C_2}(i, k)}{p_{C_1}(i)}$ and $p_{C_1|C_2}(i|k) = \frac{p_{C_1, C_2}(i, k)}{p_{C_2}(k)}$ are the conditional pmfs of C_2 and C_1 .

The asymptotic characterization that we discuss below identifies the presence or absence of an error floor in the upper bound to the misclassification probability as $\sigma^2 \rightarrow 0$, leading to conditions on the number of measurements that guarantee perfect classification in the low-noise regime, i.e.,

$$\lim_{\sigma^2 \rightarrow 0} \bar{P}_{\text{err}}(\sigma^2) = 0. \quad (15)$$

Note that the characterization of the presence or absence of an error floor in the upper bound of the misclassification probability also leads to the characterization of a phase transition region in terms of m_1 and m_2 , where within this region (15) holds and outside the region (15) is not verified. Note also that the boundaries of the region associated to the upper bound of the misclassification probability represent also lower bounds of the boundaries of the corresponding region associated with the true error probability.

Note that all the pairs of indices (i, k) such that $p_{C_1, C_2}(i, k) = 0$ clearly do not affect the phase transition region. Therefore, we can define the set of index pairs of interest as

$$\mathcal{S} = \{(i, k) \in \{1, \dots, K_1\} \times \{1, \dots, K_2\} : p_{C_1, C_2}(i, k) > 0\}. \quad (16)$$

We also define the set of index quadruples

$$\mathcal{S}_{\text{SIC}} = \{(i, k, j, \ell) : (i, k), (j, \ell) \in \mathcal{S}, i \neq j\}. \quad (17)$$

In the next theorem, we provide conditions on the number of measurements m_1 and m_2 that are sufficient to drive the upper bound to the misclassification probability to zero when $\sigma^2 \rightarrow 0$, that is, in order to achieve the phase transition of the upper bound to the misclassification probability, and, therefore, that are also sufficient to drive the true misclassification probability to zero when $\sigma^2 \rightarrow 0$.

Theorem 1: Consider the model in (1) and (2), where the input signal \mathbf{x}_1 and the side information signal \mathbf{x}_2 are drawn according to the class-conditioned joint distribution (6).

If $r_{\mathbf{x}}^{(ik,j\ell)} > r_{\mathbf{x}}^{(ik)}, r_{\mathbf{x}}^{(j\ell)}, \forall (i, k, j, \ell) \in \mathcal{S}_{\text{SIC}}$, then, with probability 1, the upper bound to the misclassification probability (14) approaches zero when $\sigma^2 \rightarrow 0$ if the following conditions hold $\forall (i, k, j, \ell) \in \mathcal{S}_{\text{SIC}}$:

- 1) if $r_{\mathbf{x}_1}^{(ik,j\ell)} > r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}_1}^{(j\ell)}$ and $r_{\mathbf{x}_2}^{(ik,j\ell)} > r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}_2}^{(j\ell)}$:

$$m_1 > \min\{r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}_1}^{(j\ell)}\} \quad \text{or} \quad m_2 > \min\{r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}_2}^{(j\ell)}\}$$

$$\text{or} \quad m_1 + m_2 > \min\{r_{\mathbf{x}}^{(ik)}, r_{\mathbf{x}}^{(j\ell)}\}; \quad (18)$$

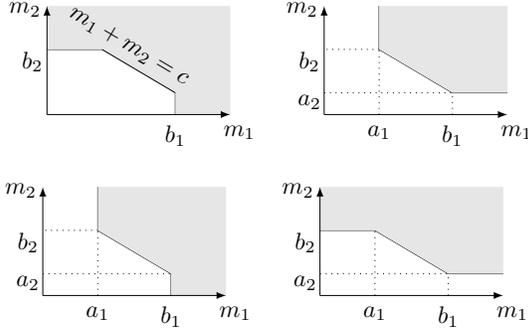


Fig. 2. Representation of the conditions on m_1 and m_2 for phase transition, for the 4 different cases encapsulated in Theorem 1. In all cases $a_1 = \min\{r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}}^{(j\ell)} - r_{\mathbf{x}_2}^{(j\ell)}\} + 1$, $b_1 = \min\{r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}_1}^{(j\ell)}\} + 1$, $a_2 = \min\{r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}}^{(j\ell)} - r_{\mathbf{x}_1}^{(j\ell)}\} + 1$, $b_2 = \min\{r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}_2}^{(j\ell)}\} + 1$ and $c = \min\{r_{\mathbf{x}}^{(ik)}, r_{\mathbf{x}}^{(j\ell)}\} + 1$. The shaded regions represent values of m_1 and m_2 that satisfy the conditions (18)–(21).

$$2) \text{ if } r_{\mathbf{x}_1}^{(ik,j\ell)} = r_{\mathbf{x}_1}^{(ik)} = r_{\mathbf{x}_1}^{(j\ell)} \text{ and } r_{\mathbf{x}_2}^{(ik,j\ell)} = r_{\mathbf{x}_2}^{(ik)} = r_{\mathbf{x}_2}^{(j\ell)}:$$

$$\begin{cases} m_1 > \min\{r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}}^{(j\ell)} - r_{\mathbf{x}_2}^{(j\ell)}\} \\ m_2 > \min\{r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}}^{(j\ell)} - r_{\mathbf{x}_1}^{(j\ell)}\} \\ m_1 + m_2 > \min\{r_{\mathbf{x}}^{(ik)}, r_{\mathbf{x}}^{(j\ell)}\} \end{cases}; \quad (19)$$

$$3) \text{ if } r_{\mathbf{x}_1}^{(ik,j\ell)} > r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}_1}^{(j\ell)} \text{ and } r_{\mathbf{x}_2}^{(ik,j\ell)} = r_{\mathbf{x}_2}^{(ik)} = r_{\mathbf{x}_2}^{(j\ell)}:$$

$$m_1 > \min\{r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}_1}^{(j\ell)}\}$$

$$\text{or } \begin{cases} m_1 > \min\{r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}}^{(j\ell)} - r_{\mathbf{x}_2}^{(j\ell)}\} \\ m_1 + m_2 > \min\{r_{\mathbf{x}}^{(ik)}, r_{\mathbf{x}}^{(j\ell)}\} \end{cases}; \quad (20)$$

$$4) \text{ if } r_{\mathbf{x}_1}^{(ik,j\ell)} = r_{\mathbf{x}_1}^{(ik)} = r_{\mathbf{x}_1}^{(j\ell)} \text{ and } r_{\mathbf{x}_2}^{(ik,j\ell)} > r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}_2}^{(j\ell)}:$$

$$m_2 > \min\{r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}_2}^{(j\ell)}\}$$

$$\text{or } \begin{cases} m_2 > \min\{r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}}^{(j\ell)} - r_{\mathbf{x}_1}^{(j\ell)}\} \\ m_1 + m_2 > \min\{r_{\mathbf{x}}^{(ik)}, r_{\mathbf{x}}^{(j\ell)}\} \end{cases}. \quad (21)$$

The characterization of the numbers of measurements m_1 and m_2 that are sufficient to achieve the phase transition in the upper bound to the misclassification probability is divided into 4 cases, depending on whether the range spaces $\text{Im}(\Sigma_{\mathbf{x}_1}^{(ik)})$ and $\text{Im}(\Sigma_{\mathbf{x}_1}^{(j\ell)})$, or the range spaces $\text{Im}(\Sigma_{\mathbf{x}_2}^{(ik)})$ and $\text{Im}(\Sigma_{\mathbf{x}_2}^{(j\ell)})$, are distinct or not⁴. Fig. 2 depicts the tradeoff between the values of m_1 and m_2 associated with these different cases. Note also that the values of m_1 and m_2 associated with the phase transition of the upper bound of the misclassification probability lie in the intersection of the regions corresponding to index quadruples $(i, k, j, \ell) \in \mathcal{S}_{\text{SIC}}$.

In case 1), the range spaces associated to the input covariance matrices are all distinct, and by observing (18) we can clearly determine the beneficial effect of the correlation

⁴We recall that, given two positive semidefinite matrices \mathbf{A} and \mathbf{B} with ranks $r_{\mathbf{A}} = \text{rank}(\mathbf{A})$, $r_{\mathbf{B}} = \text{rank}(\mathbf{B})$, $r_{\mathbf{A+B}} = \text{rank}(\mathbf{A+B})$, $\text{Im}(\mathbf{A}) = \text{Im}(\mathbf{B})$ if and only if $r_{\mathbf{A+B}} = \frac{r_{\mathbf{A}} + r_{\mathbf{B}}}{2}$ [17, Lemma 2] and then, if and only if $r_{\mathbf{A+B}} = r_{\mathbf{A}} = r_{\mathbf{B}}$.

between \mathbf{x}_1 and \mathbf{x}_2 in guaranteeing the phase transition for the upper bound to the misclassification probability. Namely, we note that the phase transition is achieved either when error-free classification is possible from the observation of \mathbf{y}_1 alone ($m_1 > \min\{r_{\mathbf{x}_1}^{(ik)}, r_{\mathbf{x}_1}^{(j\ell)}\}$) or from the observation of \mathbf{y}_2 alone ($m_2 > \min\{r_{\mathbf{x}_2}^{(ik)}, r_{\mathbf{x}_2}^{(j\ell)}\}$) cf. [16], but, more importantly, the condition $m_1 + m_2 > \min\{r_{\mathbf{x}}^{(ik)}, r_{\mathbf{x}}^{(j\ell)}\}$ shows the benefit of side information in order to obtain the phase transition with a lower number of measurements. In fact, when $r_{\mathbf{x}}^{(ik)} < r_{\mathbf{x}_1}^{(ik)} + r_{\mathbf{x}_2}^{(ik)}$, joint classification of \mathbf{y}_1 and \mathbf{y}_2 leads to a clear advantage in the number of measurements needed to achieve the phase transition with respect to the case in which classification is carried independently from \mathbf{y}_1 and \mathbf{y}_2 , despite the fact that linear measurements are taken independently from \mathbf{x}_1 and \mathbf{x}_2 .

In case 2), the range spaces associated to the input covariance matrices are such that $\text{Im}(\Sigma_{\mathbf{x}_1}^{(ik)}) = \text{Im}(\Sigma_{\mathbf{x}_1}^{(j\ell)})$ and $\text{Im}(\Sigma_{\mathbf{x}_2}^{(ik)}) = \text{Im}(\Sigma_{\mathbf{x}_2}^{(j\ell)})$ so that classification based on the observation of \mathbf{y}_1 or \mathbf{y}_2 alone yields an error floor in the upper bound of the misclassification probability [16]. In other terms, input signals and side information signals from classes (i, k) and (j, ℓ) are never perfectly distinguishable. In this case, the impact of correlation between the input signal and the side information signal is clear when observing (19). In fact, when combining measurements taken independently from the vectors \mathbf{x}_1 and \mathbf{x}_2 , it is possible to drive to zero the misclassification probability, in the low-noise regime, provided that the number of measurements m_1 and m_2 verify the conditions in (19).

Finally, cases 3) and 4) represent intermediate scenarios in which range spaces associated to \mathbf{x}_1 are distinct, but those related to \mathbf{x}_2 are completely overlapping, and vice versa. We note then how the sufficient conditions for phase transition in (20) and (21) are given by combinations of the conditions in (18) and (19).

IV. RECONSTRUCTION WITH SIDE INFORMATION

We now consider signal reconstruction in the presence of side information. We are interested in the asymptotic characterization of the minimum mean-squared error (MMSE)

$$\text{MMSE}(\sigma^2) = \text{E} [\|\mathbf{x}_1 - \hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2)\|^2], \quad (22)$$

where $\hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2)$ is the conditional mean estimator in (5). In the following, we determine conditions on the number of measurements m_1 and m_2 that guarantee perfect reconstruction in the low-noise regime, i.e., when $\sigma^2 \rightarrow 0$, that is

$$\lim_{\sigma^2 \rightarrow 0} \text{MMSE}(\sigma^2) = 0, \quad (23)$$

thus generalizing the results in [17] to the case when side information is available at the decoder.

Note the characterization of such conditions also leads to the characterization of a phase transition region in terms of m_1 and m_2 , where within this region (23) holds and outside the region (23) is not verified.

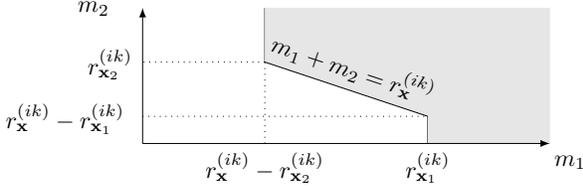


Fig. 3. Representation of the conditions on m_1 and m_2 for phase transition of the MMSE. The shaded region represents values of m_1 and m_2 that satisfy the conditions (24) for a given pair of classes (i, k) .

Theorem 2: Consider the model in (1) and (2), where the input signal \mathbf{x}_1 and the side information signal \mathbf{x}_2 are drawn according to the class-conditioned joint distribution (6). Then, with probability 1, if

$$m_1 > r_{\mathbf{x}_1}^{(ik)} \quad \text{or} \quad \begin{cases} m_1 > r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_2}^{(ik)} \\ m_1 + m_2 > r_{\mathbf{x}}^{(ik)} \end{cases}, \forall (i, k) \in \mathcal{S} \quad (24)$$

then (23) is verified. Conversely, if (23) holds, then, with probability 1, we also have

$$m_1 \geq r_{\mathbf{x}_1}^{(ik)} \quad \text{or} \quad \begin{cases} m_1 \geq r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_2}^{(ik)} \\ m_1 + m_2 \geq r_{\mathbf{x}}^{(ik)} \end{cases}, \forall (i, k) \in \mathcal{S}. \quad (25)$$

The sufficient conditions in (24) show that the numbers of measurements of \mathbf{x}_1 and \mathbf{x}_2 have to be collectively greater than the largest among the dimensions of the spaces spanned by signals $\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T]^T$ in all Gaussian components $(i, k) \in \mathcal{S}$. Moreover, the measurements of \mathbf{x}_1 need to be enough to capture signal components which are not correlated with the side information, for all Gaussian components. Finally, the condition $m_1 > r_{\mathbf{x}_1}^{(ik)}$ is obtained trivially by considering reconstruction of \mathbf{x}_1 from the measurements collected in the vector \mathbf{y}_1 , thus disregarding side information.

It is interesting to note that the necessary conditions for the phase transition of the MMSE of GMM inputs are one measurement away from the corresponding sufficient conditions, akin to our previous results for the case without side information [17], thus providing a sharp characterization of the MMSE phase transition region. Finally, note that the values of m_1 and m_2 that are sufficient for the MMSE phase transition are obtained by considering the intersection of regions akin to that in Fig. 3 for all the pairs of classes $(i, k) \in \mathcal{S}$.

V. CONCLUSIONS

We have considered a linear measurement model, where a decoder has access to noisy linear projections of both the signal of interest and the side information signal, in order to carry out either classification or reconstruction. We have also considered a model where the joint distribution of the signal of interest and the side information, conditioned on some underlying class labels, is a multivariate Gaussian; the marginal distributions of the signal and the side information conditioned on a class label are Gaussian mixtures.

We have provided a characterization of sharp sufficient conditions for a phase transition in the misclassification

probability and necessary and sufficient conditions for the phase transition of the reconstruction error (the performance quantities under consideration), as a function of the geometry of the sources, the geometry of the measurement kernels and their interplay. Our results are reminiscent of the Slepian-Wolf and the Wyner-Ziv conditions for joint source coding and source coding with side information.

REFERENCES

- [1] E. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery for incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [2] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [4] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [5] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal processing with compressive measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 445–460, 2010.
- [6] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [7] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," *arXiv preprint arXiv:1311.0576*, 2013.
- [8] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. I. Phys.*, vol. 35, no. 2, pp. 660–663, 2008.
- [9] L. Weizman, Y. C. Eldar, and D. B. Bashat, "The application of compressed sensing for longitudinal MRI," *arXiv preprint arXiv:1407.2602*, 2014.
- [10] J. F. Mota, N. Deligiannis, and M. R. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *arXiv preprint arXiv:1408.5250*, 2014.
- [11] M. Chen, J. Silva, J. Paisley, D. Dunson, and L. Carin, "Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6140–6155, 2010.
- [12] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, 2012.
- [13] M. Chen, W. Carson, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications inspired linear discriminant analysis," in *Int. Conf. Machine Learn. (ICML)*, Jun.-Jul. 2012.
- [14] J. Yang, X. Yuan, X. Liao, P. Lull, D. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using Gaussian mixture models," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4863–4878, 2014.
- [15] H. Reberdo, F. Renna, R. Calderbank, and M. R. D. Rodrigues, "Compressive classification," in *IEEE Int. Symp. Information Theory (ISIT)*, Jul. 2013.
- [16] —, "Compressive classification of a mixture of Gaussians: Analysis, designs and geometrical interpretation," *arXiv preprint arXiv:1401.6962v1*, 2014.
- [17] F. Renna, R. Calderbank, L. Carin, and M. R. D. Rodrigues, "Reconstruction of signals drawn from a gaussian mixture via noisy compressive measurements," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2265–2277, 2014.
- [18] 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 noisy features in the presence of side information," *arXiv preprint arXiv:1412.0614*, 2014.
- [19] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. New York, NY: Wiley-Interscience, 2000.
- [20] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.